DISSERTATION

LOW-COST EMBEDDED SYSTEMS FOR COMMUNITY-DRIVEN AMBIENT AIR QUALITY MONITORING

Submitted by

Eric Wendt

Department of Mechanical Engineering

In partial full fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: John Volckens

Jeffrey Pierce
Shantanu Jathar
Sudeep Pasricha

ABSTRACT

LOW-COST EMBEDDED SYSTEMS FOR COMMUNITY-DRIVEN AMBIENT AIR QUALITY MONITORING

Fine particulate matter ($PM_{2.5}$) air pollution is a leading cause of death, disease and environmental degradation worldwide. Existing $PM_{2.5}$ measurement infrastructure provides broad $PM_{2.5}$ sampling coverage, but due to high costs (>10,000 USD), these instruments are rarely broadly distributed at community-level scales. Low-cost sensors can be more practically deployed in spatial and temporal configurations that can fill the gaps left by more expensive monitors. Crowdsourcing low-cost sensors is a promising deployment strategy in which sensors are operated by interested community members. Prior work has demonstrated the potential of crowdsourced networks, but low-cost sensor technology remains ripe for improvement. Here we describe a body of work aimed toward bolstering the future of community-driven air quality monitoring through technological innovation. We first detail the development of the Aerosol Mass and Optical Depth (AMODv2) sampler, a low-cost monitor capable of unsupervised measurement of $PM_{2.5}$ mass concentration and Aerosol Optical Depth (AOD), a measure of light extinction in the full atmospheric column due to airborne particles. We highlight key design features of the AMODv2 and demonstrate that its measurements are accurate relative to standard reference monitors. Second we describe a national crowdsourced network of AMODv2s, in which we leveraged the measurement capabilities of the AMODv2 in a network of university students to analyze the relationship between $PM_{2.5}$ and AOD in the presence of wildfire smoke in the United States. Finally, we propose a cloud screening algorithm for AOD measurements using all-sky images and deep transfer learning. We found that our algorithm correctly screens over 95% of all-sky images for cloud contamination from a custom all-sky image data set. Taken as a whole, our work supports community-driven air pollution monitoring by advancing the tools and strategies communities need to better understand the air they breathe.

ACKNOWLEDGMENTS

First and foremost I would like to thank my family. My parents Eric Sr. and Susan raised me with the values and support I needed to achieve my goals. My siblings John, Kathleen, and Caroline, and my sister-in-law Julie (all CSU graduates) are my closest friends and are my daily inspiration.

I thank my advisor, John Volckens, for his guidance, generosity, and wisdom. As an undergraduate senior, John brought me on board to lead a project that would see me all the way through graduate school. John's belief in me as just an undergraduate to take on one of his ambitious projects gave me the confidence to tackle difficult problems going forward. John always made sure I had the resources I needed to succeed, while giving me the freedom to explore creative solutions within the framework of his project vision. Chapter four of this dissertation in particular exists because of the freedom and support John provided. John is also a master writer and presenter. I can only hope my work here does his example justice.

I thank my PhD committee members, professors Jeff Pierce, Shantanu Jathar, and Sudeep Pasricha for their guidance and constructive feedback. I thank Jeff for his expertise, responsiveness, passion, creativity, and attention to detail. Jeff was instrumental in my growth in the field of aerosol science. I could always rely on Jeff for prompt responses to questions and thorough feedback on design concepts and paper drafts. Through his class on aerosol physics and technology, Shantanu taught me the fundamental concepts underlying all of my graduate work. Shantanu's expertise spanned the full breadth of my research. Shantanu provided some of the best guidance on how to produce good works of both engineering and science. Sudeep is largely responsible for expanding my skill set beyond my mechanical engineering background. I took his class on embedded systems early on in my PhD program. It was one of the most rigorous and comprehensive courses I have ever taken, and provided the basis for most of the technology I developed in graduate school. After the course, he continued to provide guidance toward developing my skill in new subjects. Due in large part to Sudeep's influence, I intend to continue working in embedded systems and machine learning as I begin my professional career.

My work would not have been possible without Casey Quinn and Bonne Ford. In the summer of 2015, Casey hired me as an undergraduate assistant to work on a project involving programming sensors, at a time when I had zero programming experience! His folly was my gain, with that summer kick-starting my journey into sensors and embedded systems. More than anybody else, Casey worked with me at bizarre hours to meet seemingly unattainable deadlines and solve some of the most frustrating problems I encountered in graduate school. Casey was the single most supportive colleague for the first half of my degree, when most of the technology was built and painfully debugged. Bonne has been selfless with her time and has always shown genuine interest in my success as a scientist. As an engineer first and foremost, I often struggled to navigate the atmospheric science literature. Bonne was my guide through this unfamiliar territory and any of my contributions to atmospheric science are a result of her support. If Casey pushed me out of the starting gate, Bonne helped me past the finish line as the most supportive colleague for the latter half of my program.

I owe Dan Miller-Lionberg, Christian L'Orange, Jessica Tryner, and Todd Hochwitz a debt of gratitude. Dan, Christian, Jessica, and Todd are some of the most skilled engineers and scientists I have ever met. They were all instrumental in designing the sensors that form the backbone of my graduate work.

Many other scientists and students have contributed to this work. I thank John Mehaffy, Michael Cheeseman, David Hagan, Zoey Rosen, Marilee Long, Nick Trammell-Jamison, Alex Lynch, Ben Strauss, Gabe Neymark, Snoop, Tom Reilly, and Mollie Phillips for their unique and valuable efforts.

I thank all of the volunteers from our crowdsourced air quality monitoring campaigns including members of the Community Collaborative Rain, Hail, and Snow Network (CoCoRhas), NASA's Student Airborne Research Program (SARP), and other volunteers from our local community.

This research was supported by NASA grant 80NSSC18M0120.

DEDICATION

Dedicated to my sister, Kathleen.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

The moment we step outside, we make immediate observations about our environment. Our perceptions of heat, cold, moisture, light, and (if we are near Greely) smell direct our plans and desires for the day ahead. Modern technology allows us to experience our environment with a level of foreknowledge and precision beyond what our natural faculties would allow. With a glance at a screen, we can know what conditions to expect on the top of a ski mountain 100 miles away, within a few degrees or miles-per-hour. With this foreknowledge we can make decisions to maximize not only our short-term comfort, but also our immediate safety and long-term health. We can avoid hypothermia by packing layers before a cold front moves in and travel safely by avoiding bi-planes in gale-force winds.

But what of the air we breathe? People who live near a major industrial center or have been in the path of a wildfire smoke plume are familiar with the irritation, stench, and labored breathing that comes with poor air quality. Our instinctual revulsions toward dirty air are well founded. Polluted air is often partially composed of particulate matter smaller than 2.5 microns in diameter ($PM_{2.5}$), which, when inhaled, can come to rest deep in our lungs, leading to acute and chronic disease (Feng et al., 2016; Janssen et al., 2013; Kim et al., 2019; Pope & Dockery, 2006). Each year, millions of deaths are attributed to $PM_{2.5}$ exposure (Brauer et al., 2016; Forouzanfar et al., 2016). For most people on the planet, assessing risks posed by ambient air pollution in their area may not be as simple as checking the weather. The relative lack of accessible air pollution information has implications for individuals deciding if they ought to go for a morning run, all the way up to national and global regulatory agencies seeking to limit the adverse effects of air pollution on the populations they serve. In order to achieve a higher level of informational accessibility for air pollution, we need monitoring instruments that are accurate, scalable, and economically viable.

Air pollution is measured using ground-based and satellite-based sensors. $PM_{2.5}$ mass concentrations can be measured using a variety of methods. Some systems sample air at a controlled flow rate and mechanically partition particles into those above and below 2.5 microns (e.g. Kelleher et al., 2018;

Volckens et al., 2017). Mass from the flow stream containing $PM_{2.5}$ is deposited on a filter. The average concentration over some period of time can be calculated based on the flow rate and the $PM_{2.5}$ mass deposited. Other sensors (e.g. nephelometers and optical particle counters), measure how light emitted from a laser source is influenced by particles before it reaches a detector opposite a flow stream. Particle concentrations in a column of air spanning the atmosphere between the surface of the earth and outer space can be assessed by instruments known as sun photometers. Sun photometers measure the diminution of sunlight due to airborne particles, a quantity known as Aerosol Optical Depth (AOD).

The last several decades have seen a proliferation of consumer microelectronics. Microelectronic devices today range from plug-and-play black-box products to open-ended hardware and software platforms for do-it-yourself projects. Platforms like Arduino® empower individuals to create their own circuits and write their own software for projects that interest them. What was once the purview of professional engineers is now open to everyone from elementary school students to retirees. The widespread availability of device development platforms has manifested in the emergence of citizen science as a practical means for community members of all backgrounds to participate in air pollution science. The result of emerging technological and philosophical innovation in citizen science has been a new paradigm of air pollution measurement featuring lower-cost devices that can be readily operated by people of all backgrounds.

When air pollution monitors are sufficiently low-cost, they can be distributed at high spatial resolutions, providing local-scale air pollution information. Low-cost sensors are increasingly being distributed to citizen volunteers to form crowdsourced monitoring networks. Some crowdsourced networks are already in place and making valuable contributions to community air quality monitoring (Badura et al., 2020; Chadwick et al., 2021; Ford et al., 2019; Gupta et al., 2018; Li et al., 2020; Lin et al., 2020; Y. Lu et al., 2021). However, low-cost measurements are typically limited to a single measurement modality (i.e. laser-light scattering) and are prone to bias (Kelly et al., 2017; Zheng et al., 2018; Levy Zamora et al., 2019; Sayahi et al., 2019; Ford et al., 2019; Tryner et al., 2020; Barkjohn et al., 2021). The shortcomings of existing low-cost sensors highlight the need for innovative technology to power the future of community-driven air quality monitoring.

In this body of work, we describe three unique contributions toward community-driven air pollution measurement. First, we present the design and validation of a novel low-cost monitor capable of simultaneous measurement of $PM_{2.5}$ and AOD known as the Aerosol Mass and Optical Depth (AMODv2) sampler. Second, we describe a deployment of AMODv2s in a crowdsourced network of university students in the contiguous United States in a study of the effects of wildfire smoke on the relationship between AOD and $PM_{2.5}$. Finally, we develop a platform independent quality control procedure for AOD measurement based on images of the sky and machine vision techniques.

CHAPTER 2. A LOW-COST MONITOR FOR SIMULTANEOUS MEASUREMENT OF FINE PARTICULATE MATTER AND AEROSOL OPTICAL DEPTH: AUTOMATION AND DESIGN IMPROVEMENTS

Reproduced (or 'Reproduced in part') with permission from "A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth - Part 3: Automation and design imporvements". Eric A. Wendt, Casey Quinn, Christian L'Orange, Daniel D. Miller-Lionberg, Bonne Ford, Jeffrey R. Pierce, John Mehaffy, Michael Cheeseman, Shantanu H. Jathar, David H. Hagan, Zoey Rosen, Marilee Long, and John Volckens; Atmospheric Measurement Techniques 2021 14, 6023–6038; DOI: 10.5194/amt-14-6023-2021; Copyright 2021 Copernicus Publications.

Chapter Overview

Atmospheric particulate matter smaller than 2.5 micrometers in diameter (PM2.5) has a negative impact on public health, the environment, and Earth's climate. Consequently, a need exists for accurate, distributed measurements of surface-level PM2.5 concentrations at a global scale. Existing PM2.5 measurement infrastructure provides broad PM2.5 sampling coverage, but does not adequately characterize community-level air pollution at high temporal resolution. This motivates the development of low-cost sensors which can be more practically deployed in spatial and temporal configurations currently lacking proper characterization. In part 1 of this series we described the development and validation of a first-generation device for low-cost measurement of AOD and PM2.5: The Aerosol Mass and Optical Depth (AMODv1) sampler. Part 2 of the series describes a citizen-science field deployment of the AMODv1 device. Here in part 3, we present an updated version of the AMOD, known as AMODv2, featuring design improvements and extended validation to address the limitations of the AMODv1 work. The AMODv2 measures AOD and PM2.5 at 20-minute time intervals. The sampler includes a motorized sun-tracking system alongside a set of four optically filtered photodiodes for semi-continuous, multi-wavelength (current version at 440, 500, 675, and 870 nm) AOD sampling. Also included are a Plantower PMS5003 sensor for

time-resolved optical PM2.5 measurements and a pump/cyclone system for time-integrated gravimetric filter measurements of particle mass and composition. AMODv2 samples are configured using a smartphone application and sample data are made available via data streaming to a companion website (csu-ceams.com). We present the results of a nine-day AOD validation campaign where AMODv2 units were co-located with an AERONET (Aerosol Robotics Network) instrument as the reference method at AOD levels ranging from 0.02 ± 0.01 to 1.59 ± 0.01. We observed close agreement between AMODv2s and the reference instrument with mean absolute errors of 0.04, 0.06, 0.03, and 0.03 AOD units at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. We derived empirical relationships relating the reference AOD level with AMODv2 instrument error and found that the mean absolute error in the AMODv2 deviated by less than 0.01 AOD units between clear days and elevated-AOD days and across all wavelengths. We identified bias from individual units, particularly due to calibration drift, as the primary source of error between AMODv2s and reference units. In a test of 15-month calibration stability performed on 16 AMOD units, we observed median changes to calibration constant values of -7.14%, -9.64%, -0.75%, and -2.80% at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. We propose annual recalibration to mitigate potential errors from calibration drift. We conducted a trial deployment to assess the reliability and mechanical robustness of AMODv2 units. We found that 75% of attempted samples were successfully completed in rooftop laboratory testing. We identify several failure modes in the laboratory testing and describe design changes we have since implemented to reduce failures. We demonstrate that the AMODv2 is an accurate, stable and low-cost platform for air pollution measurement. We describe how the AMODv2 can be implemented in spatial citizen-science networks where reference-grade sensors are economically impractical and low-cost sensors lack accuracy and stability.

Introduction

Fine particulate matter air pollution (PM2.5) is a leading cause of human morbidity and mortality, and also a significant contributor to radiative climate forcing (Myhre et al., 2013; Forouzanfar et al., 2016; Brauer et al., 2016; Vohra et al., 2021). Inhaled $PM_{2.5}$ can penetrate deep into the lungs, leading to both acute and chronic health impacts (Pope & Dockery, 2006; Janssen et al., 2013; Feng et al., 2016; Kim et al., 2019). Each year, millions of deaths worldwide are attributed to $PM_{2.5}$ exposure (Brauer et al., 2016; Forouzanfar et al., 2016). In addition to public health, $PM_{2.5}$ also contributes to visual degradation of the atmosphere and affects the climate by influencing Earth's radiative budget (Myhre et al., 2013). Regions with the highest levels of air pollution often lack adequate ground level monitoring (Snider et al., 2015; Brauer et al., 2016). Thus, disease estimates for much of the world's population rely on exposure estimates where satellite data or model simulations are the best or only source of information on human exposure. Installing a global network of reference-grade surface monitors is not currently feasible due to the high installation and maintenance costs.

Satellite remote sensing, supplemented with data from surface measurements and chemical transport models (CTMs), represents the state-of-the-art for global $PM_{2.5}$ monitoring at relatively high temporal and spatial resolution (van Donkelaar et al., 2016, 2019; Hammer et al., 2020; Lee, 2020). Measurements from satellite instruments, such as the Moderate Resolution Imaging Spectrometer (MODIS) and the Multi-angle Imaging SpectroRadiometer (MISR) (Salomonson et al., 1989; Diner et al., 1998), are used to estimate surface-level $PM_{2.5}$ concentrations (e.g Liu et al., 2005), which in turn have facilitated research on the health effects associated with $PM_{2.5}$ exposure (Brauer et al., 2016; Forouzanfar et al., 2016; Li et al., 2018; X. Lu et al., 2019). Satellites equipped with aerosol remote sensing instrumentation retrieve aerosol optical depth (AOD), a measure of light extinction in the atmospheric column, which can then be converted to ground level $PM_{2.5}$ using a CTM or statistical relationship (Y. Liu et al., 2005; van Donkelaar et al., 2006, 2010, 2012, 2016; Hammer et al., 2020). The relationship between AOD and $PM_{2,5}$ can be expressed as follows (Y. Liu et al., 2005):

$$PM_{2.5} = \eta \cdot AOD \qquad (2\text{-}1)$$

where η is a conversion factor between $PM_{2.5}$ and AOD. The uncertainty of surface-level $PM_{2.5}$ concentrations derived from satellite observations has two main components: 1) the uncertainty of the satellite AOD measurement and 2) the uncertainty of the modeled $PM_{2.5}$ to AOD ratio (η) (e.g. Ford and Heald, 2016; Jin et al., 2019).

The error of the satellite AOD retrieval can be estimated using ground-level AOD measurements from instruments known as sun photometers (e.g., Sayer et al., 2012). The Aerosol Robotics Network (AERONET) provides reference-quality AOD measurements at hundreds of locations around the Earth; these data are used to constrain and reduce uncertainties in AOD values (Holben et al., 1998). AERONET instruments are rarely deployed at high spatial density (i.e. sub-city scale), outside of field campaigns (e.g. Garay et al., 2017), due to the high cost of the instrument and supporting equipment (>50,000 USD). Determining the uncertainty in the modeled $PM_{2.5}$ to AOD ratio requires co-locating AOD and $PM_{2.5}$ measurements. The Surface PARTiculate mAtter Network (SPARTAN) was established to provide co-located $PM_{2.5}$ and AOD reference measurements and to evaluate uncertainties in both AOD and the $PM_{2.5}$ to AOD ratio; however, the number of SPARTAN sites worldwide is limited by number (~20 active sites), equipment, and operational costs (Snider et al., 2015).

Networks of low-cost nephelometers (notably the Plantower PMS5003), have been suggested and deployed in large numbers as a means to provide surface $PM_{2.5}$ data at a higher spatial density than can be achieved with reference-grade monitors (Lin et al., 2020; Li et al., 2020; Badura et al., 2020; Y. Lu et al., 2021; Chadwick et al., 2021). However, low-cost sensors (or more specifically, the Plantower PMS5003 devices) tend to exhibit measurement bias (Kelly et al., 2017; Zheng et al., 2018; Levy Zamora et al., 2019; Sayahi et al., 2019; Tryner et al., 2020), requiring correction relative to reference monitors (Ford et al., 2019b; Wendt et al., 2019a). Low-cost Sun photometers have been deployed at high-spatial resolution to evaluate satellite AOD uncertainty as part of the Global Learning and Observations to Benefit the Environment (GLOBE) program (Boersma & de Vroom, 2006; Brooks & Mims, 2001a). GLOBE Sun photometers were operated by students as part of education programming, resulting in over 400 measurements between January 2002 and October 2005 in the Netherlands (Boersma & de Vroom, 2006). These data were used to evaluate satellite-derived AOD in corresponding regions. However, the authors noted difficulty coordinating with schools to achieve consistent measurements, specifically those corresponding with satellite overpasses. Collectively, these previous efforts have advanced the understanding of AOD and $PM_{2.5}$:AOD variability considerably. However, there is still demand for co-located $PM_{2.5}$ and AOD samplers deployed at higher spatial density and with greater temporal resolution (Ford & Heald, 2016; Garay et al., 2017; Jin et al., 2019). Samplers used in these networks must be sufficiently low-cost to deploy in large numbers, have manageable operational and maintenance requirements, and provide useful and reliable PM2.5 and AOD measurements (i.e., measurement data of sufficient accuracy and precision so as to support scientific inference or public decision-making). Thus, consideration should be given to the tradeoffs associated with deploying low-cost sensors such as scalability and simplicity versus accuracy and reliability.

In part 1 of this series of articles, we describe a low-cost, compact $PM_{2.5}$ and AOD ground monitor (Wendt et al., 2019a; Ford et al., 2019b). The device, known as the Aerosol Mass and Optical Depth (AMOD) sampler, featured a $PM_{2.5}$ cyclone inlet for integrated gravimetric sampling and composition analysis, a low-cost nephelometer (Plantower PMS5003, Beijing, China) for real-time $PM_{2.5}$ mass estimate, and four filtered-photodiode (Intor Inc., Socorro, NM, USA) sensors at 440, 520, 680, and 870 nm for measuring AOD. Here, we refer to this earlier instrument as the AMODv1. The assembly cost for the first manufacturing set of 25 AMODv1s was under 1,100 per unit USD (Wendt et al., 2019a). The results of a field validation campaign revealed agreement to within 10% (mean relative error) for AOD values relative to co-located AERONET instruments. The mean AOD difference was <0.01 with 95% confidence upper and lower limits of agreement of 0.03 and -0.02, respectively. With respect to $PM_{2.5}$, the AMODv1 filter measurements agreed within 8% (mean relative error) relative to Federal Equivalent Method (FEM) monitors from the Environmental Protection Agency (EPA), with a mean difference of -0.004 $\mu g\ m^{-3}$ and 95% confidence upper and lower limits of agreement of 1.84 and -1.85 $\mu g\ m^{-3}$, respectively (Wendt et al., 2019a). With respect to real-time PMS5003 $PM_{2.5}$ measurements, the mean relative error between the AMODv1 and an FEM monitor was 1.98 $\mu g\ m^{-3}$ with and mean difference of 0.04 $\mu g\ m^{-3}$ and 95% confidence upper and lower limits of agreement of 5.02 and -4.95 $\mu g\ m^{-3}$, respectively (Wendt et al., 2019a). These results indicated that the AMODv1 accurately quantified surface $PM_{2.5}$ concentrations and AOD simultaneously and at a substantially lower cost and smaller size than existing equipment. To test implementation of the AMODv1, we constructed and deployed 25 AMODv1s in a citizen-science network, as documented in part 2 in this series (Ford et al., 2019b).

Despite the promise of the AMODv1, the initial deployment highlighted several key limitations. First, the AMODv1 lacked quality control measures for misalignment or cloud contamination during the measurement period. Second, the instrument had limited temporal resolution for AOD (typically one measurement per day). Third, despite the presence of a visual alignment aid (Wendt et al., 2019), many volunteers found it difficult to align the instrument with the sun, which was compounded by inconsistent standards as to what constituted proper alignment. Fourth, data could not be transmitted wirelessly or accessed remotely. The first objective of this current work was to address these four major limitations of the AMODv1 design. Another shortcoming of our work on AMODv1 was limited stability analysis of the AOD sensors across varying atmospheric conditions and over time. The second objective of this work, therefore, was to evaluate the stability of the AOD sensors across a range of pollution levels and to assess the stability of the AOD sensors after repeated deployments over the course of a year. Here, we describe our design changes and extended validation efforts toward our research objectives. First, we summarize the design advantages of the AMODv2 relative to the AMODv1. Second, we present the results from a validation campaign where AMODv2 units were co-located with reference instruments. Third, we analyze the stability of AMODv2 AOD measurements after 15 months of use. Finally, we analyze the reliability of the AMODv2 design in a series of laboratory experiments. The results presented here demonstrate that AMODv2 is a practical option to establish spatially-dense $PM_{2.5}$ and AOD measurement networks. Applied in these networks, the AMODv2 will close gaps in the existing global aerosol measurement infrastructure of ground-based and satellite-based observations.

Materials and Methods

Instrument design

We designed the AMODv2 to sample integrated gravimetric $PM_{2.5}$ mass concentration, real-time $PM_{2.5}$ mass concentration, and AOD simultaneously. One intended application is large-scale sampling campaigns with the AMODv2 instruments operated by volunteers with little to no background in aerosol or atmospheric science (Ford et al., 2019b). Thus, we prioritized a design that is low-cost, accurate, mechanically robust, portable, automated, and user-friendly. We provide images of AMODv2 hardware in Fig. 2-1, highlighting key internal and external components.

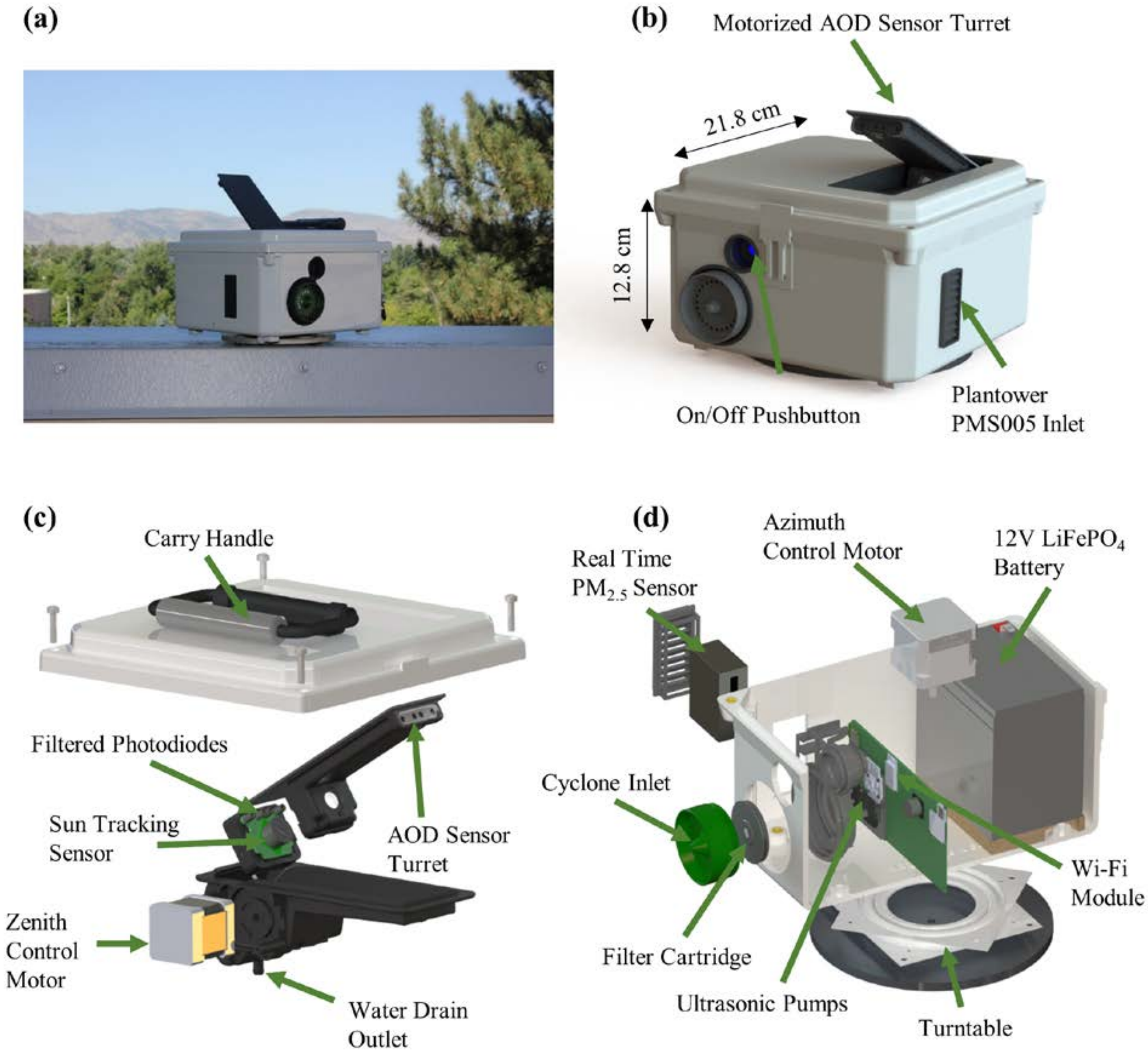

Figure 2-1: Images detailing external and internal AMODv2 design and hardware. a) Photograph of AMODv2 sampling outdoors. b) External computer animated rendering of AMODv2 features and dimensions. c) Computer generated exploded view of AOD measurement subsystem. d) Computer generated exploded section view of $PM_{2.5}$ sampling, wireless data transfer, and power subsystems.

The AMODv2 measures AOD at 440 nm, 500 nm, 675 nm, and 870 nm using optically filtered photodiodes (Intor Inc., Socorro, NM, USA) with narrow bandwidth (<15 nm at full-width half-maximum signal). The measurement process is fully automated using a solar tracking system (Section 2.3), reducing the potential for misalignment due to user error. Movement in the zenithal plane is achieved using a custom turret module embedded in the interior of the AMODv2 enclosure (Fig. 2-1a). The module was designed in

SolidWorks® (ANSYS, Inc., Canonsburg, PA, USA) and built using multi-jet fusion printing. The module houses a custom printed circuit board containing the solar tracking sensors and the filtered photodiodes. Light enters the turret through four, 4 mm apertures, and passes through 112 mm tubes to reach the filtered photodiodes (Fig. 2-1c). These proportions yield a viewing angle of approximately 2 degrees for each photodiode sensor element. A stepper motor (Stepper Online 17HS10-0704S-C2, Nanjing City, China), fixed to the turret, actuated the zenithal rotation. Movement in the azimuthal plane is actuated using a second stepper motor (Stepper Online 17HS19-1684S-C6, Nanjing City, China) fixed to a turntable and base-plate assembly (McMaster Carr 6031K16, Elmhurst, IL, USA), which enables 360 degree rotation of the AMODv2. The angular resolution of each stepper motor is tuned to 0.056 degrees using programmable drivers (Texas Instruments DRV8834RGER, Dallas, Texas, USA). Active tracking is accomplished using closed-loop control enabled by a 3-axis accelerometer (STMicroelectronics LSM6DSM, Geneva, Switzerland), a GPS module (u-blox CAM-M8, Thalwil, Switzerland), and a quadrant photodiode solar tracking sensor (Solar MEMS NANO-ISS5, Seville, Spain).

The AMODv2 measures $PM_{2.5}$ using both real-time and time-integrated techniques. Real-time $PM_{2.5}$ concentrations are measured and streamed using a light-scattering $PM_{2.5}$ sensor (Plantower PMS5003, Beijing, China). A 3D-printed fixture secured the sensor in position to sample ambient air, while downward sloping vents protect the sensor from water ingress (Fig. 2-1d). $PM_{2.5}$ concentrations are evaluated on the PMS5003 chip via a manufacturer proprietary algorithm. The AMODv2 reports the $PM_{2.5}$ values corrected by Plantower's proprietary atmospheric correction. These values are accessed by the AMODv2 microcontroller via serial communication. A flow chart detailing the $PM_{2.5}$ measurement protocol is provided in Fig. A1.

For time-integrated $PM_{2.5}$ mass concentration measurement, we leveraged a $PM_{2.5}$ cyclone design from prior studies (Volckens et al., 2017; Kelleher et al., 2018; Wendt et al., 2019a). The main circuit board features three ultrasonic pumps (Murata MZBD001, Nagaokakyo, Japan) and a mass flow sensor (Honeywell Omron D6F, Charlotte, NC, USA,) to control the flow of air through a custom aluminum cyclone and filter cartridge with a 50% cut point of 2.5 µm (Fig. 2-1d). The gravimetric sample is collected on a 37mm Teflon filter secured within a filter cartridge. Sampled particles are collected on a single filter that is pre and post weighed for each sample. During deployment, a field blank is carried along with the sampler to correct for incidental mass contamination or drift.

The AMODv2 is powered using a 12 V, 10 Ah $LiFePO_4$ battery (Dakota Lithium, Grand Fork, ND, USA) with a secondary 12 V, 3.3Ah $LiFePO_4$ (Battery Space, LFH4S4R1WR-C5, Richmond, CA, USA) battery in parallel. The battery is charged using a barrel plug inlet accessible on the side of the enclosure. A detachable rubber plug seals the inlet from the outside environment when not charging. Charging circuitry supports charging at a rate of 3.0 A, enabling a full charge in approximately eight hours. A full charge can power the AMODv2 for over 120 hours.

The AMODv2 records and wirelessly transfers meteorological and quality-control data in real time. Meteorological data include ambient temperature (°C), ambient pressure (hPa), and relative humidity (%). Quality control metrics include sample duration (s), sample flow rate (L min^{-1}), total sampled volume (L), battery temperature (°C), battery voltage (V), battery state of charge (%), current draw (mA), and wireless signal strength (RSSI).

The external housing of the AMODv2 (Fig. 2-1b) is made from a weather-resistant NEMA electrical enclosure (Polycase, YQ-080804, Avon, Ohio, USA). The dimensions of a fully assembled AMODv2 are 21.8 cm W × 21.8 cm L × 12.8 cm H, with a weight of 3.1 kg. A folding carry handle is fixed to the upper surface of the enclosure to aid transport (Fig. 2-2b). The total cost of the AMODv2 was 1,175 USD per unit, for a production run of 100 units (Table A1). This tabulation includes an estimated three hours of assembly at a rate of 25 USD per hour.

We developed the AMODv2 control software using an online, open-source platform (mbed™; ARM® Ltd., Cambridge, UK). The software was written in C++ and executed by a 64-bit microcontroller (STMicroelectronics STM32L476RG, Geneva, Switzerland). We implemented wireless data transfer using a Wi-Fi and Bluetooth™ module (Espressif Systems ESP32-C3-WROOM, Shanghai, China). A MicroSD card stores all data for data backup or offline deployment (Molex 5031821852, Lisle, IL, USA). We integrated software modules for AOD, real-time $PM_{2.5}$, gravimetric $PM_{2.5}$, data logging, and wireless data transfer using a real-time operating system (RTOS) for pseudo-simultaneous software execution.

AOD measurement and solar tracking

The AMODv2 applies the Beer-Lambert-Bouguer law to calculate AOD (a). This relationship, expressed in terms of measurable parameters, is as follows:

$$\tau_a(\lambda) = \frac{1}{m}\left(ln\left(\frac{V_0}{R^2}\right) - ln(V)\right) - \tau_R(\lambda, p) - \tau_{O3} \quad (2\text{-}2)$$

where *m* is the unitless air mass factor, which accounts for the increased air mass that light passes through as the sun approaches the horizon, *R* is the Earth-sun distance in astronomical units (AU), *V* is the signal produced by the light detector in volt, τ_R accounts for Rayleigh scattering by air molecules, *p* is the pressure at the sensor in Pa, λ is the sensor wavelength in m, τ_{O3} accounts for ozone absorption, and V_0 is the extraterrestrial constant in volts, which is the sensor signal if measured at top-of-atmosphere and is determined via calibration. AOD values at 440 nm, 500 nm, 675 nm, and 870 nm are calculated using Eq. (2-2). The Earth-Sun distance, R, is computed directly from GPS data and the solar positioning algorithm. *V* is the signal produced by the photodiode and V_0 is accessed from on-chip memory. The relative optical air mass factor is computed as a function of solar zenith angle (θ) as follows (Young, 1994):

$$m = \frac{1.002432 \cdot cos^2(\theta) + 0.148386 \cdot cos(\theta) + 0.0096467}{cos^2(\theta) + 0.149864 \cdot cos^2(\theta) + 0.0102963 \cdot cos(\theta) + 0.000303978} \quad (2\text{-}3)$$

AOD calibration procedure

The extraterrestrial constants for all AMODv2s were evaluated via calibration relative to AERONET sun photometers (Cimel CE318, Paris, France) (Holben et al., 1998). AERONET instruments report AOD at 340 nm, 380 nm, 440 nm, 500 nm, 675 nm, 870 nm, 1020 nm, and 1640 nm (Holben et al., 1998). We selected the four AMODv2 AOD wavelengths in part for direct comparison with AERONET instruments. We conducted calibrations at the MAXAR-FUTON site in Fort Lupton, Colorado (40.036 N, 104.885 W) between November 2019 and February 2020. AMODv2 units were co-located within 50 m of the AERONET instrument and sampled for 2 to 3 hours at a rate of one sample every 2.5 to 3 minutes (note: AERONET instruments are programmed to record AOD every 15 minutes so we oversampled the AMODv2 to achieve sufficient temporal overlap with AERONET). AMODv2 and AERONET level 1.0 measurements concurrent within 60 seconds of each other were included in the calibration data set (Holben et al., 1998). For each set of concurrent measurements, we calculated the extraterrestrial constant by applying Eq. (2-2) solved for V_0, where V was the raw voltage reported by the AMODv2, and τ_a was the AOD reported by the AERONET instrument. The AMODv2 calibration constants were the average value of V_0 for a given instrument and wavelength.

User operation and measurement procedure

We designed the AMODv2 to be operated by individuals without a background in aerosol sampling. We developed a standard procedure that is detailed in a user manual provided as supplementary material. After the initial setup, the AMODv2 requires no operator inputs for the duration of the sample. A flow chart outlining the manual and automatic steps to perform an AMODv2 measurement is provided in Fig. 2-2.

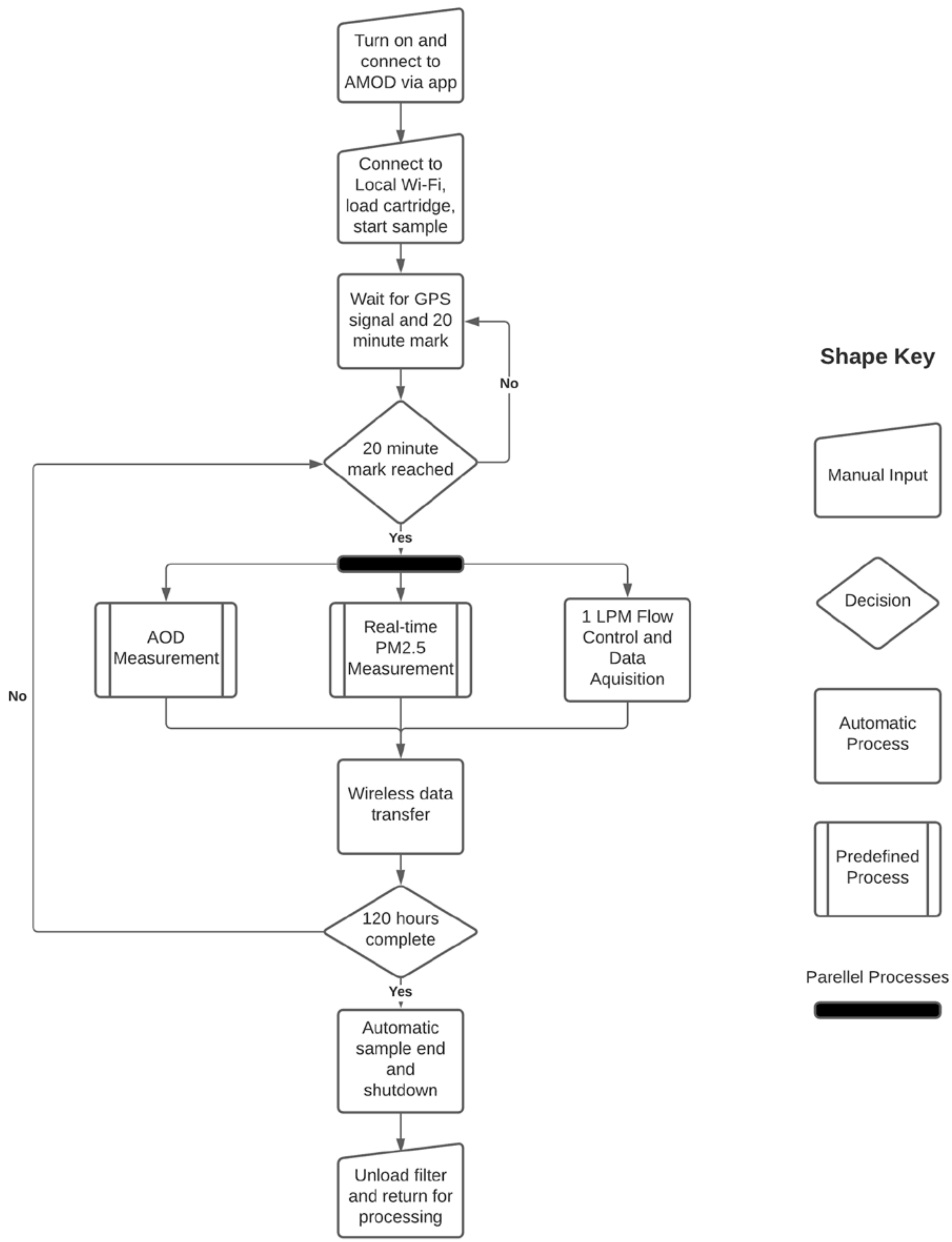

Figure 2-2: Overall device operation flow diagram for a single sample. After each sample, the AMODv2 must be recharged for at least eight hours. Manual inputs require operator intervention. Automatic processes are executed with no operator intervention. Predefined processes are detailed in supplemental Figs. A1-A6. Parallel processes are executed pseudo-simultaneously using a real-time operating system.

Materials needed to initiate a sample include an AMODv2, a cartridge loaded with a pre-weighed filter, and a smartphone with the AMODv2 control application installed ("CEAMS"; available on the Apple App Store and Google Play). A detailed description of the mobile application is in the user manual, which is included as a supplement to this work. After executing an initialization routine by selecting "Scan for Device", the operator may connect to their device via Bluetooth™ using the mobile application. The operator can select a wireless network and input the proper credentials to connect the AMODv2 to the internet. The application then prompts the operator to scan the QR code on the back of the filter cartridge to link the filter with the upcoming sample in the data log. After the cartridge is manually loaded into the compartment behind the inlet (Fig. 2-1b), the AMODv2 should be placed on a flat surface with an unobstructed view of the sun. The operator then starts the sample from the mobile application. After an initial data push, the sample begins at the next 20 minute mark (e.g. 12:00, 12:20, or 12:40). The AMODv2 begins sampling air through the inlet at 1 L min^{-1} and continues to do so for the remainder of the 120-hour sampling period. Real-time $PM_{2.5}$ and AOD measurements are initiated at each 20 minute mark from the start of the sample. The $PM_{2.5}$ reported at each 20 minute interval is the average of measurements taken every 10 seconds over a period of 3 minutes. If the sun is less than 10 degrees above the horizon, the motors do not activate and the solar tracking algorithm is not executed. After each AOD and $PM_{2.5}$ measurement is completed, data are uploaded to the affiliated website (csu-ceams.com), where real-time visualizations of AOD and $PM_{2.5}$ are available. Data reported to the website are accessible with a map-based user interface. Quality-control data are available to research staff via a private administrator portal. A snapshot example of the website is provided in Fig. A7. At the conclusion of a sample, the operator removes the filter cartridge. Upon receipt of the filters, the CEAMS team stored the filters in the refrigerator until mailed to minimize loss of volatile compounds. Complete data files can be downloaded from the website or accessed via a MicroSD card. Individual measurements of AOD and $PM_{2.5}$, from which averages are derived, are available in the complete file, facilitating post-sample uncertainty analysis of $PM_{2.5}$ and AOD measurements.

Validation, stability, and reliability studies

We assessed precision and bias of AMODv2 AOD sensors relative to an AERONET monitor at the NEON-CVALLA site in Longmont, Colorado (40.160 N, 105.167 W) between June 2020 and December 2020 (Holben et al., 1998). We co-located our instruments within 50 m of the reference instrument (and within 5 m of each other) on nine separate days with varying atmospheric conditions (e.g. wildfire smoke and clean air) using a total of 14 unique AMODv2 units. Each test consisted of 2 to 4 hours of sampling at a rate of one sample approximately every 3 minutes. The AERONET reference monitor sampled at a frequency of one sample approximately every 15 minutes. AMODv2 and AERONET measurements concurrent within 2 minutes were included in the validation data set. The accuracy of AMODv2 AOD measurements was assessed via Deming regression.

We evaluated the long-term stability of the AOD sensors by re-calibrating a set of 16 AMODv2 units 15 months after their initial calibration. Original calibrations for the units tested were conducted at the MAXAR-FUTON site in Fort Lupton, Colorado, USA (40.036 N, 104.885 W) on February 21, 2020. Re-calibrations were conducted at the NEON-CVALLA site on May 27, 2021 (The MAXAR-FUTON site was indefinitely unoperational at the time of the second calibration).

We tested the reliability of AMODv2 instruments in a series of 5-day, outdoor samples on the roof of a Colorado State University laboratory facility (430 N College Avenue, Fort Collins, Colorado, USA). All units were co-located within a 10 m radius. We started tests on January 16, 2021, January 30, 2021, and March 31, 2021, which included 34, 27, and 15 unique AMODv2 units respectively, for a total of 76 samples. We assessed the reliability of the AMOD according to the rate at which samples terminated prematurely. Samples that failed to reach at least 115 hours of the intended 120 hour sample duration were designated as premature terminations. We specifically assessed the mechanical robustness of AMODv2 units by visually inspecting failed units for evidence of water ingress and electrical component damage. We also analyzed the AOD data from these samples to evaluate the automatic solar alignment procedure and quality control algorithm.

Compared with our prior work (Wendt et al., 2019a), we tested the AMODv2 AOD measurement system under a broader range of atmospheric conditions. A sizable portion of validation measurements were taken under heavy smoke caused by the Cameron Peak and East Troublesome fires of 2020. We conducted additional testing under more moderate smoke and clear conditions. AOD values reported by AERONET during validation experiments ranged from 0.035 ± 0.01 to 1.59 ± 0.01 at 440 nm, 0.030 ± 0.01 to 1.51 ± 0.01 at 500 nm, 0.021 ± 0.01 to 1.130 ± 0.01 at 675 nm, and 0.016 ± 0.01 to 0.770 ± 0.01 at 870 nm.

Results and discussion

Summary of design improvements

With the AMODv2 design presented here, we addressed the key shortcomings that we identified with AMODv1 enumerated in the Introduction. First, AOD quality control was addressed with motorized solar tracking and a cloud screening protocol. AMODv2 AOD measurements are taken as triplets, facilitating the application of screening protocols based on temporal variation (Smirnov et al., 2000; Giles et al., 2019). The availability of full data files at the end of each sample facilitates additional screening based on hourly and daily variations in AOD values, beyond the immediate quality controls applied to triplets. Second, insufficient temporal resolution was addressed by automating AOD measurement and increasing the sample rate. With automatic sampling in place, units measure every 20 minutes of daylight for up to five days. This updated protocol increases the likelihood that measurements will be available at the desired times of day (e.g. satellite overpass times). Third, we reduced the potential for operator error by eliminating the manual alignment requirement present in the prior design via solar tracking. Fourth, we improved data accessibility through the integration of a Wi-Fi module and a user-friendly website interface. These design changes were achieved while adding only 75 USD to the manufacturing cost, relative to AMODv1 (Table A2). The most important design changes from AMODv1 to AMODv2 are summarized in Table 2-1.

Table 2-1: Design comparison between AMODv1 and AMODv2

Design specification	AMODv1	AMODv2
Sample interval	48 hours	120 hours
Sample flow rate	2 L min^{-1}	1 L min^{-1}
Sun alignment procedure	Manual using pinhole aperture target	Automatic dual-axis closed-loop sun tracking system
AOD cloud screening	None available	Automatic AOD triplet measurement screening protocol
AOD measurement frequency	1 measurement per day	1 measurement every 20 minutes during daytime hours
Data logging	MicroSD card	MicroSD card, wireless data transfers every 20 minutes, and complete file wireless data transfer at the end of each sample
Data visualization	None available	Real-time $PM_{2.5}$ and AOD plots on website
Real-time debugging information	None available	Sample flow rate, total sampled volume, battery temperature, battery voltage, state of charge, current draw, and wireless signal strength
Manufacturing Cost in USD (As of July 2019)	1,100	1,175

We conducted a sample deployment of 10 AMOD units during a wildfire smoke event in Fort Collins, Colorado in October of 2020. The purpose of this deployment was to highlight the design advantages of the AMODv2 in the context of rapidly changing air quality. The results of the deployment are detailed in the first supplement to this work (Figs. A8 and A9).

AOD sensor validation and calibration stability

Here, we present results of co-located validation studies for the AOD measurement system. Our cyclone-based gravimetric $PM_{2.5}$ sampling system has been validated extensively in prior work and shown to agree closely with reference $PM_{2.5}$ monitors (Volckens et al., 2017; Arku et al., 2018; Kelleher et al., 2018; Pillarisetti et al., 2019; Wendt et al., 2019a). Plantower light scattering sensors have likewise been evaluated extensively in prior work (Kelly et al., 2017; Zheng et al., 2018; Levy Zamora et al., 2019; Sayahi et al., 2019; Wendt et al., 2019a; Bulot et al., 2019; Tryner et al., 2020).

We observed close AOD agreement between AMODv2 and AERONET instruments. Correlation plots on the full set of measurement pairs are provided in Fig. 2-3 (n = 426 paired measurements per wavelength). Summary statistics calculated on the full set of measurement pairs across all measurement conditions are provided for each wavelength in Table 2-2.

Table 2-2: Summary statistics for AMODv2 vs. AERONET co-located tests

Wavelength (nm)	Mean absolute error (AOD)	Deming slope coefficient	R^2	AOD Precision (AOD)
440	0.04	0.953	0.987	0.02
500	0.06	0.985	0.978	0.03
675	0.03	1.011	0.995	0.01
870	0.03	1.015	0.977	0.02

Summary statistics on the data set partitioned into clear and elevated-AOD samples are presented in Table A1. The definitions of clear and elevated-AOD samples are explained in the description of Table A1. The mean absolute errors for the full data set were 0.04, 0.06, 0.03, and 0.03 AOD units at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. The Deming regression slope coefficients were 0.953, 0.985, 1.011 and 1.015 at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. The squares of Pearson correlation coefficients were 0.987, 0.978, 0.995, and 0.977 at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. With respect to precision, the average differences from the mean for units measuring coincidentally (i.e. the average amount an individual unit deviated from the mean of all units measuring at the same time) were 0.02, 0.03, 0.01, and 0.02 AOD units at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. With respect to stability across AOD magnitude, the mean absolute error deviated by less than 0.011 between clear days and elevated-AOD days across all wavelengths (Table A1).

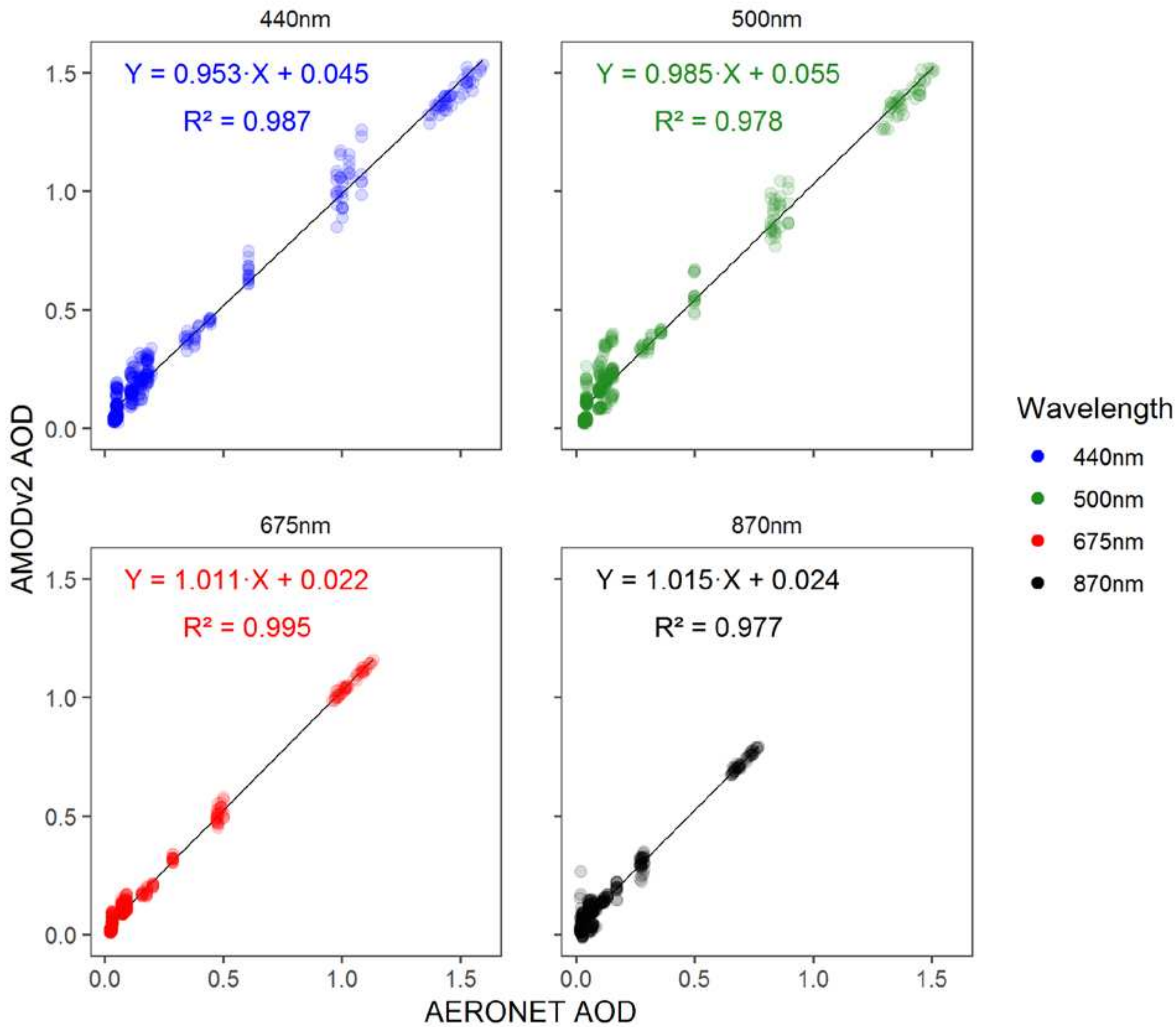

Figure 2-3: AERONET (MAXAR-FUTON site in Fort Lupton, Colorado, USA) vs. AMODv2 AOD co-located comparison (n=426) results with panels separated by wavelength. Lines of best fit were calculated via deming regression analysis.

Due to the broad range of AOD levels during testing, global summary statistics do not fully capture how error and precision scales with increasing AERONET AOD, as these figures of merit are not constant across the range of measured AOD values (Fig. 2-4). Measurements at high AOD impact the mean absolute error disproportionately, while measurements at low AOD impact the mean percent error disproportionately. We derived expected error (EE) equations to constrain the error of AMODv2 measurements relative to AERONET as a function of AOD (following the form used in the validation of satellite AOD products compared to AERONET AOD). We derived the equations iteratively by adjusting the constant and linear terms until the bounds defined by Eqs. (2-4) through (2-7) each contained 85% of the co-located measurement pairs for each wavelength.

$$EE_{440} = \pm(0.080 + 0.050 \cdot AOD_{AERONET440}) \tag{2-4}$$

$$EE_{500} = \pm(0.090 + 0.040 \cdot AOD_{AERONET500}) \tag{2-5}$$

$$EE_{675} = \pm(0.045 + 0.020 \cdot AOD_{AERONET675}) \tag{2-6}$$

$$EE_{870} = \pm(0.050 + 0.010 \cdot AOD_{AERONET870}) \tag{2-7}$$

A logarithmic plot illustrating how the error bounds scale with increasing AOD is provided in Fig. 2-4.

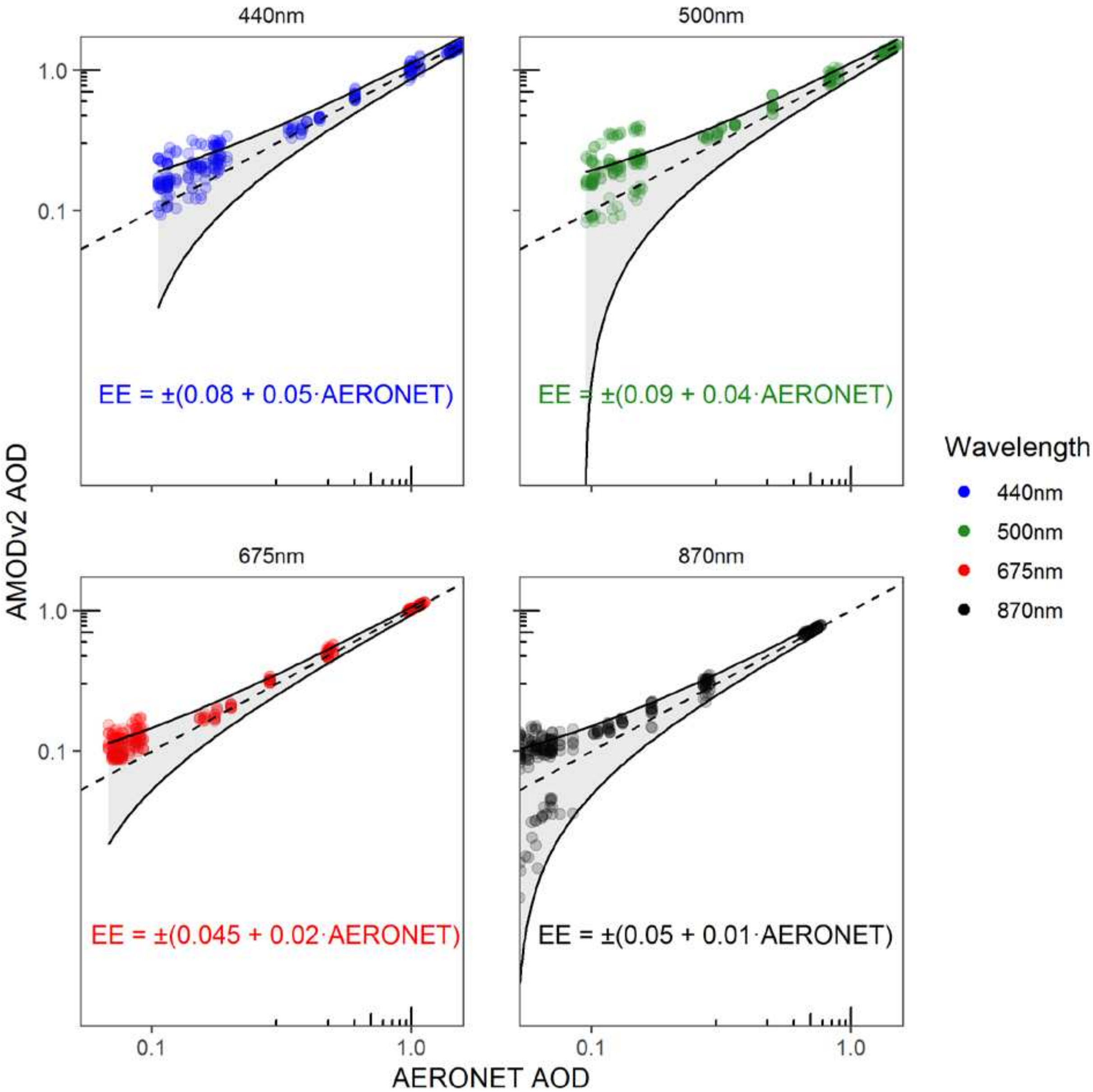

Figure 2-4: Logarithmic AERONET vs. AMODv2 AOD co-located results with expected error (EE; AOD units) bounds, with panels separated by wavelength. Equation bounds contain 85% of co-located measurements.

Equations (2-4) through (2-7) indicate a low dependence of the AOD magnitude on the AMODv2 error relative to AERONET for all wavelengths. Existing error between AMODv2 and AERONET measurements was explained primarily by the constant term. These findings are consistent with the summary statistics presented in Table A1 and demonstrate the stability of AMODv2.

AMODv2 bias relative to AERONET was primarily dependent on the specific unit, rather than systemic design uncertainty. A mean-difference plot colored by AMODv2 unit ID is provided in Fig. 2-5.

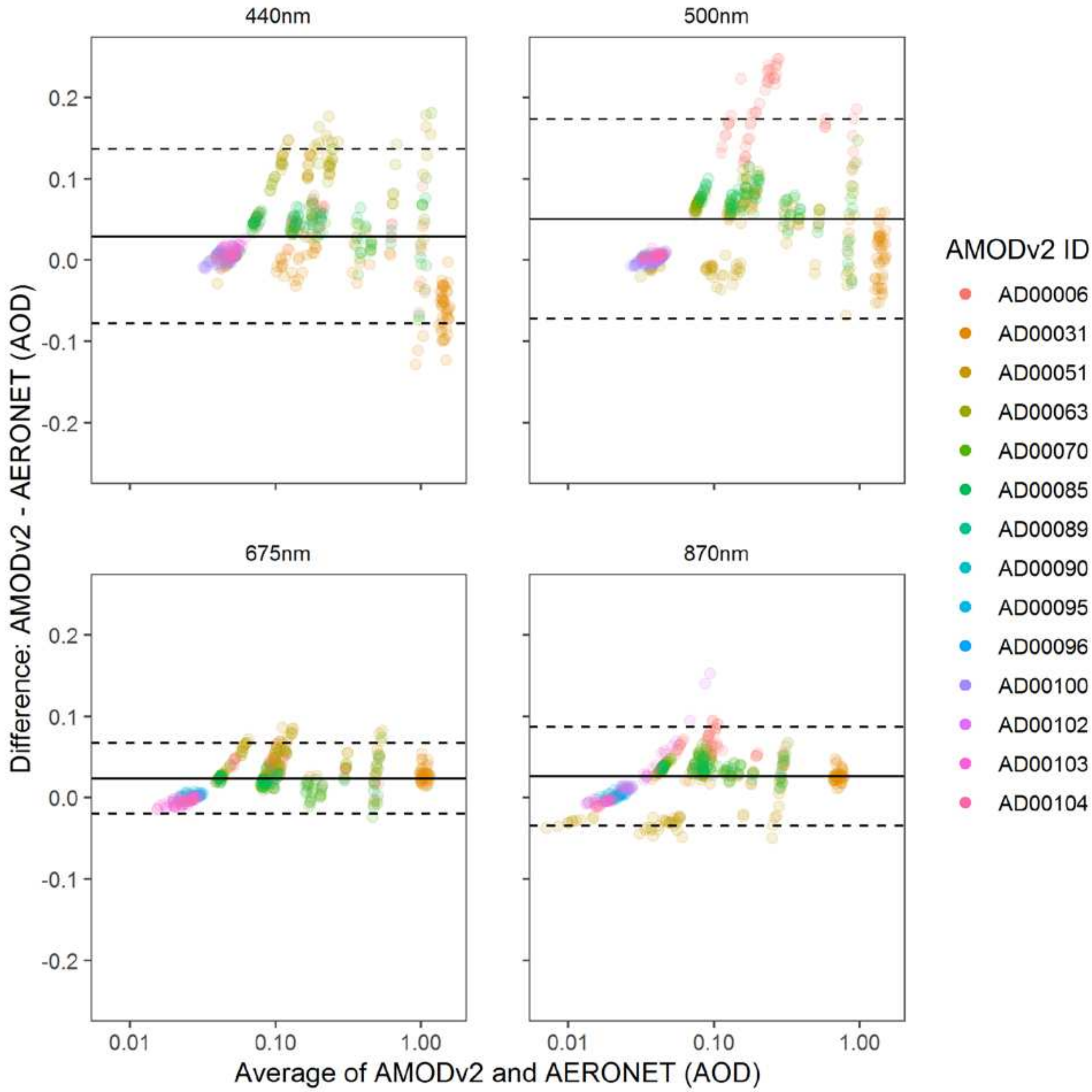

Figure 5: Mean-difference plot for measurements taken by AERONET and AMODv2 instruments, with panels separated by wavelength. Paired AERONET and AMODv2 under both clear and biomass burning conditions (as defined in Table A1) are included. Points represent paired AMODv2 and AERONET measurements with the average of the measurement pair on the x-axis in log scale and the difference on the y-axis. The top and bottom dashed lines represent the upper and lower limits of agreement, respectively, evaluated at 95% confidence. The solid line in between the limits of agreement is the mean difference between the two measurement techniques. Points are colored according to the AMODv2 unit ID.

Units AD00006 and AD00051 exhibited the highest bias at 440 nm and 500 nm, respectively. With units AD00051 and AD00006 removed from the data set, mean absolute errors were reduced by 0.011, 0.013, 0.008, and 0.004 AOD units at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. Bias from units AD00006 and AD00051 also impacted the EE derivations. With units AD00006 and AD00051 omitted, Eqs. (2-4) through (2-7) bound 92.5%, 94.6%, 97.6% and 92.2% of the co-located pairs, respectively. Individual unit bias was most likely caused by faulty calibration or optical sensor drift over time.

Previous work has noted the tendency for optical interference filters to degrade over time, changing the accuracy of the most recent calibration (Brooks & Mims, 2001a; Giles et al., 2019). We quantified the long-term stability of the AMODv2 AOD sensors by re-calibrating 16 AMODv2 units 15 months after their initial calibration. Summary statistics quantifying the change calibration constant (V_0) changes are provided in Table 2-3.

Table 2-3: Summary statistics for AMODv2 calibration stability test. All summary statistics refer to the change in V_0 (Eq. 2-2). Note that the absolute value of the maximum change refers to the single unit with the highest percent change for each wavelength.

Wavelength (nm)	Average absolute value of change (%)	Median change (%)	Absolute value of maximum change (%)
440	13.84	-7.14	62.72
500	11.80	-9.64	37.08
675	6.66	-0.75	29.40
870	14.63	-2.80	50.72

A plot illustrating the voltage change undergone by each of the 16 AMODv2 units is provided in Fig. 2-6.

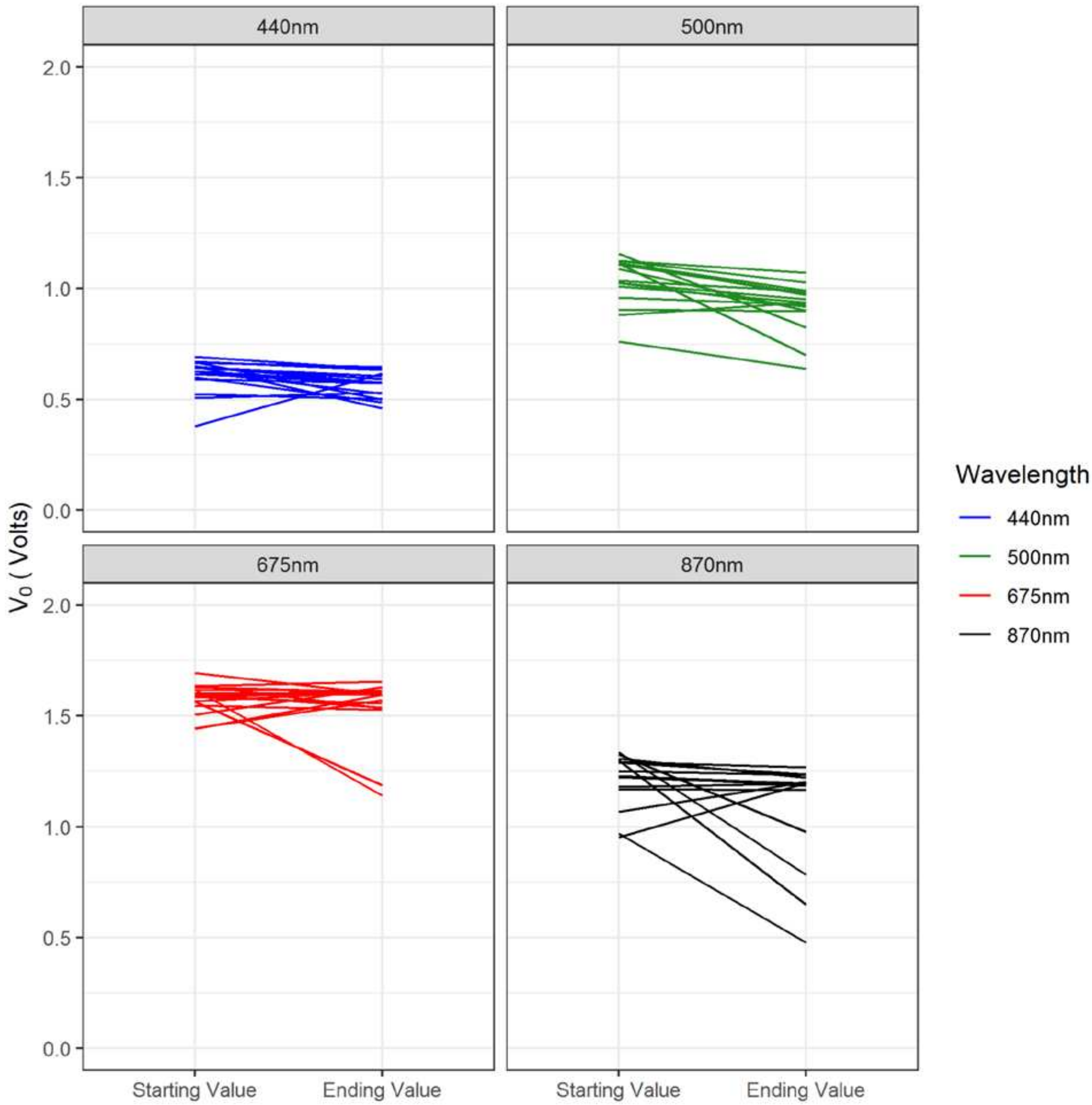

Figure 2-6: Linear change plots illustrating the change in calibration voltage, V_0 (Eq. 2-2), from the initial calibration to a follow up test calibration of 16 AMODv2 units. Each instrument is represented by a separate line with starting and ending calibration voltage values delineated on the vertical axis. Panels are separated by wavelength. Each line represents the change after 15 months of a single wavelength channel of an AMODv2 unit.

The results presented in Fig. 2-6 illustrate that the calibration constants (V_0 in Eq. 2-2) remained relatively stable (changes of 5% or less) for most AMODv2 units over the course of 15 months. However, several units exhibited relatively large changes (in excess of 30%) in their calibration constants, indicating calibration changes may vary considerably by unit. Boersma and de Vroom (2006) present theoretical analyses and conclude that the calculation of AOD is most sensitive to errors in the calibration constant, V_0 (Boersma & de Vroom, 2006). Their theoretical analyses combined with the results in Fig. 2-6, point to drift in V_0 as a likely source for large, unit specific errors in AOD AMODv2 measurements. To limit errors due to calibration drift, we recommend that AMODv2 V_0 values be re-calibrated on an annual basis. Determining the source of changes to the calibration constants of some AMODv2 units is the subject of ongoing investigation. Potential sources include changes in sensitivity or drift of the photodiode sensor element, degrading of the optical interference filters, and/or clouding of the protective glass window element in the light path of the sensors.

Reliability testing

AMODv2 sensor validation results from this work and prior work indicate that the instrument can accurately measure AOD and PM2.5 when operating properly. However, for effective large-scale deployments, AMODv2 units must reliably complete their intended sampling protocol when deployed outdoors for 120 hours. Potential causes of premature sample failure included, premature battery drainage, damage to mechanical or electrical components (e.g. water ingress into motors or sensors), and firmware related crashes (e.g. memory overflow errors). In a series of reliability tests on the rooftop of our laboratory facility, we found that of 76 attempted samples, 75% were successfully completed, 16% failed due to premature battery drainage, 8% failed due to water damage, and 1% (one unit) failed due to a firmware crash. To address failures due to premature battery drainage, we replaced batteries that would not fully charge and replaced motors that were drawing excess current. To address failures due to water damage, we replaced damaged boards and applied additional sealant to key mechanical interfaces. We addressed the firmware crash issue by reconfiguring the memory allocation to grant more memory to the wireless data push functionality, which proved to be the most memory intensive sub-system. Overheating was not an issue in the testing discussed here, as the testing was conducted in winter months. We will test the AMODv2 under warmer conditions to evaluate heating effects on the performance of the instrument.

We also verified that AMODv2 units were attempting AOD measurements and applying the prescribed data screening protocols. In the 76 test samples, AMODv2 units attempted 22,419 AOD measurements per wavelength. Units detected the sun and took at least one measurement toward forming a triplet 4,763 times per wavelength. The results partitioned by quality control designation are provided in Table 2-4. Instances where an AMODv2 reported a numerical AOD value were considered valid AOD measurements. Instances where an AMODv2 failed to acquire three AOD measurements for a single measurement sequence (Fig. A6) were designated as incomplete with a unique error code. Cloud-screened measurements were those where the solar alignment is achieved for 3 measurements but the triplet failed to meet the acceptance criteria (Fig. A6).

Table 2-4: Results from the AMODv2 quality control algorithm from 4.763 AOD measurements taken in laboratory rooftop testing. Attempts where zero measurements were logged for a triplet attempt are omitted from the table.

Wavelength (nm)	Proportion of valid AOD measurements	Proportion of invalid AOD measurements	
		Incomplete AOD triplets	Cloud-screened measurements
440	33%	20%	46%
500	34%	20%	45%
675	35%	20%	44%
870	33%	20%	46%

The results of this study indicate the AMODv2 automatically acquired solar alignment for a complete measurement triplet on 80% of attempted measurements. However, among the completed triplets, approximately 45% of measurements were identified as cloud-contaminated and subsequently screened. The screening algorithm did not reach consistent results across all wavelengths, as evident by slight deviations in the proportion of screened data across wavelengths. In this work, we applied the same exclusion criteria to each wavelength (Fig. A6). These results indicate unique exclusion criteria may be necessary for each wavelength to achieve consistent results, particularly when there is substantial deviation in magnitude between two measurement wavelengths (e.g. 440 nm AOD much higher than 870 nm AOD for a single measurement).

Discussion and conclusions

In the current study, we evaluated the AMODv2 under a wide range of atmospheric pollution levels and observed close agreement between the AMODv2 and AERONET AOD measurements, with mean absolute errors of 0.04, 0.06, 0.03, and 0.03 AOD units at 440 nm, 500 nm, 675 nm, and 870 nm, respectively. The agreement between AMODv2 and AERONET was stable across AOD levels ranging from 0.016 ± 0.01 to 1.590 ± 0.01. We identified unit-specific changes to AOD calibration constants over time as a potential source of error in AOD measurements and recommended annual re-calibration (in line with recommendations for AERONET instruments) to mitigate those errors. While the AMODv2 was designed to be deployed by citizens, here the evaluation was done with data collected by team members. In Parts 1 and 2, we noted that there could be potential user errors that may impact the data quality. These were not analyzed in the present study. Even though the AMODv2 was designed to reduce these errors by automating the AOD process, there is still the potential for errors (i.e., improper placement). Future work describing the deployment of AMODv2s by citizen scientists should also include analysis of these issues.

The AMOD was designed to be a low-cost, user-friendly, and high-performance instrument for $PM_{2.5}$ and AOD measurements to be deployed in citizen-science campaigns. Citizen-led sampling is a promising approach to produce large-scale data sets to quantify air pollution concentrations at spatiotemporal resolution unachievable by more-expensive reference monitors (e.g., Brooks and Mims, 2001; Boersma and de Vroom, 2006; Ford et al., 2019). In Parts 1 and 2 of this series, we detailed the design and deployment of the AMODv1. In these previous studies, we noted several limitations of the instrument design that limited the amount of data (specifically AOD) collected by participants. Here, we present the improvements made to the AMOD measurement system and the implementation of wireless data transfer and real-time visualization, which were the primary areas of improvement compared with the previous design. The new design of the AMODv2 allows for unsupervised measurement and quality control protocols that reduce the operational demands on a study volunteer, particularly compared with AMODv1 and other low-cost AOD sensors, while increasing the amount of data that can be collected. Deployments with citizen scientists are ongoing and data from those campaigns will be the subject of future studies. The portability, performance, and low cost of the AMODv2 make it a practical option to establish spatially-dense $PM_{2.5}$ and AOD measurement networks. Applied in these networks, the AMODv2 will close gaps in the existing global aerosol measurement infrastructure of ground-based and satellite-based observations.

CHAPTER 3. A NATIONAL CROWDSOURCED NETWORK OF LOW-COST FINE PARTICULATE MATTER AND AEROSOL OPTICAL DEPTH MONITORS: RESULTS FROM THE 2021 WILDFIRE SEASON IN THE UNITED STATES

Chapter Overview

Fine particulate matter ($PM_{2.5}$) air pollution is a leading cause of premature death, disease, and environmental degradation globally. Wildfire smoke is a primary source of ambient air pollution, particularly in the United States. In regions lacking adequate ground monitoring, health impact assessments and epidemiological studies on the effects of wildfire smoke may rely heavily on satellite-based instruments, which estimate ground-level $PM_{2.5}$ based on Aerosol Optical Depth (AOD), a measure of light extinction through the atmosphere. However, reference-grade ground monitors are cost prohibitive to deploy and maintain at the spatial scales needed to assess the spatial variability of wildfire smoke. Low-cost $PM_{2.5}$ sensors have been deployed at large scales and high spatial resolution in crowdsourced networks. However, these sensors typically lack AOD measurement capability. In prior work, we designed the Aerosol Mass and Optical Depth (AMODv2) sampler, which is capable of simultaneously measuring $PM_{2.5}$ (optical and gravimetric filter-based) and AOD. In this work, we distributed AMODv2s to student volunteers in the contiguous United States, forming a nationwide crowdsourced monitoring network from June 15, 2021 through August 8, 2021. A majority of our network were successful, with 86.6% of them resulting in a valid filter sample. We found our AOD sensors agreed closely with reference AOD monitors within 25 km on measurements within 180 seconds, with mean absolute error results in AOD units of 0.04 at 440 nm, 0.06 at 500 nm, 0.03 at 675 nm, and 0.03 at 870 nm, even with these imperfect co-location criteria. In a regional analysis of the effects of wildfire smoke on crowdsourced measurements, we observed elevated $PM_{2.5}$ and AOD on smoky days in most regions in the contiguous United States. These increases are manifested in similar $PM_{2.5}$:AOD ratio values in these regions regardless of the presence of smoke. However, in California, median $PM_{2.5}$ remained similar on smoky days relative to non-smoky days, while median AOD increased on smoky days, implying lofted smoke away from the surface. In California, the median $PM_{2.5}$:AOD ratio was 67.2 µg m^{-3} on non-smoky days, compared with 30.2 µg m^{-3} on smoky days. We show that paired $PM_{2.5}$ and AOD measurements collected by a crowdsourced network can highlight anomalies in ambient air quality during smoke events.

Introduction

Fine particulate matter ($PM_{2.5}$) air pollution is a leading cause of human death and disease globally (Brauer et al., 2016; Forouzanfar et al., 2016; Fuller et al., 2022; Vohra et al., 2021). $PM_{2.5}$ can penetrate deep into a person's lungs, causing acute and chronic disease (Feng et al., 2016; Janssen et al., 2013; Kim et al., 2019; Pope & Dockery, 2006). Each year, millions of premature deaths worldwide are attributed to $PM_{2.5}$ exposure (Brauer et al., 2016; Forouzanfar et al., 2016). Ambient $PM_{2.5}$ also impacts the Earth's climate by contributing to radiative climate forcing (Myhre et al., 2013).

As a primary source of ambient $PM_{2.5}$, smoke from wildfires has been linked to negative health outcomes (Liu et al., 2015; Reid et al., 2016; Cascio, 2018 and references therein). Worldwide, an estimated hundreds of thousands of deaths per year are attributable to ambient smoke (Johnston et al., 2012). In the United States, tens of thousands premature deaths are attributable to wildfire smoke per year (Ford et al., 2018; O'Dell et al., 2021). Wildfire smoke has been linked to respiratory, cardiovascular, and asthma-related morbidity (e.g., Henderson et al., 2011; Gan et al., 2017), resulting in thousands of hospital admissions per year (Fann et al., 2018; Ford et al., 2018; Neumann et al., 2021; O'Dell et al., 2021).

The impact of wildfire smoke plumes can vary considerably at relatively small spatial scales (e.g., Reid et al., 2015). In many areas affected by wildfire smoke, ground monitors are not distributed with sufficient spatial density for public health assessment, due to cost constraints. Satellite-based instruments, which can estimate ground-level $PM_{2.5}$ at relatively high spatial resolution, are used for assessment in areas lacking sufficient ground monitors (Hammer et al., 2020; Lee, 2020; van Donkelaar et al., 2016). Recent studies on the impact of wildfire smoke have incorporated data from satellite-based instruments in addition to surface monitors and simulation results from chemical transport models (CTMs) (e.g. Reid et al., 2015; Lassman et al., 2017; Ford et al., 2017; O'Dell et al., 2019; Cleland et al., 2020; Cheeseman et al., 2020).

Satellite-based instruments, such as the Moderate Resolution Imaging Spectrometer (MODIS) and the Multi-angle Imaging SpectroRadiometer (MISR) (Diner et al., 1998; Salomonson et al., 1989), have been used to estimate ground-level $PM_{2.5}$ by associating aerosol optical depth (AOD), a measure of light extinction through the atmosphere, with ground-level $PM_{2.5}$ (Hammer et al., 2020; Y. Liu et al., 2005; van Donkelaar et al., 2006, 2010, 2012, 2013, 2016). Often, these studies have translated AOD to $PM_{2.5}$ using a simple proportional relationship, expressed as follows (Y. Liu et al., 2005):

$$PM_{2.5} = \eta \cdot AOD \quad (1)$$

where η is an empirically-derived conversion factor between $PM_{2.5}$ and AOD. The uncertainty of satellite-based $PM_{2.5}$ estimates consists of the uncertainties of satellite-derived AOD and the conversion factor η (Ford & Heald, 2016; Jin et al., 2019).

Ground monitors can be used to constrain the uncertainties of satellite-based $PM_{2.5}$ estimates by accurately measuring $PM_{2.5}$ and AOD at the Earth's surface (Hammer et al., 2020; Sayer et al., 2012; van Donkelaar et al., 2012, 2013, 2016). Sun photometers measure AOD from the Earth's surface by quantifying the extinction of sunlight in the atmosphere due to atmospheric aerosols (Holben et al., 1998). The Aerosol Robotics Network (AERONET) consists of reference-quality sun photometers distributed throughout the planet, including hundreds of active sites in the contiguous United States (Holben et al., 1998). Ground-level $PM_{2.5}$ can be measured by a variety of different methods. For example, $PM_{2.5}$ can be measured gravimetrically by instruments that sample air at a known flow rate and isolate particles with diameters smaller than 2.5 µm from the flow stream, which are deposited on a filter (e.g. Volckens et al., 2017; Kelleher et al., 2018) or by using light-scattering sensors, measuring aerosol concentrations and size distributions based on how a controlled light source is scattered and absorbed by sampled air (e.g. Tryner et al., 2020). The Environmental Protection Agency (EPA) maintains the Air Quality System (AQS), which includes reference-quality gravimetric and light-scattering PM samplers (https://aqs.epa.gov/aqsweb/airdata.html). A complete evaluation of satellite-derived $PM_{2.5}$ requires co-locating ground-based AOD and $PM_{2.5}$ monitors. The Surface Particulate Matter Network (SPARTAN) features sites throughout the world where reference-quality AOD and $PM_{2.5}$ monitors are co-located and operate simultaneously (Snider et al., 2015). SPARTAN and deployments of similar instruments have been used to evaluate satellite AOD and $PM_{2.5}$ at specific sites (Green et al., 2009; Snider et al., 2015). Due to the relatively high costs of reference equipment (>10,000 USD for $PM_{2.5}$ and >50,000 USD for AERONET), there are relatively few active SPARTAN sites worldwide (~20 active sites). Outside of specific field campaigns (Garay et al., 2017; Holben et al., 2018; Sorek-Hamer et al., 2020), AERONET monitors are rarely deployed long-term at sub-city scales, which may be necessary to capture the $PM_{2.5}$ variability during wildfire smoke events (Reid et al., 2015). Lower-cost instruments have the potential to fill spatial gaps in ground monitoring networks left by reference monitors (Gupta et al., 2018).

Networks of low-cost nephelometers (e.g. the Plantower PMS5003) have been deployed to acquire ground-level $PM_{2.5}$ data at finer spatial resolution (Badura et al., 2020; Chadwick et al., 2021; Gupta et al., 2018; Li et al., 2020; Lin et al., 2020; Y. Lu et al., 2021). However, Plantower PMS5003 sensors are known to exhibit relatively high measurement bias, requiring field correction relative to reference methods (Barkjohn et al., 2021; Ford et al., 2019b; Kelly et al., 2017; Levy Zamora et al., 2019; Sayahi et al., 2019; Tryner et al., 2020; Zheng et al., 2018). Networks of low-cost, hand-held sun photometers have been deployed in crowdsourced studies (Boersma & de Vroom, 2006; Brooks & Mims, 2001b). However, these studies encountered difficulties coordinating with study participants to take measurements during satellite overpasses in addition to inconsistent measurement quality control practices (Boersma & de Vroom, 2006). To date, crowdsourced deployments of low-cost air pollution monitors have typically been limited to one measurement modality ($PM_{2.5}$ or AOD). This limitation has motivated our work toward the development of novel, low-cost monitors capable of measuring both $PM_{2.5}$ and AOD(Wendt et al., 2019b, 2021).

In prior studies, we developed and tested a low-cost monitor for simultaneous measurement of AOD and $PM_{2.5}$ called the Aerosol Mass and Optical Depth (AMOD) sampler (Ford et al., 2019b; Wendt et al., 2019b, 2021). We deployed AMODv1 samplers in a crowdsourced network called Citizen Enabled Aerosol Measurements for Satellites (CEAMS) in northern Colorado and found that non-scientist volunteers could effectively operate the instruments to acquire data used to assess satellite measurements at relatively small spatial scales (<5 km) (Ford et al., 2019b). In a subsequent study, we developed and validated the AMODv2 sampler. The AMODv2 maintained the $PM_{2.5}$ measurement capabilities of the AMODv1, while achieving longer runtimes through a larger battery, automated AOD measurement through a solar tracking system, and real-time data transmission through a Wi-Fi module (Wendt et al., 2021).

In this work, we describe our second CEAMS deployment featuring a crowdsourced $PM_{2.5}$ and AOD monitoring network of AMODv2s spanning the contiguous United States in the summer of 2021. We specifically focus our analysis on the regional variability of $PM_{2.5}$ and AOD under the influence of wildfire smoke. First, we summarize the data collected by volunteers using AMODv2s. Second, we assess the performance of our AOD sensors in a crowdsourced context by comparing our results to nearby AERONET monitors. Third, we evaluate the regional variability of $PM_{2.5}$ and AOD in the presence and absence of smoke. Finally, we highlight results from California, which exhibited unique smoke-dependent variability compared with the rest of the contiguous United States. With our crowdsourced network of low-cost $PM_{2.5}$ and AOD monitors, we seek to expand the availability of ground-level $PM_{2.5}$ measurements for direct application to Earth science and public health research, and to advance the understanding of the regional variability of $PM_{2.5}$ and AOD toward improving satellite-based air quality monitoring.

Materials and methods

Nationwide crowdsourced monitoring network

Our nationwide CEAMS network consisted of 29 undergraduate students and two administrators from the Student Airborne Research Program (SARP) through the National Aeronautics and Space Administration (NASA). Participants operated AMODv2s at their homes or workplaces. A map illustrating the distribution of sampling locations is provided in Fig. 3-1. We partitioned sites in the contiguous United States into four regions: 1) California/West Coast, 2) Mountain West, 3) Midwest, and 4) Northeast (Fig. 3-1). The sampling campaign began on June 15, 2021 and ended on August 8, 2021.

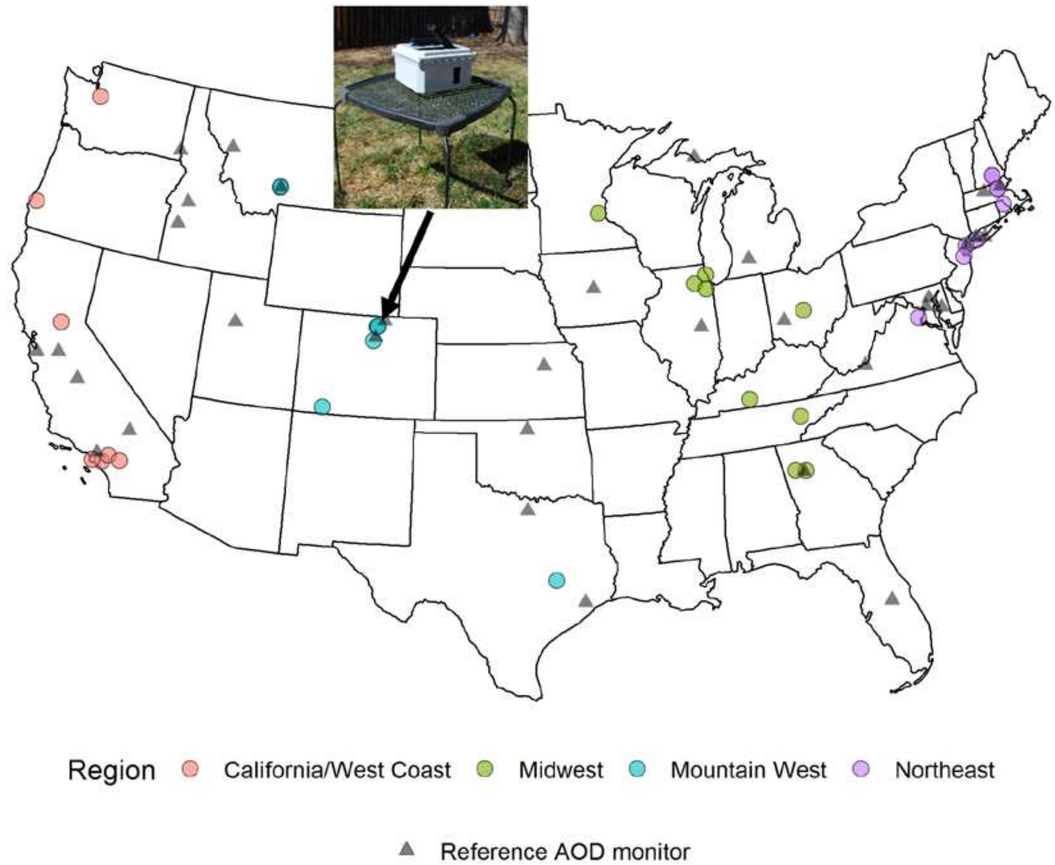

Figure 3-1: Locations of CEAMS and AERONET sites for summer 2021 field campaign. Sites are colored by region. An example photograph of an AMODv2 sampling AOD in Laporte, Colorado is provided (Photo Credit: Bonne Ford).

Participant training and AMODv2 operation

Here we summarize how participants operated the AMODv2 in our crowdsourced network. A detailed description of the instrument and prior validation work are provided in Wendt et al., (2021). The AMODv2 measures gravimetric time-integrated $PM_{2.5}$ mass concentration, real-time $PM_{2.5}$ mass concentration, and AOD simultaneously (Wendt et al., 2021). For gravimetric $PM_{2.5}$, the AMODv2 sampled air at 1 L min^{-1} through a custom cyclone (Kelleher et al., 2018; Volckens et al., 2017), which isolated $PM_{2.5}$ from incoming flow. $PM_{2.5}$ was continuously deposited on the filter until the conclusion of the sample, when the participant placed the cartridge in a sealed container and refrigerated their samples until the end of the campaign. AMODv2s also reported semi-continuous $PM_{2.5}$ concentrations using a light-scattering

$PM_{2.5}$ sensor (Plantower PMS5003, Beijing, China). AOD was measured at 440 nm, 500 nm, 675 nm, and 870 nm semi-continuously (during daylight hours) by AMODv2s using a solar tracking subsystem (Wendt et al., 2021). Instruments were programmed to report $PM_{2.5}$ and AOD at each 20-minute mark throughout the day (e.g. 1:00, 1:20, 1:40, etc.). At night or when the sun was obstructed from view, only $PM_{2.5}$ was reported. At the conclusion of each sampling period, participants brought their AMODv2 inside for charging and to initiate a wireless full-file data transfer via the smartphone application. The full log files contained meteorological (e.g temperature, relative humidity, and barometric pressure) and quality control (e.g. battery state of charge and wireless signal strength) data (Wendt et al., 2021).

We distributed AMODv2 sampling kits to study participants via mail. Each sampling kit included an AMODv2, eight pre-weighed gravimetric sampling filters, two pre-weighed blank filters, a charging cable, a user manual (Wendt et al., 2021), a microSD card reader, and a sampler repair kit. The external housing of the AMODv2 is made from a weather-resistant NEMA electrical enclosure (Polycase, YQ-080804, Avon, Ohio, USA). We trained participants on how to operate their AMODv2s in a series of online video training sessions. We coordinated participants to begin their 96-hour samples each week on Tuesday morning and sample through Saturday morning. We held online "office hours" on Mondays to facilitate repairs of damaged samplers and provide further instruction for interested participants. We assigned each participant with a unique identification code to link their instrument with our project website (csu-ceams.com). Participants with internet access connected their instruments to their site's Wi-Fi using our smartphone application ("CEAMS"; available on the Apple App Store and Google Play). They could then visualize their data on our website in real time. Those lacking internet access removed their sampler's microSD card at the end of each sampling period to download data, which was subsequently uploaded to our server via FTP. To prepare a sample, participants loaded a pre-weighed, 37mm Teflon filter contained in a cartridge (Wendt et al., 2021). Participants then placed their AMODv2 in an outdoor location with, to the greatest extent possible, an unobstructed view of the sun during the day. Participants then initiated a 96-hour sampling run using our smartphone application.

$PM_{2.5}$ and AOD data processing and quality control

At the conclusion of the nationwide CEAMS campaign, participants mailed back their AMODv2s along with their filter samples. We first post-weighed sample filters and blanks. We evaluated the mass limit-of-detection as the mean change in filter blank mass plus three times the standard deviation of the blank filter masses. We discarded AMODv2 log files based on the validity of the accompanying filter samples. Filter samples with less mass than the limit-of-detection were not included in subsequent analyses. We applied blank correction to filter masses by subtracting the mean blank mass from the mass of each sample filter from a given participant. Following Ford et al., (2019) and Tryner et al., (2020), we applied a filter correction to each PMS5003 $PM_{2.5}$ measurement. We scaled each PMS5003 $PM_{2.5}$ measurement such that the sample period average was equal to the time-weighted average $PM_{2.5}$ concentration measured via the filter sample.

AOD measurements made in our study followed a triplet protocol. At each 20 minute mark, the AMODv2 AOD subsystem sampled sunlight intensity three times at 30-second intervals. Both the raw measurements and the triplet average were saved in the sample log file. We screened our triplet measurements for cloud contamination based on the AERONET cloud screening algorithm, which, in part, identifies measurements with intra-triplet variation above an empirically determined threshold (Giles et al., 2019; Smirnov et al., 2000). To account for differences in the electrical stability of our optical sensors compared with those in AERONET monitors, we derived our own empirical intra-triplet variability thresholds. Specifically, at AOD below 0.7, we classified triplets with maximum variability greater than 0.1 to be cloud contaminated. At AOD greater than or equal to 0.7, the threshold was set to 0.15. We also omitted incomplete triplets. For analysis of $PM_{2.5}$:AOD ratio, we used the 500 nm AOD channel.

The gravimetric $PM_{2.5}$ subsystem of the AMODv2 have been extensively evaluated in laboratory and field settings in prior work (Arku et al., 2018; Kelleher et al., 2018; Pillarisetti et al., 2019; Volckens et al., 2017; Wendt et al., 2019b). The light-scattering $PM_{2.5}$ subsystem has likewise underwent extensive laboratory and field evaluations (Kelly et al., 2017; Zheng et al., 2018; Levy Zamora et al., 2019; Sayahi et al., 2019; Wendt et al., 2019a; Bulot et al., 2019; Tryner et al., 2020). By contrast, prior to this work, the AOD subsystem had been evaluated only in controlled experiments (Wendt et al., 2021) and not in a crowdsourced field campaign. To evaluate the performance of the AOD subsystem in our crowdsourced network, we compared valid AMODv2 AOD measurements with AERONET AOD measurements taken within 180 seconds and 25 km of each other. A total of four CEAMS/AERONET sites satisfied these criteria yielding 493 valid measurement pairs in total including all wavelengths.

We obtained daily smoke plume information from National Oceanic and Atmospheric Administration (NOAA) Hazardous Mapping System (HMS) polygons (Ruminski et al., 2006). The HMS product identifies that there is smoke in the atmospheric column, without specifying where in the column the smoke is located (Ruminski et al., 2006). CEAMS $PM_{2.5}$ and AOD measurements occurring within the boundaries of daily smoke polygons were labeled as smoke-impacted.

Results and discussion

Data overview

Participants in our nationwide CEAMS network collected a total of 192 unique log files using AMODv2 samplers in the summer of 2021. These log files included meteorological and quality control data at 30-second intervals, along with $PM_{2.5}$ and AOD measurements at 20-minute intervals. They also include summary runtime metrics such as total sampling time, total volume of air sampled, and the reason for runtime termination. Sampling runs concluded at the end of the specified 96-hour runtime or prematurely due to one of the following mechanisms: depleted battery, user-initiated manual shutdown, or unknown error. Based on prior testing (Wendt et al., 2021), we suspect failures with unknown cause were most likely due to water damage from either heavy rain or home irrigation systems. For a detailed analysis of prematurely terminated sampling runs, see Text S1 in the supplementary material. Of the 192 sample log files, 189 had an accompanying filter returned at the end of the study. Others were lost or damaged in the unloading process or in transit. We received a total of 60 blank filters at the conclusion of the study. The limit-of-detection for filter samples was 17.3 μg, which translates to a time-averaged $PM_{2.5}$ mass concentration of 3 μg m^{-3}, based on a 96-hour sample duration at 1 L min^{-1}. Of the returned filters 149 (78.8%) contained $PM_{2.5}$ mass in excess of the limit-of-detection. Of the 43,071 total PMS5003 $PM_{2.5}$ measurements collected, 41,381 (96.2%) were non-negative (post-sampling weight exceeded pre-sampling weight) and classified as valid. The mean ratio of the filter $PM_{2.5}$ to Plantower PMS5003 $PM_{2.5}$ (i.e., PMS5003 scaling factor) was 1.7 with a standard deviation of 1.9. Box and whisker plots of these PMS5003 scaling factors as a function of the number of smoke-impacted days is provided in Fig. B1, and these correction-factor distributions for each individual AMODv2 device are provided in Fig. B2. Log files without a corresponding filter measurement were excluded from all analyses, leaving a total of 38,699 filter-corrected PMS5003 measurements. For AOD measurements, there were 3,658; 3,760; 3,681; and 3,083 triplets marked as valid for 440 nm, 500 nm, 675 nm, and 870 nm channels; respectively. For additional information on AOD quality-control results, see Text S1 in the supplementary material. In Fig. 3-2, we provide the distribution of sample runtimes associated with CEAMS log files labeled by failure mode and filter status.

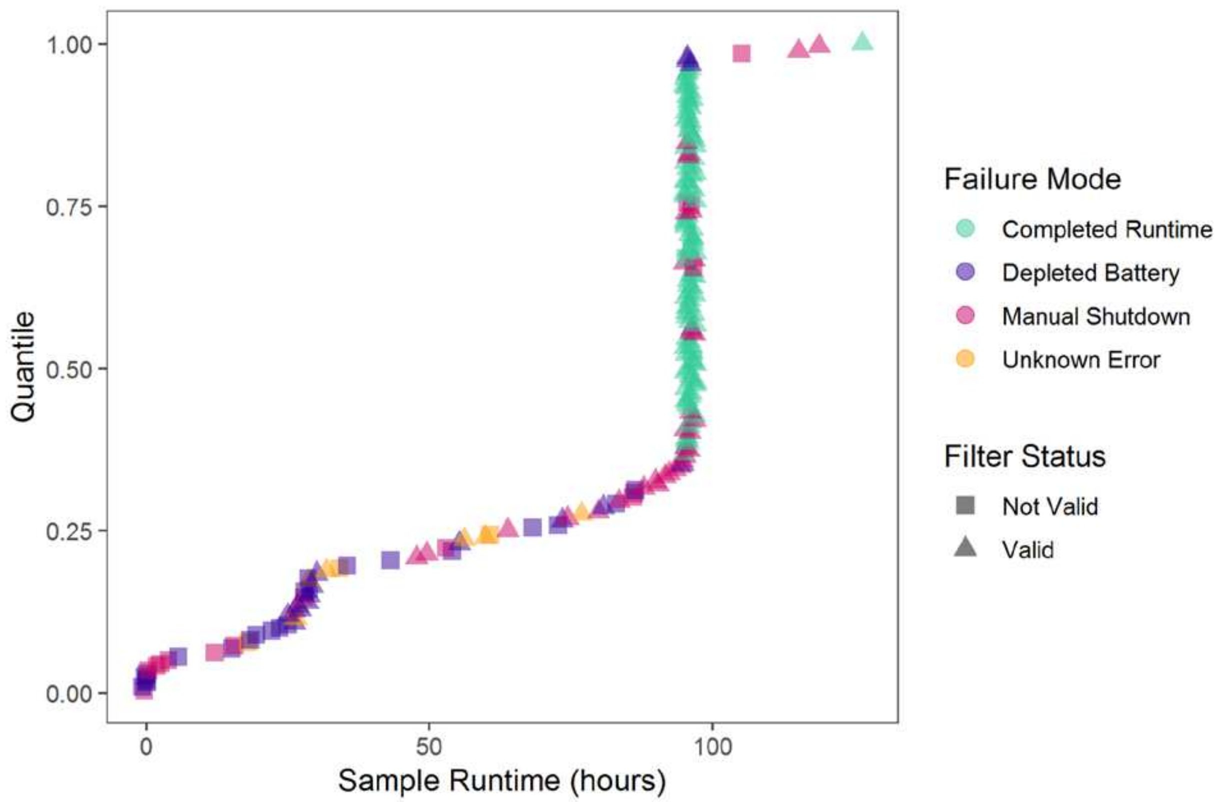

Figure 3-2: Distribution of CEAMS sample logs ordered by runtime with failure mode and filter status labeled. We introduced a horizontal random jitter with one hour width for visualization purposes. Points labeled "Completed Runtime" completed the full 96-hour runtime without interruption. Sample runtimes above 96 hours were caused by participants shutting off their AMODv2 mid-sample and restarting it shortly thereafter. Such samples with valid filters were included in subsequent analyses.

In Fig. 3-3, we provide histograms illustrating the distributions of filter-corrected PMS5003 $PM_{2.5}$, AOD at 500 nm, and $PM_{2.5}$:AOD ratios for all data from the study. In Fig. B3, we provide histograms for the remaining three AOD wavelengths.

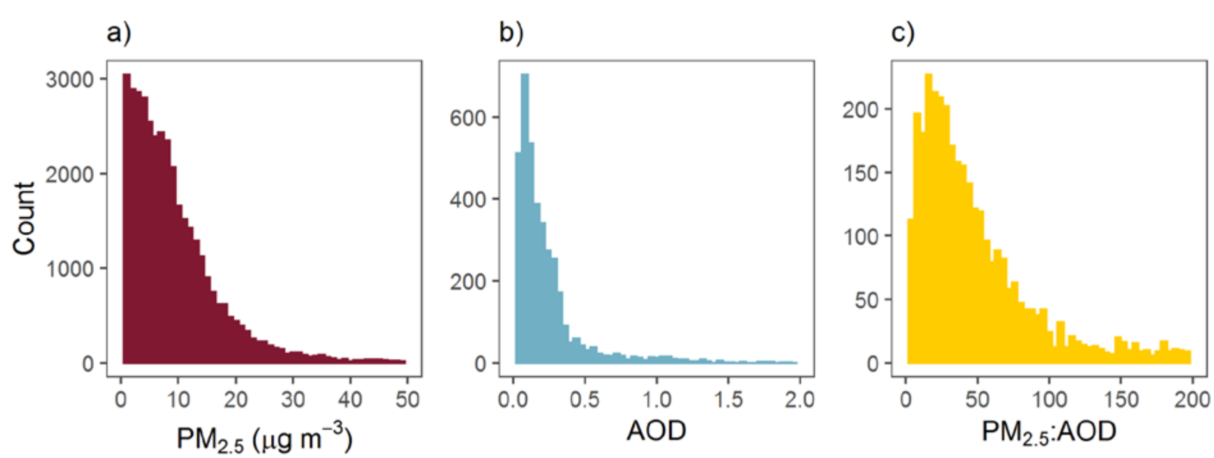

Figure 3-3: Histograms of a) $PM_{2.5}$, b) AOD at 500 nm, and c) $PM_{2.5}$:AOD ratio for all valid measurements across all CEAMS sites for June -August 2021 deployment. Counts represent measurements collected at 20-minute intervals.

Across all CEAMS measurements, the mean (median) values of PM2.5 concentration, AOD at 500 nm, and PM2.5:AOD ratio were 9.6 (7.2) µg m-3, 0.24 (0.12), and 60.1 (37.9) µg m-3, respectively. These average results are of similar order of magnitude to those in preliminary SPARTAN sites in North America (Snider et al., 2015).

In-field AERONET comparison

Here we analyze the bias and accuracy of crowdsourced AOD data from the relatively low-cost CEAMS instruments compared to reference AERONET monitors. In Fig. 3-4, we provide a scatter plot of CEAMS and AERONET AOD measurements by wavelength and HMS smoke designation. Paired measurements in Fig. 3-4 were located within 25 km and coincident within 180 seconds.

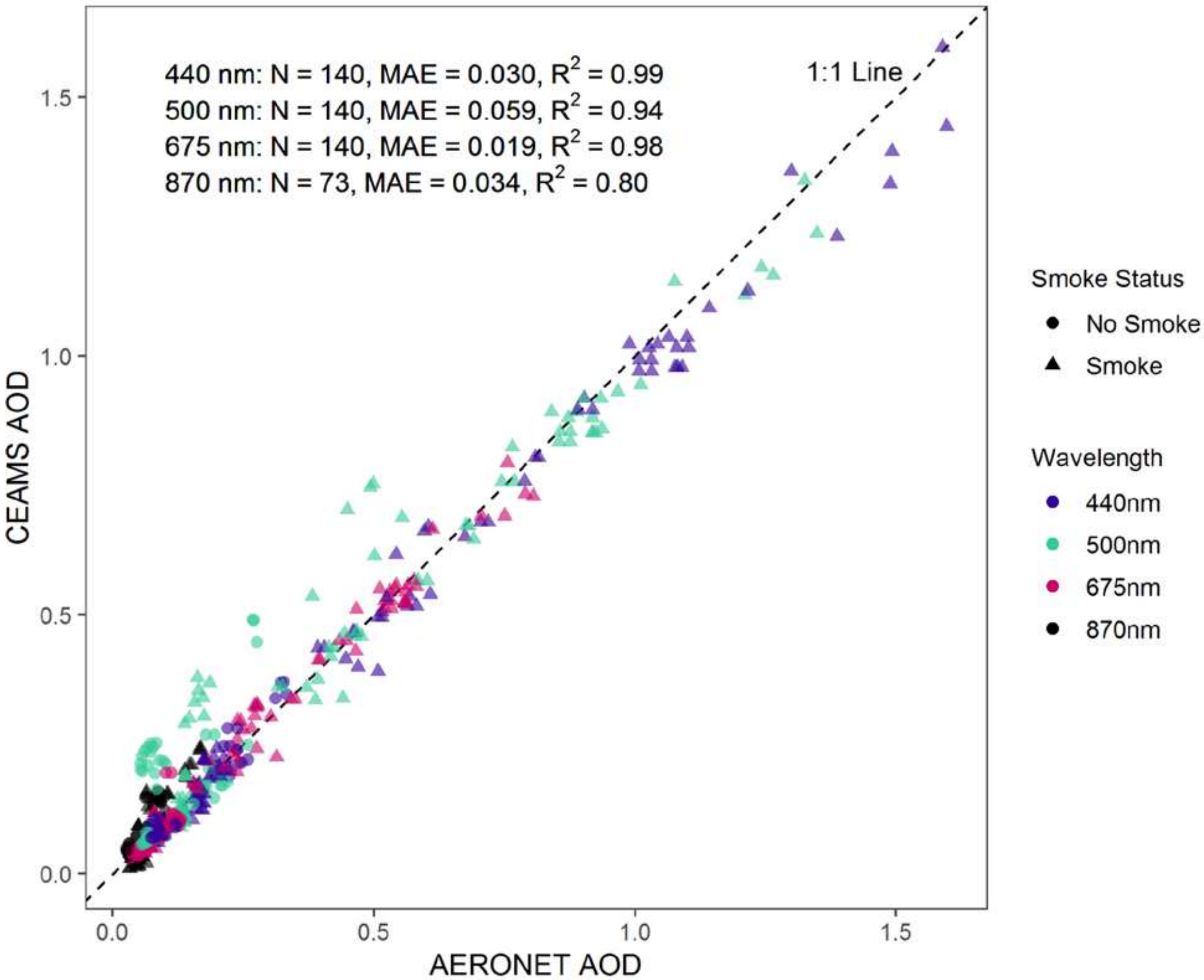

Figure 3-4: Scatter plot of CEAMS and AERONET AOD measurement pairs occurring within 180 seconds and 25 km of each other. N = number of paired triplicate measurement points; MAE = mean absolute error in dimensionless AOD units.

AERONET AOD values ranged from 0.057 to 1.60 at 440 nm, 0.053 to 1.35 at 500 nm, 0.034 to 0.81 at 675 nm, and 0.026 to 0.17 at 870 nm. The mean absolute errors in AOD units were 0.030 at 440 nm, 0.059 at 500 nm, 0.019 at 675 nm, and 0.034 at 870 nm. These results are consistent with our prior study of AMODv2 accuracy (Wendt et al., 2021), in which the mean absolute errors in AOD units were 0.04 at 440 nm, 0.06 at 500 nm, 0.03 at 675 nm, and 0.03 at 870 nm. These results indicate that crowdsourced AMODv2 measurements can achieve similar agreement with reference AERONET measurements across a broad range of AOD values.

The agreement between CEAMS AOD and AERONET AOD varied as the distance between two monitors increased. In Fig. B3, we illustrate how the mean absolute percent error varies inconsistently as the distance between the CEAMS and AERONET instruments increases. We observe relatively close agreement for measurements within 25 km of each other. Between 25 km and 37 km, the mean absolute percent error is elevated and then falls again between 37 km and 53 km (Fig. B4). Paired AMODv2 and AERONET AOD measurements from locations between 25 km and 37 km away from the nearest AERONET monitor were impacted by smoke, which may have large gradients in AOD over these lengthscales. The relatively close agreement between AMODv2s and AERONET monitors separated by between 37 km and 53 km, of which measurements were primarily not smoke-impacted, indicates the presence of smoke can have a substantial impact on the agreement of AMODv2 and AERONET monitors within 53 km. For AMODv2 and AERONET monitors separated by more than 53 km, the mean absolute percent error was relatively high for all coincident measurements (Fig. B1).

Regional variability of $PM_{2.5}$ and AOD due to smoke

To assess the relationship between $PM_{2.5}$ and AOD, we isolated measurements with valid filter-corrected $PM_{2.5}$ and AOD at 500 nm, yielding a total of 3,391 paired measurement data points. Each measurement included a region designation along with a binary HMS smoke designation (yes/no smoke). There were 1,217 (655); 88 (194); 490 (412); and 158 (177) measurements with no smoke (yes smoke) according to HMS in California, the Midwest, the Mountain West, and the Northeast respectively. In Fig. 3-5, we present box and whisker plots of $PM_{2.5}$, AOD, and $PM_{2.5}$:AOD ratio as functions of region and HMS smoke designation. We present summary statistics for CEAMS based on region and HMS smoke designation in Table 3-1.

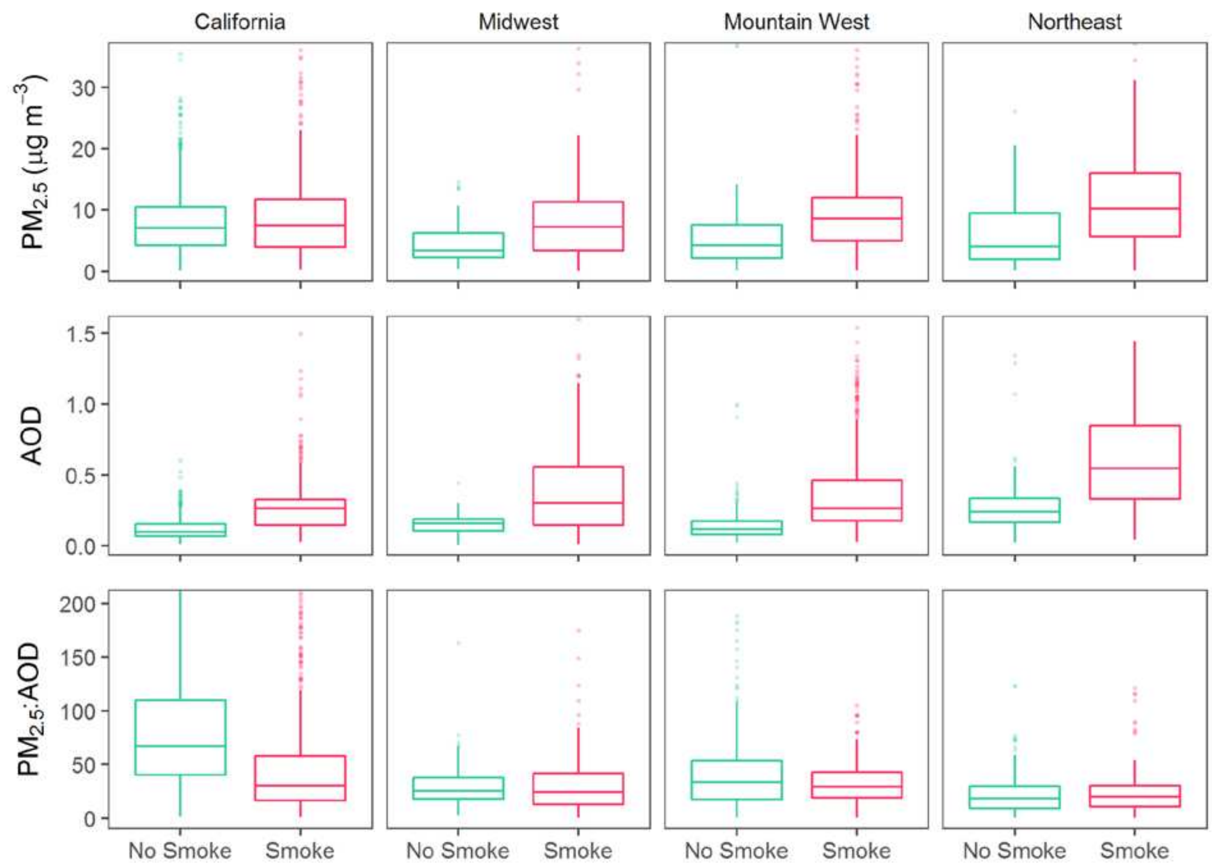

Figure 3-5: Box and whisker plots of $PM_{2.5}$, AOD, and $PM_{2.5}$:AOD ratio separated by region and HMS smoke status. The box represents the 25th-75th percentile range with the line within the box denoting the median. Note that here we refer to the "California/West Coast" region as "California" because sites in Washington and Oregon did not collect valid AOD measurements.

Table 3-1: Regional $PM_{2.5}$, AOD, and $PM_{2.5}$:AOD ratio quantiles separated by HMS smoke status.

Region	HMS Smoke Designation	$PM_{2.5}$: Median (25%, 75%) ($\mu g\ m^{-3}$)	AOD: Median (25%, 75%) ($\mu g\ m^{-3}$)	$PM_{2.5}$:AOD ratio: Median (25%, 75%) ($\mu g\ m^{-3}$)
California	Not smoky	7.1 (4.2, 10.5)	0.098 (0.069, 0.15)	67.2 (40.2, 109.6)
California	Smoky	7.5 (4.0, 11.7)	0.26 (0.15, 0.33)	30.7 (16.4, 57.9)
Midwest	Not smoky	3.47 (2.2, 6.3)	0.16 (0.11, 0.19)	25.4 (18.0, 37.7)
Midwest	Smoky	7.21 (3.4, 11.3)	0.30 (0.15, 0.56)	24.2 (12.8, 42.1)

Mountain West	Not smoky	4.23 (2.2, 7.6)	0.11 (0.079, 0.17)	33.50 (17.6, 53.8)
Mountain West	Smoky	8.55 (4.9, 12.0)	0.26 (0.18, 0.47)	29.4 (19.1, 42.7)
Northeast	Not smoky	4.1 (2.0, 9.5)	0.24 (0.17, 0.33)	18.4 (9.3, 29.9)
Northeast	Smoky	10.2 (5.6, 16.0)	0.55 (0.33, 0.85)	20.3 (10.7, 30.8)

In the Midwest, Mountain West, and Northeast regions; median $PM_{2.5}$ and AOD values were both higher on smoky days compared with non-smoky during the summer of 2021 (Fig. 3-5). These simultaneous increases from smoke manifested in relatively stable $PM_{2.5}$:AOD ratio values in those three regions regardless of the presence of smoke (Fig. 3-5 and Table 3-1). For sites in California, median AOD was elevated on days with smoke present compared with days without smoke (Fig. 3-5). The magnitude of the median increase was similar to what we observed in the Midwest and Mountain West regions (Table 3-1). However, in California, $PM_{2.5}$ concentrations were relatively unaffected by the presence of smoke, with the median concentration only 0.41 $\mu g\ m^{-3}$ higher on smoky days compared to non-smoky days. Higher AOD with unaffected $PM_{2.5}$ concentrations on smoky days in California manifested in median $PM_{2.5}$:AOD ratio being higher on non-smoky days compared to smoky days.

In a decadal climatology (2008 to 2017) of the Pacific West, South West, and Southeast regions of the United States, Bian et al., (2020) show that median $PM_{2.5}$ concentrations were higher in the presence of wildfire smoke in all regions. Our results in the Midwest, Mountain West, and Northeast regions of the United States are consistent with this trend. However, in California, our results differ from climatological expectations for the Pacific West (Bian et al., 2020), with median $PM_{2.5}$ concentrations remaining stable across smoky and non-smoky conditions. We found these anomalous results were driven primarily by four CEAMS locations in southern California (Fig. 3-1). In Fig. B5, we present maps of fire locations and smoke plumes from the HMS product illustrating evolving smoke plumes in the California region for selected transition periods between smoky and non-smoky conditions. In Fig. 3-6, we present time series data from the four devices in Southern California between July 5, 2021 and July 25, 2021, featuring transient wildfire smoke. A version of Fig. 3-6 including the 440 nm to 675 nm Angstrom Exponent is provided in Fig. B6. A version of Fig. 3-6 for the CEAMS location in Sacramento, California is provided in Fig. B7.

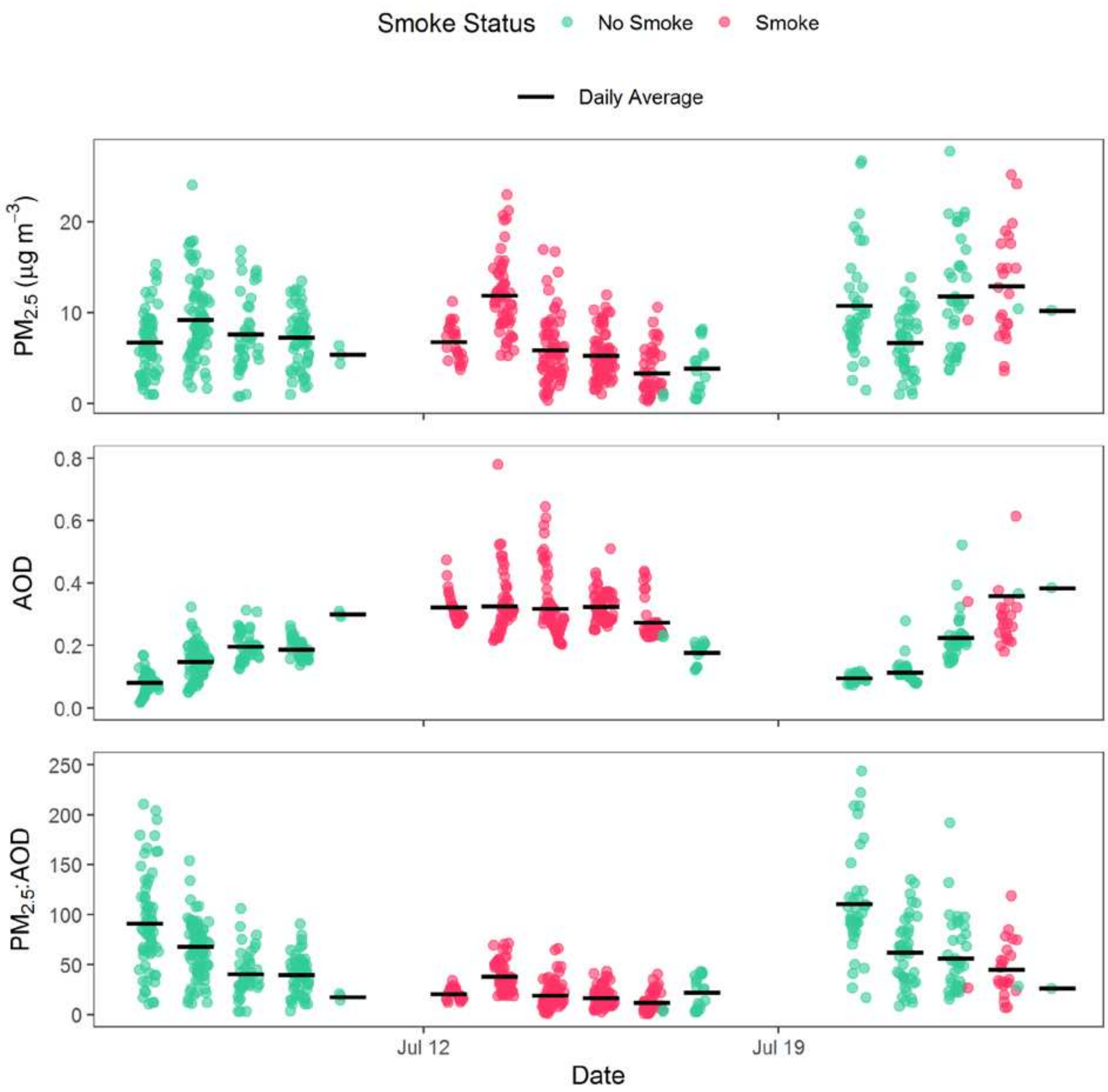

Figure 3-6: Time series plots of $PM_{2.5}$, AOD, and $PM_{2.5}$:AOD ratio from July 6, 2021 to July 24, 2021 from four CEAMS sites in Southern California, colored by HMS-derived smoke status. Measurements are from four AMODv2s (Fig. 3-1). Note that measurements on July 12, 2022 were from a single AMODv2 started early by the CEAMS participant.

In the sampling periods of July 6 through July 10 and July 12 through 17, we observed similar week-to-week $PM_{2.5}$ concentrations, despite the presence of wildfire smoke from July 12 through July 16

(Fig. 3-6). During the same sampling periods, the magnitude of AOD was affected by wildfire smoke, with AOD values increasing substantially during the smoky week relative to the non-smoky week (Fig. 3-6). In the sampling period of July 20 through July 24, smoke impacted measurements were observed on July 23. Between July 20 and July 24 $PM_{2.5}$ concentrations in Southern California trended upward, without significantly deviating from the previous two weeks in magnitude. AOD reached its peak average magnitude on July 23 in the presence of wildfire smoke (Fig. 3-6). The site near Sacramento (not included in southern California category) exhibited similar behavior during the first two week (Fig S7). However, on July 23 and 24, the site near Sacramento exhibited a prominent increase in both $PM_{2.5}$ concentration and AOD (Fig S7).

Prior studies using low-cost sensors in California during wildfire events have reported distinct increases in ground-level $PM_{2.5}$ concentrations during wildfire smoke events (Gupta et al., 2018; Y. Lu et al., 2021). Using satellite-based AOD measurements, Gupta et al., (2018) observed a similarly distinct increase in AOD during the same wildfire smoke events. We observed similar results near Sacramento on July 23 and 24 (Fig. B7). However, in Southern California only AOD was consistently higher on smoky days, with $PM_{2.5}$ concentrations often remaining similar to levels on non-smoky days (Fig 6). These results are consistent with smoke aloft during some wildfire smoke events specified by HMS. With smoke aloft, the AOD sensors would detect additional light diminution caused by suspended particles, where surface nephelometers would be minimally affected. In Fig. 3-7, we highlight potential smoke aloft in California by providing regional histograms for $PM_{2.5}$:AOD ratios as a function of HMS smoke status.

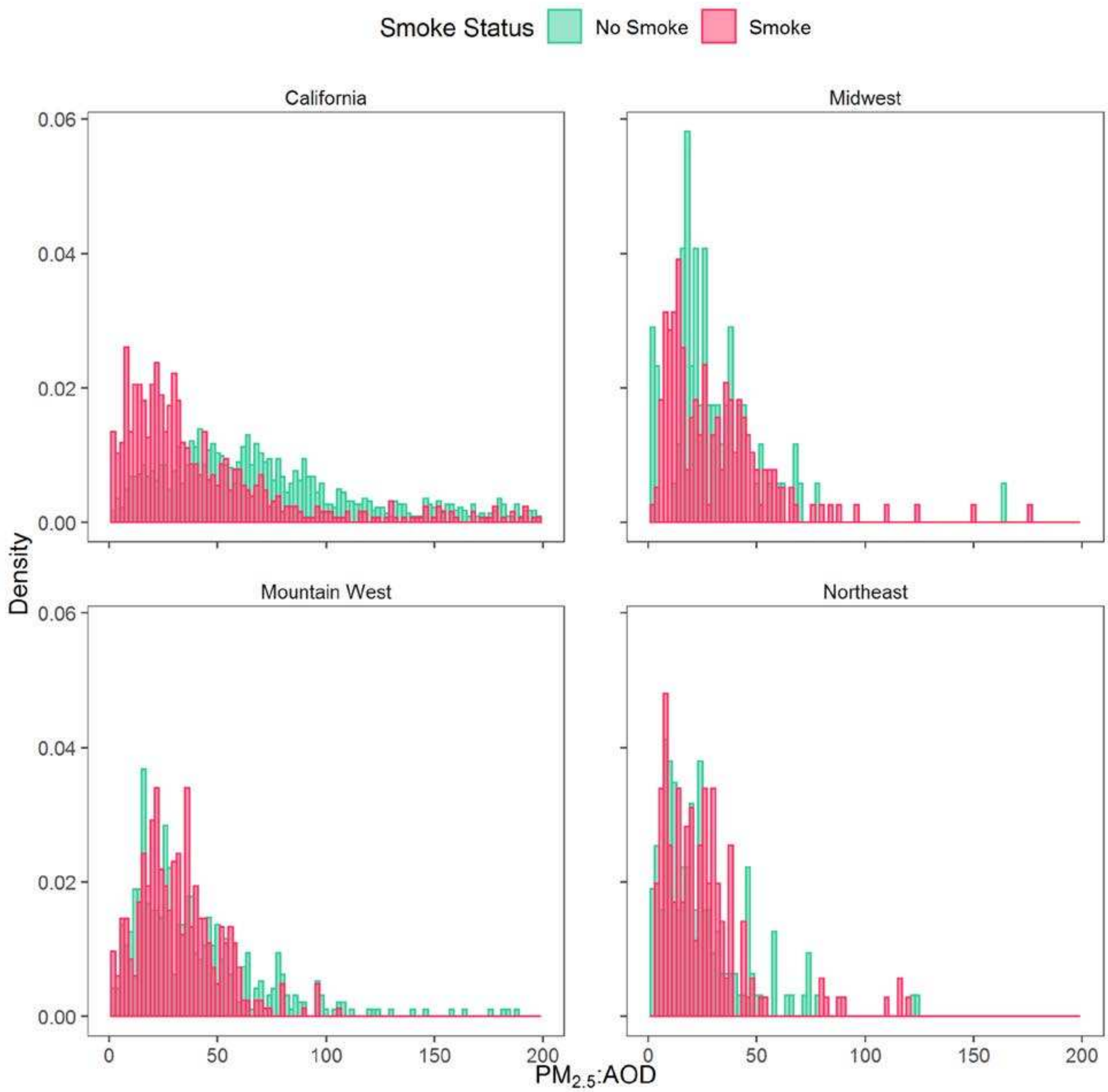

Figure 3-7: Regional density histograms for $PM_{2.5}$:AOD ratio separated by HMS smoke status. Note that here we refer to the "California/West Coast" region as "California" because sites in Washington and Oregon did not collect valid AOD measurements.

In the Midwest, Mountain West, and Northeast, distributions of $PM_{2.5}$:AOD ratios were relatively consistent from non-smoky to smoky days (Fig. 3-7). In California, the distribution of $PM_{2.5}$:AOD ratios on non-smoky days differed from other regions in our study. There were more instances of $PM_{2.5}$:AOD ratios above 50 μg m^{-3} on non-smoky days in California than for all other regions and smoke statuses (Fig. 3-7). If the regional trend of similarity between smoky and non-smoky days held true in California, we would expect the distribution of $PM_{2.5}$:AOD ratios on smoky days to skew high relative to other regions. However, we observed $PM_{2.5}$:AOD ratios in California skewed relatively low on smoky days (Fig. 3-7). Like daily results presented in Fig. 3-6, these results are consistent with smoke aloft events, where surface $PM_{2.5}$ are relatively unaffected compared with AOD in the presence of elevated plumees. The presence of smoke aloft may introduce a positive bias to AOD-based $PM_{2.5}$ quantification methods, including satellite-based instruments, if the vertical distribution of smoke is not accounted for (Cheeseman et al., 2020; Ford & Heald, 2016). A majority of wildfire smoke impacting California is produced by local sources, creating wildfire events that can evolve at relatively small spatial and temporal scales (Brey et al., 2018). Our work highlights the utility of deploying co-located $PM_{2.5}$ and AOD monitors during wildfire events.

Limitations

The primary limitation of this study was the relatively small number of unique sites in our national crowdsourced network. While we were able to acquire substantially more $PM_{2.5}$ and AOD measurements than in our previous local CEAMS deployment (Ford et al., 2019b), our network was still not near the scale of low-cost networks of nephelometers (Badura et al., 2020; Chadwick et al., 2021; Gupta et al., 2018; Li et al., 2020; Lin et al., 2020; Y. Lu et al., 2021). To better realize the advantages of low-cost monitors and crowdsourcing future networks similar to CEAMS should feature more sites in a given locality than EPA-AQS and AERONET.

In addition to a limited number of deployment sites, our results were also limited by AMODv2 failures due to damaged components and inconsistent operation. Eight of 31 sites did not contribute AOD measurements and were thus not included in the analysis depicted in Fig. 3-5. These instruments sustained damage to the AOD subsystem either in the shipping process or early on in the deployment. We instructed participants with malfunctioning AOD subsystems to disconnect the AOD sensors on their AMODv2 and complete the remainder of the deployment collecting filter and light-scattering $PM_{2.5}$ measurements. The crowdsourcing aspect of this deployment also introduced unique difficulties, with 28.2% of premature shutdowns being caused by operator intervention. For example, two AMODv2s were stolen from participants' yards, prompting other participants to prematurely end some samples when they felt their AMODv2 was not in a secure location. Other reasons for early manual shutdowns included transportation of instruments mid-sample, uncertainty about inclement weather, and uncertainty about potentially malfunctioning components. These incomplete sampling attempts were discarded in the final analysis if they did not meet limit-of-detection thresholds for the filter samples. Irregularities in sampler operation have impacted prior studies involving crowdsourced data collection (e.g. Boersma and de Vroom, 2006; Ford et al., 2019). In this study, we found that the increased degree of automation of the AMODv2 relative to the AMODv1 led to more reliable results from our participants (Ford et al., 2019b). In future studies, we will continue to improve the usability and reliability of our sensors to reduce potential complications for participants without a background in aerosol sampling.

Conclusions

Recent studies have leveraged crowdsourced networks of low-cost sensors toward a greater understanding of ambient air quality (e.g. Boersma and de Vroom, 2006; Gupta et al., 2018; Ford et al., 2019; Lu et al., 2021). In this study, we build upon these efforts by establishing a nationwide crowdsourced network of integrated $PM_{2.5}$ and AOD monitors. AMODv2s performed similarly in our crowdsourced field campaign compared with laboratory validation experiments with respect to runtime reliability and AOD quality control (Wendt et al., 2021). AMODv2 AOD measurements in the CEAMS network agreed closely with AERONET monitors for measurements co-located within 25 km and coincident within 180 seconds. The magnitude of the agreement was similar to that of our prior instrument validation study co-locating (<1 km) AMODv2s with an AERONET monitor in northern Colorado (Wendt et al., 2021). Results from our network indicated that median $PM_{2.5}$:AOD ratio was relatively unimpacted by the presence or absence of smoke at sites in the Midwest, Mountain West, and Northeast regions of the United States. However, median $PM_{2.5}$:AOD ratio was higher on smoke-free days in California relative to days with smoke present. In California, median $PM_{2.5}$ concentration trends on smoky vs. non-smoky days differed from climatological expectations (Bian et al., 2020). We used data collected by participants to identify probable instances of smoke aloft, which low-cost $PM_{2.5}$ monitors alone could fail to properly characterize. As networks of crowdsourced low-cost $PM_{2.5}$ monitors expand, further work in expanding access to co-located, low-cost AOD monitors strengthen these networks' ability to monitor evolving smoke events. By crowdsourcing our low-cost instruments we empower local communities to stay informed on wildfire events impacting their daily lives.

CHAPTER 4. A CLOUD SCREENING ALGORITHM FOR GROUND-BASED AEROSOL OPTICAL DEPTH MEASUREMENTS USING ALL-SKY IMAGES AND DEEP TRANSFER LEARNING

Chapter Overview

Aerosol optical depth (AOD) is used to characterize aerosol loadings within Earth's atmosphere. However, AOD measurements can be biased by cloud obstruction. We present a novel deep transfer learning model on all-sky images to support more accurate AOD retrievals. We used three independent image datasets for training and testing: the novel Northern Colorado All-Sky Image (NCASI), the Whole Sky Image SEGmentation (WSISEG), and the METCRAX-II datasets from the National Center for Atmospheric Research (NCAR). The model is intended to classify whole-sky images as: 1) clear sky, 2) thin cirrus obstructing the solar disk, and 3) thick, non-cirrus clouds obstructing the solar disk. The best-performing model successfully classified 95.5%, 96.9%, and 89.1% of testing images from NCASI, METCRAX-II and WSISEG datasets, respectively. Our results demonstrate that all-sky imaging with deep transfer learning can be applied toward cloud screening, which would aid ground-based AOD measurements.

Introduction

The abundance of aerosols in the atmosphere can be quantified optically from surface-based instruments called sun photometers, which measure aerosol optical depth (AOD), a dimensionless metric of light extinction by particles. Accurate AOD measurement requires a clear view of the solar disk. If the solar disk is partially obscured by clouds, measured AOD will be biased high. Thus, reliable AOD measurements from sun photometers require robust cloud screening.

Prior work has implemented quality control algorithms to reduce errors from cloud cover. The Aerosol Robotics Network (AERONET) provides automatic, multi-wavelength AOD measurements at hundreds of locations (Holben et al., 1998). Smirnov et al. (2000) leveraged the functionality of AERONET sun photometers to develop an automated cloud screening protocol for AOD measurements. Each AOD measurement is the average of a triplet of measurements, with 30 seconds between each measurement (Smirnov et al., 2000). The triplet is classified as cloud-contaminated if the maximum difference within the triplet exceeds an empirically derived threshold, based on the assumption that AOD variability within short time periods is more likely due to clouds than rapid changes in aerosol (Smirnov et al., 2000). Additional screening steps incorporated all AOD measurements across a day, followed by screening by an analyst. This latter process often lagged the initial measurement by months and may be affected by analyst inconsistencies (Smirnov et al., 2000).

Alexandrov et al. (2004) proposed an algorithm based on the change in AOD as a function of the Azithumal position of the sun for Multi-Filter Rotating Shadowband Radiometers (MFRSRs). The algorithm had a 3.4% false negative rate, wherein the data point was identified as clear when actually cloud-contaminated, and a 4.0% false positive rate (Alexandrov, 2004). This approach is less conservative than the AERONET algorithm, (i.e. it will classify fewer points as cloud-contaminated) but is also independent of calibration, enabling effective cloud screening during calibration (Alexandrov, 2004). A subsequent study (Giles et al., 2019), used LIDAR to detect cirrus clouds and derive empirical thresholds, improving the removal of cirrus contamination by the AERONET algorithm.

A weakness of sensor-based cloud screening algorithms is their instrument-specific nature (Smirnov et al., 2000). Algorithms for AERONET instruments are likely not applicable to different instruments, which have different hardware (Wendt et al., 2021). When porting algorithms, unique empirical thresholds must be determined to remain effective, assuming the necessary sensors are present at all (Wendt et al., 2021). Previous studies using handheld sun photometers relied on operator observations of the sky to assess clouds (Boersma & de Vroom, 2006; Ford et al., 2019a; Wendt et al., 2019a). This approach requires consistent operator attention, which defeats the practical and personnel advantages gained via automated measurements, and relies on the subjective opinion of the operator.

To date, image-based analysis of clouds has not been used in support of ground-based AOD measurement quality control. However, prior work in the area of cloud classification of sky images can be leveraged toward this aim. Long et al. (2006) used whole-sky images (collected using fish-eye lenses to achieve at least 160 degrees field of view) to evaluate cloud brokenness, uniformity, and solar obstruction (Long et al., 2006). Their algorithm, which separated images into their red, green, and blue (RBG) components, and classified pixels based on their R/B values, performed well under uniform sky conditions (>95% accuracy for solar disk obstruction), but was less accurate for images with more irregular cloud coverage (<85% accuracy for solar disk obstruction) (Long et al., 2006).

Calbo & Sabburg (2008) mathematically defined sky conditions based on whole sky images, determined using six image features: mean, standard deviation, smoothness, third moment, uniformity, and entropy determined using the R/B values and an intensity parameter. The Calbo & Sabburg (2008) algorithm was least effective at discriminating between cases with thin clouds present and covering or not covering the sun (Calbó & Sabburg, 2008). For cloud classification, this distinction may not be significant; however, for AOD cloud screening applications, the presence of thin cirrus clouds can substantially bias a measurement (Smirnov et al., 2000; Alexandrov, 2004; Giles et al., 2019); thus, determining if thin clouds are present (and covering the sun) is critical.

Other cloud classification schemes aim to better distinguish cirrus from clear sky. Heinle et al. (2010) use R – B rather than R/B along with additional image features (difference, energy, contrast, and homogeneity); and a non-parametric k-Nearest-Neighbors classifier. Li et al. (2011) found better results for cirrus clouds when they first classified sky cases as unimodal (only clouds or only sky) or bimodal (mixture of cloud and sky) before applying their algorithm. Liu et al. (2013) proposed the use of multiple images taken over the course of several minutes to form a Tensor Ensemble of images (Shuang Liu et al., 2013), an approach which leverages the dynamic nature of cloud movement to help distinguish it from forward scattering around the solar disk.

Recent research has applied machine learning techniques to evaluate sky condition and cloud coverage from images (Gu et al., 2018). Taravat et al. (2015) applied both a multilayer perceptron neural network (MLP) and a support vector machine classification to whole-sky image classification, which both showed classification accuracy than previous thresholding approaches. Xia et al. (2015) proposed the use of a hybrid method using an extreme learning machine and kNN. Prior to classification, textural, color, and shape features were extracted. The classification accuracy increased with increasing numbers of features (Xia et al., 2015). Deep convolutional neural network (CNN) models are particularly effective for feature extraction and classification on image data (Gu et al., 2018). Shi et al. (2017) presented a CNN model for cloud identification based on common image classification architectures and demonstrated that their model outperformed prior feature-based models on the same dataset. Liu et al. (2018) incorporated temperature, humidity, pressure, wind speed, and maximum wind speed parameters with visual image data to further refine cloud type classification on whole sky images. Zhang et al. (2018) presented CloudNet, a deep CNN model tailored to extract cloud features and classify images based on cloud type; it has achieved the best results on both partial and whole-sky images. The success of prior work in image-based cloud identification supports the use of similar algorithms for solar obstruction screening for AOD quality control.

In previous machine learning cloud classification models, training and testing data were drawn from the same domain, namely sky images (S. Liu et al., 2018; Shi et al., 2017; Zhang et al., 2018). Recent research into deep learning has explored transfer learning, in which data from outside the application domain are incorporated into model development (Pan and Yang, 2010; Zhuang et al., 2021 and references therein). Transfer learning for image classification leverages the most effective pre-trained CNN models to compute features useful for building classifiers for a wide variety of image classes and then applies the models to specific domains (Zhuang et al., 2021). Effective deep CNN models can have over 10 million trainable parameters (Simonyan & Zisserman, 2015); however, in transfer learning, these parameters are trained in advance, outside of the application domain (Pan & Yang, 2010; Zhuang et al., 2021), thus reducing the time and resource requirements. In practice, pre-trained parameters may be frozen or left trainable. If frozen, application-specific training is reduced to the classification layer parameters, which are a small fraction of the total parameters. If left trainable, application-specific training is accelerated as parameters will likely be initialized closer to their optimal values.

The Visual Geometry Group (VGG) at Oxford University developed the VGG-16 model, a deep CNN model designed for image classification (Simonyan & Zisserman, 2015). VGG-16 was trained on ImageNet, an image database consisting of over 14 million images from 20,000+ unique image classes

(Deng et al., 2009; Simonyan & Zisserman, 2015). VGG-16 has been used effectively for transfer learning in image classification applications (e.g. Tammina, 2019; Guan et al., 2019; Kaur and Gandhi, 2019).

Here, we develop a transfer learning model based on VGG-16 for cloud screening on all sky images. We first present a new all-sky-image data set collected using a low-cost prototype imaging module. We then describe the image pre-processing algorithm used to prepare images for input into our classification model and the design and training of our classification model. Finally, we evaluate the performance of our algorithm on three independent all-sky-image sets.

Materials and methods

All-sky images

Image-based cloud screening for AOD measurements requires the co-location or hardware integration of an all-sky camera with a sun photometer. All-sky cameras suitable for AOD cloud screening must image the solar disk and the sky surrounding the solar disk while preserving the edge detail of the solar disk and nearby clouds. Images cannot be used for AOD screening if the image area surrounding the solar disk is saturated with sunlight. For this reason, all-sky cameras used in previous cloud identification studies were designed to block or attenuate incident sunlight (e.g Calbó & Sabburg, 2008; Fa et al., 2019; Xie et al., 2020).

Accordingly, we located two pre-existing sets of all-sky image data with which to train and test our model, in addition to a third data set that was collected specifically for this project (see the top row of Fig. 4-1 for example images). The Whole Sky Image SEGmentation (WSISEG) data set includes 400 all-sky images captured on the 25 rooftop of Anhui Air Traffic Management Bureau, Civil Aviation Administration of China in July 2018 (Xie et al., 2020). The resolution of images in WSISEG is 2,000 pixels × 1,944 pixels. The National Center for Atmospheric Research (NCAR) published METCRAX-II ISS All Sky Camera Imagery (UCAR/NCAR - Earth Observing Laboratory, 2016). The images in METCRAX-II were collected on the island of Diego Garcia between September 2011 and February 2012. The resolution of these images is 640 pixels × 480 pixels. Due to the large data storage requirements (6,128 MB), we did not save the entire data set. We manually selected all day-time images that were later evaluated for inclusion into training and testing data.

Finally, we created a collection of all sky images called the Northern Colorado All-Sky Image (NCASI) set, using a custom-designed imaging module (see Text C1, Fig. C1, and Table C1 in Appendix C). We collected a total of 3,544 images between 1-21 September 2021. Images from the 1st, 20th, and 21st of September 2021 were collected near a private residence in Boulder, Colorado, USA. Images from all other days were collected at the Powerhouse Energy Campus at Colorado State University (430 N. College Avenue, Fort Collins, Colorado, USA). The image resolution is 1,920 pixels × 1,200 pixels. The module was configured such that when started, it captured an all-sky image every 30 seconds.

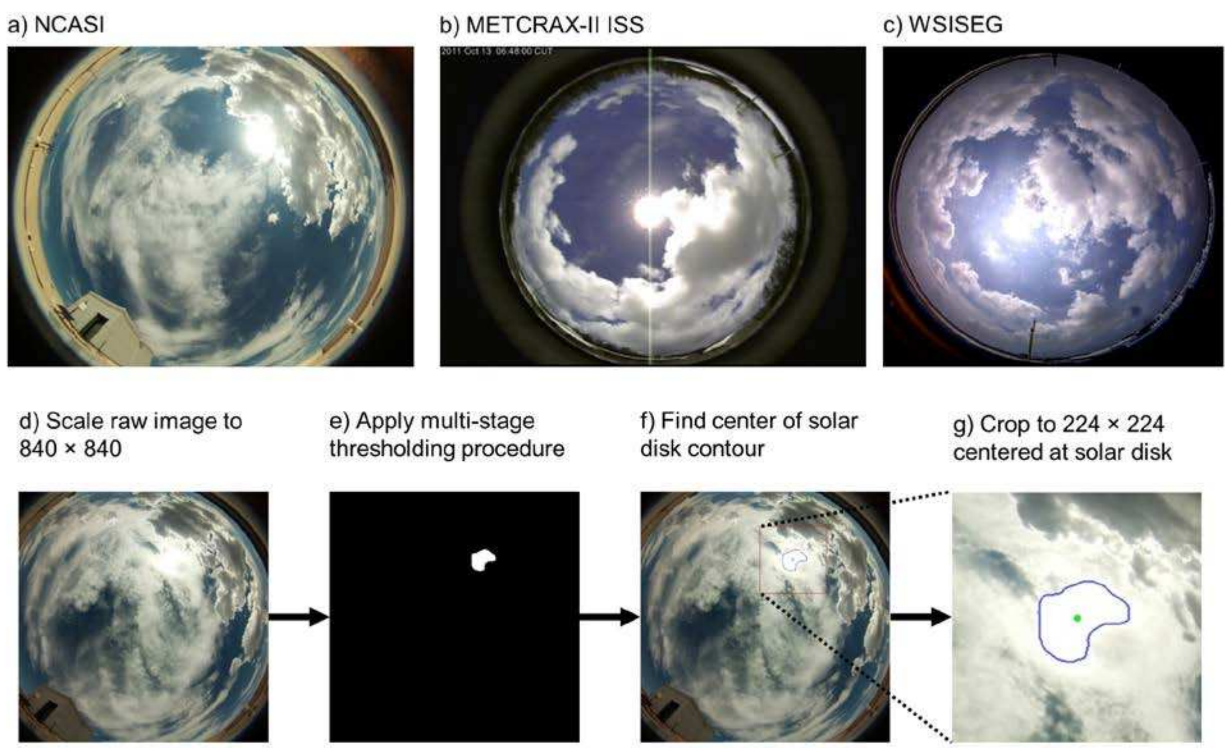

Figure 4-1: Top: example images from the a) NCASI, b) METCRAX, and c) WSISEG datasets. Bottom: Transformations performed in image preparation algorithm. d) Raw image files are scaled and cropped to a uniform size 840 pixels × 840 pixels). e) A multi-stage thresholding algorithm isolates sunlit pixels. f) From the sunlit contour, we calculate the center of the sunlit pixels in the image. g) Using the center, we crop the image (224 pixels × 224 pixels).

We built our algorithm to classify three types of sky conditions: 1) sun not obscured by clouds; 2) sun obscured by thin cirrus clouds; and 3) sun obscured by thick, non-cirrus clouds. We manually labeled images according to these designations. We discriminated between cirrus and thick, non-cirrus clouds based on how the cloud cover impacted the shape of the solar disk. Images where the solar disk presented as circular, despite the presence of thin cloud cover, were designated as cirrus. Images where the shape of the solar disk deviated from circular due to the presence of cloud cover were designated as non-cirrus. We built and selected our model training and testing datasets to include samples from all three enumerated sky conditions. We provide example images from each data set under each sky condition in Fig. C2.

Image preparation

Initially, input images of potentially varying sizes are scaled to a common a size of 840 pixels $\times$ 840 pixels using an area pixel model as illustrated in Fig. 4-2a (Chun-Ho Kim et al., 2003). This operation normalizes important features such as the size of the solar disk in square pixels. From the scaled image, we isolate the blue color channel and perform a binary threshold operation, where 8-bit pixel values greater than or equal to 252 are set to the maximum value of 255 (white) and all other pixels are set to zero (black). We then apply smoothing to the thresholded image using 15 $\times$ 15 Gaussian kernel with the standard deviations in the horizontal and vertical axes set to zero (Burt, 1981). We found that for images with sunlit clouds near, but not obstructing, the solar disk, a Gaussian filter alone was insufficient to smooth the edges of sunlit contours. Therefore, we applied an additional bilateral filter with a pixel-neighborhood diameter of 15 pixels with color and space standard deviation values set to 25 (Tomasi & Manduchi, 1998). The smoothing steps blend black pixels with white pixels, leaving pixels near the contour edge with values between 0 and 255. To restore the image to a binary image, we apply an additional binary thresholding operation with the threshold set to a pixel value of 50. The result of these thresholding operations is illustrated in Fig. 4-2b. We then apply a contour detection algorithm to the binary image to derive contour arrays for each area of contiguous white pixels (Suzuki & be, 1985). For images with a high number of optically saturated clouds (common in the METCRAX-II data set), there can be multiple sunlit contours that are not the solar disk. To isolate the solar disk contours from surrounding sunlit clouds, we apply pixel area and circularity criteria. Across all three datasets, the solar disk for scaled images greater than 2750 square pixels. Contours greater than that threshold are evaluated based on their circularity, defined as follows:

$$C = \frac{4 \cdot \pi \cdot A}{L^2} \tag{3-1}$$

where C is the circularity, A is the area, and L is the arc length of the contour. Among contours within the acceptable area range, the contour with the highest circularity is considered the contour of the solar disk. We then calculate the centroid of the solar disk contour and crop the image to a region sized 224 pixels $\times$ 224 pixels centered at the centroid of the solar disk, as depicted in Fig. 4-2c and Fig. 4-2d. For images where the solar disk is fully obscured and no pixels pass the binary thresholding tests (most often due to heavy cloud cover obscuring the solar disk), the center of the cropped image is placed at the center of the original scaled image and the resulting image is given the “cloud” label. In other cases, as in Fig. 4-2, the solar disk is obscured by clouds such that its true center cannot be determined. For these images, the approximate center is used and the image is given the cloud label. For images where the calculated solar disk center is within 112 pixels of an edge, the image cannot be cropped to 224 pixels $\times$ 224 pixels. In these cases, the edge of the scaled image (Fig. 4-2a) is used as the edge of the cropped image, and the cropped image will be smaller than 224 pixels $\times$ 224 pixels.

Model Design

For our feature extraction layers, we used the VGG-16 deep CNN model with parameters pre-trained on ImageNet (Deng et al., 2009; Simonyan & Zisserman, 2015). We used the Tensorflow implementation of VGG-16 (Abadi et al., 2015). VGG-16 expects input tensors sized 224 $\times$ 224 $\times$ 3, with the third dimension representing RGB color channels present in colored images (Simonyan & Zisserman, 2015). After image preparation, most images were suitable for input into the VGG-16 model without further resizing. Images that were cropped to smaller proportions (i.e., center of the solar disk was close to the image edge) were scaled to the proper input size and then passed to a data generator implemented in Tensorflow's VGG-16 model (Abadi et al., 2015). The output of the pre-trained VGG-16 model is a $7 \times 7 \times 512$ tensor representing the features learned from the ImageNet database (Deng et al., 2009; Simonyan & Zisserman, 2015). To interface with VGG-16, the output tensor from the feature extraction layers is flattened to a one-dimensional vector, which is interfaced with a three-node dense classification layer with a softmax activation function (Bridle, 1989). For a particular input sample, the output layer gives a probability estimate for each of the three possible sky condition classes. The class associated with the highest probability value is the classification of the model.

Model training and evaluation

We created training and testing subdatasets for NCASI, METCRAX-II, and WSISEG using a random split of approximately two-thirds training and one-third testing data. Training-testing data partitions for each set and class designation are provided in Table 4-1. The images in the training and testing sets were pre-processed using our image preparation algorithm. We built seven datasets, each with training and testing subsets: one for each data set, one for each of the possible combinations, and a single combination of all three.

Table 4-1: Training and testing data partitions by data set and class designation.

Image data set	Class designation	Number of training samples	Number of testing samples
NCASI	Clear	59	27
NCASI	Cirrus	70	24
NCASI	Cloud	52	38
METCRAX-II	Clear	204	118
METCRAX-II	Cirrus	124	73
METCRAX-II	Cloud	239	133
WSISEG	Clear	35	18
WSISEG	Cirrus	26	22

WSISEG	Cloud	198	98

We trained all models using the same training parameters. We used the categorical cross-entropy loss function and the Adam optimizer with a learning rate of 0.00012 (Kingma & Ba, 2017). We trained for 100 epochs with a batch size of 32. We did not modify pre-trained weights of VGG-16. The weight and bias parameters of the output layer were the only trainable parameters, which comprised 0.51% of the total model parameters.To limit overfitting, we also applied data augmentation, which supplements training data by producing batches of randomly modified images created via transformation operations. We implemented data augmentation in Tensorflow using the ImageDataGenerator module (Abadi et al., 2015) allowing rotation of 20 degrees, width/height shifting of 10%, a zooming range of 20%, and random horizontal and vertical flips. We trained our model using a GPU (NVIDIA, Tesla K80, Santa Clara, California, USA) on the Google Colaboratory platform (Google, Mountain View, California, USA).

Results and discussion

Model evaluation

Holding model and training parameters constant, we trained seven different models using the seven possible combinations of the three training datasets. We evaluated each of the seven models according to classification accuracy on the three testing datasets. Separating cirrus and non-cirrus cloud classifications is useful for interpreting classification results, though it is not strictly necessary for AOD quality control. We assessed the model's performance on the binary classification problem typically addressed by AOD quality control algorithms by combining cirrus and non-cirrus cloud designations. The results of our analysis are presented in Table 4-2.

Table 4-2: Three-class (two-class) classification accuracy of models trained on seven different training datasets. Three-class results are from cirrus, clear, and cloud categories. Two-class results are from clear and cloud categories. Accuracy metrics for each model were calculated using the testing data from NCASI, METCRAX-II, and WSISEG individually. The best performing model on each test data set is given in bold text.

Training data set(s)	Model Number	Accuracy on NCASI (%)	Accuracy on METCRAX-II (%)	Accuracy on WSISEG (%)
NCASI	1	**97.8** (**100.0**)	76.2 (81.2)	75.4 (87.0)
METCRAX-II	2	84.3 (84.2)	95.7 (95.7)	79.0 (79.0)
WSISEG	3	64.0 (78.7)	62.0 (68.5)	89.1 (94.2)
NCASI and METCRAX-II	4	95.4 (**100.0**)	95.5 (97.5)	82.6 (87.0)
NCASI and WSISEG	5	92.1 (**100.0**)	70.0 (72.5)	**90.0** (**95.7**)
METCRAX-II and WSISEG	6	84.3 (96.6)	95.7 (97.8)	89.1 (94.9)
All	7	95.5 (**100.0**)	**96.9** (**98.4**)	89.1 (94.2)

The model trained on all training data (model 7) generalized the best to the testing data. Model 7 correctly classified 95.5%, 96.9% and 89.1% of testing samples from NCASI, METCRAX-II, and WSISEG images, respectively (Table 4-2). The three-class confusion matrix for model 7 is given in the top row of Fig. 4-2 (three-class confusion matrices for the remaining six models are provided in the top rows of Figs. C4 through C9). The accuracies of model 7 for the binary classification were 100.0%, 98.4%, and 94.2% for NCASI, METCRAX-II and WSISEG, respectively (Table 4-2). We present the binary confusion matrix for model 7 in the bottom row of Fig. 4-2 (see bottom rows of Figs B4-B9 for the remaining six models).

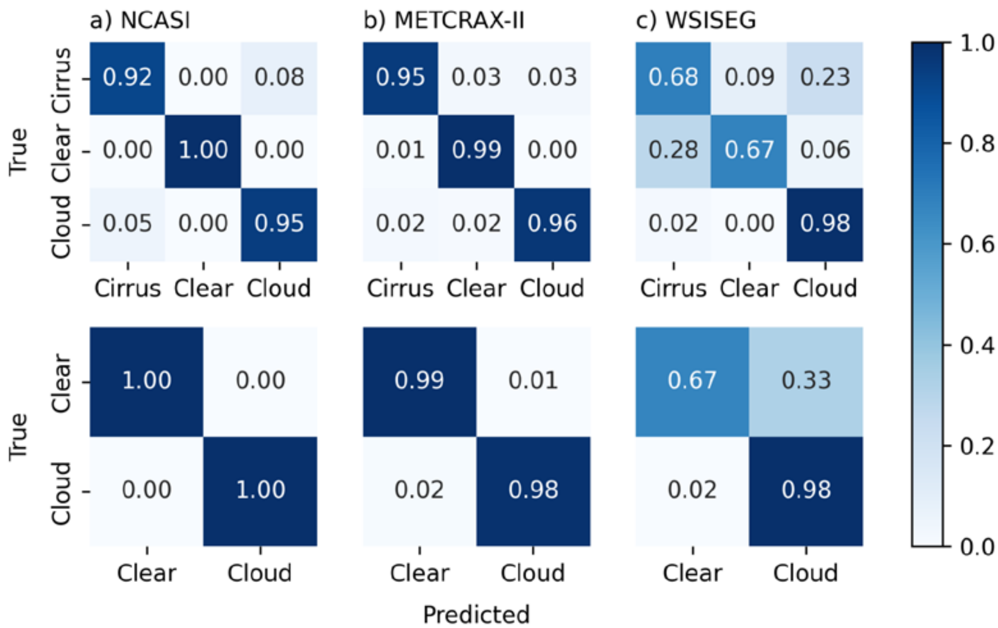

Figure 4-2: Top row: three-class (cirrus, clear, and cloud) confusion matrices for transfer learning model trained on NCASI, METCRAX-II, and WSISEG training datasets (model 7). Bottom row: Two-class (clear and cloud) confusion matrices for transfer learning model trained on NCASI, METCRAX-II, and WSISEG training datasets (model 7). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. "Predicted" refers to the model output and "True" refers to the observed class designation.

For both the three-class and binary classification problems, model 7 performed best on the NCASI images, followed by METCRAX-II, and WSISEG, respectively (Fig. 4-2). Three of the four models at least partially trained on the NCASI training set achieved over 95.0% accuracy on the NCASI testing set (Table 4-2). On both the NCASI and METCRAX-II testing images, model 7 exhibited the highest error rate on images manually labeled as cirrus. On the NCASI data, 8.0% of cirrus images were classified as cloud (not cirrus) images and 0.0% were classified as clear images. For METCRAX-II, 3.0% of the cirrus images were classified as cloud (not cirrus) images, and 3.0% were classified as clear images. However, in the two-class problem, both cirrus and non-cirrus clouds are classified identically, rendering errors between cirrus and cloud categories immaterial for the inherently binary problem of AOD cloud screening. When the two cloud types are combined, model 7 correctly classifies 100.0% of samples from the NCASI set. For METCRAX-II data, model 7 incorrectly classifies 1.0% of clear images as cloud, and incorrectly classifies 2.0% of clear images as cloud. For WSISEG data, high rates of confusion between clear and cirrus images led to relatively poor results in the two-class problem, with 33.0% of clear images being classified as cloud images, but only 2.0% of cloud images being classified as clear.

Model 7 performed well relative to prior AOD screening algorithms, but these results partially depended on the testing dataset. The algorithm proposed in Alexandrov et al. (2004) had a false negative (i.e. cloud classified as clear) rate of 3.4% and a false negative (i.e. clear classified as cloud) rate of 4.0% for the binary cloud screening problem. Our model performed better than the prior algorithm on NCASI and METCRAX-II images, and worse on WSISEG images (Fig. 4-2). The false negative rates on NCASI, METCRAX-II, and WSISEG testing data were 0.0%, 2.0%, and 2.0%, respectively (Fig. 4-2). The false positive rates on NCASI, METCRAX-II, and WSISEG testing data were 0.0%, 1.0%, and 33.0%, respectively (Fig. 4-2). In Alexandrov et al. (2004), the authors do not specify the relative proportions of cirrus and non-cirrus cloud samples present in the 575 cloud-contaminated samples they analyzed, precluding a direct comparison of performance on cirrus cases. The generalizability of our models to images similar to those in NCASI is supported by the performance of model 6 (trained only on METCRAX-II and WSISEG) on NCASI. Despite not seeing any NCASI images during training, model 6 correctly classified 96.6% of NCASI testing images in the two-class problem (Table C3).

Model 7 performed well on NCASI images because, in part, thin and thick clouds not directly in front of the sun were less likely to be saturated (i.e. maximum RGB values) with light (e.g. Fig. 4-1a). In the METCRAX-II and WSISEG datasets, clouds outside of direct sunlight, yet still in the 224 pixel $\times$ 224 pixel frame, were more likely to be fully saturated with light (e.g. Figs. 4-1b and 4-1c). In the WSISEG data set, there were apparent camera glare spots around the solar disk (Figs. C10 through C12 show all images misclassified by model 7). We suspect these glare spots were a result of sunlight reflection off of plastic or glass protective coverings over the imaging systems. The METCRAX-II data set has less severe glare spots, and the NCASI data set lacks glare spots entirely.

Limitations

A limitation of this work was the subjectivity of the class label designations, particularly for the METCRAX-II and WSISEG datasets (Figs. C10 through C12). For the NCASI data that was collected specifically for this project, we had the advantage of observing the actual condition of the sky as the images were collected. However, with the METCRAX-II and WSISEG datasets, we could only assign class designations using images. This complicated class designations, particularly between cirrus and clear images in the WSISEG data set, where it was difficult to distinguish between glare and cirrus clouds (e.g. Fig. C12). Glare in WSISEG images may have contributed to relatively poor results distinguishing between clear and cirrus designations for all models (Fig. 4-2 and Figs. C4 through C9). However, issues with the WSISEG data did not impact the performance of models 5-7, which partially used WSISEG data. Models 5-7 generalized well to NCASI and METCRAX-II despite potential mislabeling of WSISEG samples. Clarity issues in WSISEG images emphasize the importance of high image fidelity around the solar disk for AOD screening application. In future work, we will test the generalization ability of model 7 on additional independent data without additional training.

Our model generally performed well on a variety of different cloud types from three independent datasets. However, we did not test our model on images with high aerosol loading. Wildfire smoke, volcanic ash, heavy industrial emissions and other sources of high atmospheric aerosol concentrations could be erroneously classified as clouds. Additional images with the sun obscured by high levels of aerosol are needed to test the model's sensitivity to aerosol concentrations.

Our datasets were also limited to images from three unique camera configurations. Different imaging systems produce images with different hues, resolutions, and saturation levels, among other visual differences. As demonstrated on the NCASI and METCRAX-II datasets (Tables C2 and C3), our model performs best on images with limited lens glare spots. To improve performance on images from different imaging systems, we recommend further training of our output layer parameters using additional labeled images. Models 1-7 models and their respective weights are publicly available (see data availability section).

Conclusions

In this work, we presented a novel approach for cloud screening that could be applied to AOD measurements and that builds on the literature surrounding cloud classification from whole sky images. We designed an imaging module to capture all-sky images with high-fidelity, particularly around the solar disk. Using this imaging module, we produced the NCASI data set, a novel collection of all-sky images from Northern Colorado that includes images of a variety of cloud and sky conditions. Combining our NCASI data set with two other independent all-sky image datasets and the pre-trained VGG-16 model, we applied transfer learning to develop an effective cloud screening model. Our model performed well classifying the solar disk as unobstructed (clear), obstructed by thin cirrus clouds, or obstructed by optically thick, non-cirrus clouds.

Image-based cloud screening is agnostic of sun photometer hardware, making it especially applicable for lower-cost sun photometers (Wendt et al., 2019a, 2021), which may lack the hardware and personnel required to fully implement state-of-the-art sensor-based cloud screening. When paired with AOD measurements, all-sky images may be used for relatively simple quality control and descriptive purposes. Our work suggests that cameras configured to produce images like the NCASI datasets will perform well in quality control applications. The performance of our model on cirrus cloud cover represents a promising advancement in AOD quality control for cloud cover most difficult to distinguish from elevated aerosol.

CHAPTER 5. CONCLUSIONS

In this work, we present novel low-cost air quality measurement technologies and their applications. We first describe the development and validation of the AMODv2, a low-cost instrument capable of simultaneous measurement of $PM_{2.5}$ mass concentration and AOD. We then detail our deployment of AMODv2s in a crowdsourced air pollution monitoring network. There we show AMODv2s can be operated by student volunteers to monitor local air quality and assess the regional variability of ambient air pollution in the presence of wildfire smoke. Finally, we present an image-based machine learning algorithm for screening AOD measurements for cloud contamination.

From a technological perspective, the AMODv2 is a step toward the synthesis of reference-grade and low-cost air quality monitors. Electronic components that were cost-prohibitive or non-existent in recent decades are now widely available at low cost. In many other industries (e.g. telephones and consumer appliances), the proliferation of low-cost and high-performance components has produced performance improvements with simultaneous downward price pressure. The AMODv2 represents a major turn away from the cost-performance tradeoff that has limited air pollution monitoring. Pressure for greater performance at lower cost is often driven by consumer demands. By incorporating human factor design in the development of the AMODv2, we were cognizant of the preferences of a broader range of potential end-users. The AMODv2 is for both the specialized aerosol scientist and the interested community member. With this in mind, we sought to maximize the degree of automation of the AMODv2 while developing a simple and familiar configuration interface. Collecting air quality data-with the AMODv2 requires only basic filter handling and executing familiar tasks such as operating a smartphone application and connecting a device to Wi-Fi. Our crowdsourced sampling campaigns demonstrated that, after a one-hour training session, non-scientists were ready to operate AMODv2s. With our validation results, we show that measurements collected by AMODv2 users are accurate relative to reference monitors.

The modernization of air quality monitoring is not only a matter of hardware. Modern developments in algorithms and software have helped hardware systems realize their full potential. Deep learning algorithms in particular are becoming the dominant solution for a range of software tasks including text, audio, and image analysis. With our transfer learning model on all-sky images, we present a platform-independent algorithm for AOD cloud screening. High-quality image sensors are now available at relatively low-cost (< $50), making it more affordable than ever to incorporate all-sky imaging module into air quality monitoring systems. We demonstrated our model accurately discriminates between clear and cloud-contaminated all-sky images.

Our technology points to a future of community-driven air pollution monitoring. Crowdsourced networks of low-cost nephelometers (e.g. Plantower PMS5003) have made a substantial impact on air pollution research in recent years (Gupta et al., 2018; Lin et al., 2020; Li et al., 2020; Badura et al., 2020; Lu et al., 2021; Chadwick et al., 2021). Our crowdsourced network of AMODv2s highlighted how $PM_{2.5}$ and AOD measurements, collected in tandem, can illuminate details of air pollution events beyond what is possible with $PM_{2.5}$ measurements alone. Our image-based quality control approach can ensure that crowdsourced AOD measurements free from cloud-contamination. We have facilitated a future where accurate and fully-automated AOD sensors could be distributed to citizen operators at similar scales to $PM_{2.5}$ sensors. Our work on image-based cloud screening provides a simple and AOD-sensor independent approach to ensure future crowdsourced AOD measurements are of high quality.

Additional steps toward our overall goal of improving community-driven air quality monitoring include additional design improvements to the AMODv2. The primary limitation of the AMODv2 was mechanical robustness, particularly in heavy rain. Any potential AMODv3 design will feature improved material selection and manufacturing processes to improve durability in long-term deployments. The integration of a fisheye lens into future AMOD iterations would also represent a significant improvement into the technology. Hardware and software integration of a fisheye lens on a future AMODv3 design would facilitate one-shot AOD quality control for instruments in the field, even for instruments without internet access. Future work is needed to evaluate the feasibility of executing our transfer learning algorithm on relatively limited processors. We would seek to maintain the accuracy of the algorithm while scaling back the computational complexity, a goal facilitated by ongoing work in embedded processing and deep learning complexity optimization. An additional limitation of our work was the relative scale of our national crowdsourced network. Though we were able to detect air pollution anomalies during wildfire events, our crowdsource network did not include enough sites and did not span a great enough period of time to evaluate broad air quality trends in the United States. However, there are already thousands of low-cost nephelometers actively measuring $PM_{2.5}$ concentrations across the globe. What is far less common, are low-cost AOD sensors that could be paired with these nepholomenters. The AOD measurement subsystem of the AMODv2 is at least one order of magnitude less expensive than prior automated AOD measurement devices. The design and production of low-cost AOD sensors based on the AMODv2 subsystem could bolster $PM_{2.5}$ measurement networks, moving us closer to long-term and broad-scale air pollution monitoring by crowdsourced networks. With a low-cost nephelometer, a low-cost AOD sensor, and sky camera, volunteers across the world could contribute air pollution data at scopes previously unachievable.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2015). *TensorFlow: A system for large-scale machine learning*. 21.

Alexandrov, M. D. (2004). Automated cloud screening algorithm for MFRSR data. *Geophysical Research Letters*, *31*(4), L04118. https://doi.org/10.1029/2003GL019105

Arku, R. E., Birch, A., Shupler, M., Yusuf, S., Hystad, P., & Brauer, M. (2018). Characterizing exposure to household air pollution within the Prospective Urban Rural Epidemiology (PURE) study. *Environment International*, *114*, 307–317. https://doi.org/10.1016/j.envint.2018.02.033

Badura, M., Sówka, I., Szymański, P., & Batog, P. (2020). Assessing the usefulness of dense sensor network for PM2.5 monitoring on an academic campus area. *Science of The Total Environment*, *722*, 137867. https://doi.org/10.1016/j.scitotenv.2020.137867

Barkjohn, K. K., Gantt, B., & Clements, A. L. (2021). Development and application of a United States-wide correction for PM_{2.5} data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques*, *14*(6), 4617–4637. https://doi.org/10.5194/amt-14-4617-2021

Bian, Q., Ford, B., Pierce, J. R., & Kreidenweis, S. M. (2020). A Decadal Climatology of Chemical, Physical, and Optical Properties of Ambient Smoke in the Western and Southeastern United States. *Journal of Geophysical Research: Atmospheres*, *125*(1). https://doi.org/10.1029/2019JD031372

Boersma, K. F., & de Vroom, J. P. (2006). Validation of MODIS aerosol observations over the Netherlands with GLOBE student measurements. *Journal of Geophysical Research*, *111*(D20), D20311. https://doi.org/10.1029/2006JD007172

Brauer, M., Freedman, G., Frostad, J., van Donkelaar, A., Martin, R. V., Dentener, F., Dingenen, R. van, Estep, K., Amini, H., Apte, J. S., Balakrishnan, K., Barregard, L., Broday, D., Feigin, V., Ghosh, S., Hopke, P. K., Knibbs, L. D., Kokubo, Y., Liu, Y., … Cohen, A. (2016). Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. *Environmental Science & Technology*, *50*(1), 79–88. https://doi.org/10.1021/acs.est.5b03709

Brey, S. J., Ruminski, M., Atwood, S. A., & Fischer, E. V. (2018). Connecting smoke plumes to sources using Hazard Mapping System (HMS) smoke and fire location data over North America. *Atmospheric Chemistry and Physics*, *18*(3), 1745–1761. https://doi.org/10.5194/acp-18-1745-2018

Bridle, J. S. (1989). Training Stochastic Model Recognition Algorithms as Networks can Lead to Maximum Mutual Information Estimation of Parameters. *Proceedings of the 2nd International Conference on Neural Information Processing Systems*, 211–217.

Brooks, D. R., & Mims, F. M. (2001a). Development of an inexpensive handheld LED-based Sun photometer for the GLOBE program. *Journal of Geophysical Research: Atmospheres*, *106*(D5), 4733–4740. https://doi.org/10.1029/2000JD900545

Brooks, D. R., & Mims, F. M. (2001b). Development of an inexpensive handheld LED-based Sun photometer for the GLOBE program. *Journal of Geophysical Research: Atmospheres*, *106*(D5), 4733–4740. https://doi.org/10.1029/2000JD900545

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., & Loxham, M. (2019). Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Scientific Reports*, *9*(1), 7497. https://doi.org/10.1038/s41598-019-43716-3

Burt, P. J. (1981). Fast Filter Transforms for Image Processing. *Computer Graphics and Image Processing*, *16*(1), 20–51. https://doi.org/10.1016/0146-664X(81)90092-7

Calbó, J., & Sabburg, J. (2008). Feature Extraction from Whole-Sky Ground-Based Images for Cloud-Type Recognition. *Journal of Atmospheric and Oceanic Technology*, *25*(1), 3–14. https://doi.org/10.1175/2007JTECHA959.1

Cascio, W. E. (2018). Wildland fire smoke and human health. *Science of The Total Environment*, *624*, 586–595. https://doi.org/10.1016/j.scitotenv.2017.12.086

Chadwick, E., Le, K., Pei, Z., Sayahi, T., Rapp, C., Butterfield, A. E., & Kelly, K. E. (2021). Technical note: Understanding the effect of COVID-19 on particle pollution using a low-cost sensor network. *Journal of Aerosol Science*, *155*, 105766. https://doi.org/10.1016/j.jaerosci.2021.105766

Cheeseman, M., Ford, B., Volckens, J., Lyapustin, A., & Pierce, J. R. (2020). The Relationship Between MAIAC Smoke Plume Heights and Surface PM. *Geophysical Research Letters*, *47*(17). https://doi.org/10.1029/2020GL088949

Chun-Ho Kim, Si-Mun Seong, Jin-Aeon Lee, & Lee-Sup Kim. (2003). Winscale: An image-scaling algorithm using an area pixel model. *IEEE Transactions on Circuits and Systems for Video Technology*, *13*(6), 549–553. https://doi.org/10.1109/TCSVT.2003.813431

Cleland, S. E., Serre, M. L., Rappold, A. G., & West, J. J. (2021). Estimating the Acute Health Impacts of Fire-Originated PM $_{2.5}$ Exposure During the 2017 California Wildfires: Sensitivity to Choices of Inputs. *GeoHealth*, *5*(7). https://doi.org/10.1029/2021GH000414

Cleland, S. E., West, J. J., Jia, Y., Reid, S., Raffuse, S., O'Neill, S., & Serre, M. L. (2020). Estimating Wildfire Smoke Concentrations during the October 2017 California Fires through BME Space/Time Data Fusion of Observed, Modeled, and Satellite-Derived PM $_{2.5}$. *Environmental Science & Technology*, *54*(21), 13439–13447. https://doi.org/10.1021/acs.est.0c03761

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J., Myneni, R. B., Sellers, P. J., Pinty, B., & Verstraete, M. M. (1998). Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1072–1087. https://doi.org/10.1109/36.700992

Fa, T., Xie, W., Wang, Y., & Xia, Y. (2019). Development of an all-sky imaging system for cloud cover assessment. *Applied Optics*, *58*(20), 5516. https://doi.org/10.1364/AO.58.005516

Fann, N., Alman, B., Broome, R. A., Morgan, G. G., Johnston, F. H., Pouliot, G., & Rappold, A. G. (2018). The health impacts and economic value of wildland fire episodes in the U.S.: 2008–2012. *Science of The Total Environment*, *610–611*, 802–809. https://doi.org/10.1016/j.scitotenv.2017.08.024

Feng, S., Gao, D., Liao, F., Zhou, F., & Wang, X. (2016). The health effects of ambient PM2.5 and potential mechanisms. *Ecotoxicology and Environmental Safety*, *128*, 67–74. https://doi.org/10.1016/j.ecoenv.2016.01.030

Ford, B., Burke, M., Lassman, W., Pfister, G., & Pierce, J. R. (2017). Status update: Is smoke on your mind? Using social media to assess smoke exposure. *Atmospheric Chemistry and Physics*, *17*(12), 7541–7554. https://doi.org/10.5194/acp-17-7541-2017

Ford, B., & Heald, C. L. (2016). Exploring the uncertainty associated with satellite-based estimates of premature mortality due to exposure to fine particulate matter. *Atmospheric Chemistry and Physics*, *16*(5), 3499–3523. https://doi.org/10.5194/acp-16-3499-2016

Ford, B., Pierce, J. R., Wendt, E., Long, M., Jathar, S., Mehaffy, J., Tryner, J., Quinn, C., van Zyl, L., L'Orange, C., Miller-Lionberg, D., & Volckens, J. (2019a). A low-cost monitor for measurement of fine particulate matter and aerosol optical depth – Part 2: Citizen-science pilot campaign in northern Colorado. *Atmos. Meas. Tech.*, 15.

Ford, B., Pierce, J. R., Wendt, E., Long, M., Jathar, S., Mehaffy, J., Tryner, J., Quinn, C., van Zyl, L., L'Orange, C., Miller-Lionberg, D., & Volckens, J. (2019b). A low-cost monitor for measurement of fine particulate matter and aerosol optical depth – Part 2: Citizen-science pilot campaign in northern Colorado. *Atmospheric Measurement Techniques*, *12*(12), 6385–6399. https://doi.org/10.5194/amt-12-6385-2019

Ford, B., Val Martin, M., Zelasky, S. E., Fischer, E. V., Anenberg, S. C., Heald, C. L., & Pierce, J. R. (2018). Future Fire Impacts on Smoke Concentrations, Visibility, and Health in the Contiguous

United States. *GeoHealth*, *2*(8), 229–247. https://doi.org/10.1029/2018GH000144

Forouzanfar, M. H., Afshin, A., Alexander, L. T., Anderson, H. R., Bhutta, Z. A., Biryukov, S., Brauer, M., Burnett, R., Cercy, K., Charlson, F. J., Cohen, A. J., Dandona, L., Estep, K., Ferrari, A. J., Frostad, J. J., Fullman, N., Gething, P. W., Godwin, W. W., Griswold, M., … Murray, C. J. L. (2016). Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), 1659–1724. https://doi.org/10.1016/S0140-6736(16)31679-8

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., Cropper, M., Ferraro, G., Hanna, J., Hanrahan, D., Hu, H., Hunter, D., Janata, G., Kupka, R., Lanphear, B., … Yan, C. (2022). Pollution and health: A progress update. *The Lancet Planetary Health*, S2542519622000900. https://doi.org/10.1016/S2542-5196(22)00090-0

Gan, R. W., Ford, B., Lassman, W., Pfister, G., Vaidyanathan, A., Fischer, E., Volckens, J., Pierce, J. R., & Magzamen, S. (2017). Comparison of wildfire smoke estimation methods and associations with cardiopulmonary-related hospital admissions: Estimates of Smoke and Health Outcomes. *GeoHealth*, *1*(3), 122–136. https://doi.org/10.1002/2017GH000073

Garay, M. J., Kalashnikova, O. V., & Bull, M. A. (2017). Development and assessment of a higher-spatial-resolution (4.4 km) MISR aerosol optical depth product using AERONET-DRAGON data. *Atmospheric Chemistry and Physics*, *17*(8), 5095–5106. https://doi.org/10.5194/acp-17-5095-2017

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., & Lyapustin, A. I. (2019). Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques*, *12*(1), 169–

209. https://doi.org/10.5194/amt-12-169-2019

Green, M., Kondragunta, S., Ciren, P., & Xu, C. (2009). Comparison of GOES and MODIS Aerosol Optical Depth (AOD) to Aerosol Robotic Network (AERONET) AOD and IMPROVE PM $_{2.5}$ Mass at Bondville, Illinois. *Journal of the Air & Waste Management Association*, *59*(9), 1082–1091. https://doi.org/10.3155/1047-3289.59.9.1082

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guan, Q., Wang, Y., Ping, B., Li, D., Du, J., Qin, Y., Lu, H., Wan, X., & Xiang, J. (2019). Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *Journal of Cancer*, *10*(20), 4876–4882. https://doi.org/10.7150/jca.28769

Gupta, P., Doraiswamy, P., Levy, R., Pikelnaya, O., Maibach, J., Feenstra, B., Polidori, A., Kiros, F., & Mills, K. C. (2018). Impact of California Fires on Local and Regional Air Quality: The Role of a Low-Cost Sensor Network and Satellite Observations. *GeoHealth*, *2*(6), 172–181. https://doi.org/10.1029/2018GH000136

Hammer, M. S., van Donkelaar, A., Li, C., Lyapustin, A., Sayer, A. M., Hsu, N. C., Levy, R. C., Garay, M. J., Kalashnikova, O. V., Kahn, R. A., Brauer, M., Apte, J. S., Henze, D. K., Zhang, L., Zhang, Q., Ford, B., Pierce, J. R., & Martin, R. V. (2020). Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). *Environmental Science & Technology*, *54*(13), 7879–7890. https://doi.org/10.1021/acs.est.0c01764

Henderson, S. B., Brauer, M., MacNab, Y. C., & Kennedy, S. M. (2011). Three Measures of Forest Fire Smoke Exposure and Their Associations with Respiratory and Cardiovascular Health Outcomes in a Population-Based Cohort. *Environmental Health Perspectives*, *119*(9), 1266–1271. https://doi.org/10.1289/ehp.1002288

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., & Smirnov, A. (1998). AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sensing of Environment*, *66*(1), 1–16. https://doi.org/10.1016/S0034-4257(98)00031-5

Holben, B. N., Kim, J., Sano, I., Mukai, S., Eck, T. F., Giles, D. M., Schafer, J. S., Sinyuk, A., Slutsker, I., Smirnov, A., Sorokin, M., Anderson, B. E., Che, H., Choi, M., Crawford, J. H., Ferrare, R. A., Garay, M. J., Jeong, U., Kim, M., … Xiao, Q. (2018). An overview of mesoscale aerosol processes, comparisons, and validation studies from DRAGON networks. *Atmospheric Chemistry and Physics*, *18*(2), 655–671. https://doi.org/10.5194/acp-18-655-2018

Janssen, N. A. H., Fischer, P., Marra, M., Ameling, C., & Cassee, F. R. (2013). Short-term effects of PM2.5, PM10 and PM2.5–10 on daily mortality in the Netherlands. *Science of The Total Environment*, *463–464*, 20–26. https://doi.org/10.1016/j.scitotenv.2013.05.062

Jin, X., Fiore, A. M., Curci, G., Lyapustin, A., Civerolo, K., Ku, M., van Donkelaar, A., & Martin, R. V. (2019). Assessing uncertainties of a geophysical approach to estimate surface fine particulate matter distributions from satellite-observed aerosol optical depth. *Atmospheric Chemistry and Physics*, *19*(1), 295–313. https://doi.org/10.5194/acp-19-295-2019

Johnston, F. H., Henderson, S. B., Chen, Y., Randerson, J. T., Marlier, M., DeFries, R. S., Kinney, P., Bowman, D. M. J. S., & Brauer, M. (2012). Estimated Global Mortality Attributable to Smoke from Landscape Fires. *Environmental Health Perspectives*, *120*(5), 695–701. https://doi.org/10.1289/ehp.1104422

Kaur, T., & Gandhi, T. K. (2019). Automated Brain Image Classification Based on VGG-16 and Transfer Learning. *2019 International Conference on Information Technology (ICIT)*, 94–98. https://doi.org/10.1109/ICIT48102.2019.00023

Kelleher, S., Quinn, C., Miller-Lionberg, D., & Volckens, J. (2018). A low-cost particulate matter (PM_{2.5}) monitor for wildland fire smoke. *Atmospheric Measurement Techniques*, *11*(2), 1087–1097. https://doi.org/10.5194/amt-11-1087-2018

Kelly, K. E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., & Butterfield, A.

(2017). Ambient and laboratory evaluation of a low-cost particulate matter sensor. *Environmental Pollution*, *221*, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039

Kim, I., Lee, K., Lee, S., & Kim, S. D. (2019). Characteristics and health effects of PM2.5 emissions from various sources in Gwangju, South Korea. *Science of The Total Environment*, *696*, 133890. https://doi.org/10.1016/j.scitotenv.2019.133890

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Lassman, W., Ford, B., Gan, R. W., Pfister, G., Magzamen, S., Fischer, E. V., & Pierce, J. R. (2017). Spatial and temporal estimates of population exposure to wildfire smoke during the Washington state 2012 wildfire season using blended model, satellite, and in situ data: Multimethod Estimates of Smoke Exposure. *GeoHealth*, *1*(3), 106–121. https://doi.org/10.1002/2017GH000049

Lee, C. (2020). Impacts of multi-scale urban form on PM2.5 concentrations using continuous surface estimates with high-resolution in U.S. metropolitan areas. *Landscape and Urban Planning*, *204*, 103935. https://doi.org/10.1016/j.landurbplan.2020.103935

Levy Zamora, M., Xiong, F., Gentner, D., Kerkez, B., Kohrman-Glaser, J., & Koehler, K. (2019). Field and Laboratory Evaluations of the Low-Cost Plantower Particulate Matter Sensor. *Environmental Science & Technology*, *53*(2), 838–849. https://doi.org/10.1021/acs.est.8b05174

Li, J., Liu, H., Lv, Z., Zhao, R., Deng, F., Wang, C., Qin, A., & Yang, X. (2018). Estimation of PM2.5 mortality burden in China with new exposure estimation and local concentration-response function. *Environmental Pollution*, *243*, 1710–1718. https://doi.org/10.1016/j.envpol.2018.09.089

Li, J., Zhang, H., Chao, C.-Y., Chien, C.-H., Wu, C.-Y., Luo, C. H., Chen, L.-J., & Biswas, P. (2020). Integrating low-cost air quality sensor networks with fixed and satellite monitoring systems to study ground-level PM2.5. *Atmospheric Environment*, *223*, 117293. https://doi.org/10.1016/j.atmosenv.2020.117293

Lin, C., Labzovskii, L. D., Leung Mak, H. W., Fung, J. C. H., Lau, A. K. H., Kenea, S. T., Bilal, M., Vande Hey, J. D., Lu, X., & Ma, J. (2020). Observation of PM2.5 using a combination of satellite remote sensing and low-cost sensor network in Siberian urban areas with limited reference monitoring. *Atmospheric Environment*, *227*, 117410. https://doi.org/10.1016/j.atmosenv.2020.117410

Liu, J. C., Pereira, G., Uhl, S. A., Bravo, M. A., & Bell, M. L. (2015). A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. *Environmental Research*, *136*, 120–132. https://doi.org/10.1016/j.envres.2014.10.015

Liu, S., Li, M., Zhang, Z., Xiao, B., & Cao, X. (2018). Multimodal Ground-Based Cloud Classification Using Joint Fusion Convolutional Neural Network. *Remote Sensing*, *10*(6), 822. https://doi.org/10.3390/rs10060822

Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., & Koutrakis, P. (2005). Estimating Ground-Level PM $_{2.5}$ in the Eastern United States Using Satellite Remote Sensing. *Environmental Science & Technology*, *39*(9), 3269–3278. https://doi.org/10.1021/es049352m

Long, C. N., Sabburg, J. M., Calbó, J., & Pagès, D. (2006). Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images. *Journal of Atmospheric and Oceanic Technology*, *23*(5), 633–652. https://doi.org/10.1175/JTECH1875.1

Lu, X., Lin, C., Li, W., Chen, Y., Huang, Y., Fung, J. C. H., & Lau, A. K. H. (2019). Analysis of the adverse health effects of PM2.5 from 2001 to 2017 in China and the role of urbanization in aggravating the health burden. *Science of The Total Environment*, *652*, 683–695. https://doi.org/10.1016/j.scitotenv.2018.10.140

Lu, Y., Giuliano, G., & Habre, R. (2021). Estimating hourly PM2.5 concentrations at the neighborhood scale using a low-cost air sensor network: A Los Angeles case study. *Environmental Research*, *195*, 110653. https://doi.org/10.1016/j.envres.2020.110653

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Zhang, H., Aamaas, B., Boucher, O., Dalsøren, S. B., Daniel, J. S., Forster, P., … Shine, K. (2013). *8 Anthropogenic and Natural Radiative Forcing*. 82.

Neumann, J. E., Amend, M., Anenberg, S., Kinney, P. L., Sarofim, M., Martinich, J., Lukens, J., Xu, J.-W., & Roman, H. (2021). Estimating PM2.5-related premature mortality and morbidity associated with future wildfire emissions in the western US. *Environmental Research Letters*, *16*(3), 035019. https://doi.org/10.1088/1748-9326/abe82b

O'Dell, K., Bilsback, K., Ford, B., Martenies, S. E., Magzamen, S., Fischer, E. V., & Pierce, J. R. (2021). Estimated Mortality and Morbidity Attributable to Smoke Plumes in the United States: Not Just a Western US Problem. *GeoHealth*, *5*(9). https://doi.org/10.1029/2021GH000457

O'Dell, K., Ford, B., Fischer, E. V., & Pierce, J. R. (2019). Contribution of Wildland-Fire Smoke to US $PM_{2.5}$ and Its Influence on Recent Trends. *Environmental Science & Technology*, *53*(4), 1797–1804. https://doi.org/10.1021/acs.est.8b05430

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pillarisetti, A., Carter, E., Rajkumar, S., Young, B. N., Benka-Coker, M. L., Peel, J. L., Johnson, M., & Clark, M. L. (2019). Measuring personal exposure to fine particulate matter (PM2.5) among rural Honduran women: A field evaluation of the Ultrasonic Personal Aerosol Sampler (UPAS). *Environment International*, *123*, 50–53. https://doi.org/10.1016/j.envint.2018.11.014

Pope, C. A., & Dockery, D. W. (2006). Health Effects of Fine Particulate Air Pollution: Lines that Connect. *Journal of the Air & Waste Management Association*, *56*(6), 709–742. https://doi.org/10.1080/10473289.2006.10464485

Reid, C. E., Brauer, M., Johnston, F. H., Jerrett, M., Balmes, J. R., & Elliott, C. T. (2016). Critical Review of Health Impacts of Wildfire Smoke Exposure. *Environmental Health Perspectives*, *124*(9), 1334–1343. https://doi.org/10.1289/ehp.1409277

Reid, C. E., Jerrett, M., Petersen, M. L., Pfister, G. G., Morefield, P. E., Tager, I. B., Raffuse, S. M., & Balmes, J. R. (2015). Spatiotemporal Prediction of Fine Particulate Matter During the 2008 Northern California Wildfires Using Machine Learning. *Environmental Science & Technology*, *49*(6), 3887–3896. https://doi.org/10.1021/es505846r

Ruminski, M., Kondragunta, S., Draxler, R., Zeng, J., & Rm, A. (2006). *Recent Changes to the Hazard Mapping System*. 16.

Salomonson, V. V., Barnes, W. L., Maymon, P. W., Montgomery, H. E., & Ostrow, H. (1989). MODIS: Advanced facility instrument for studies of the Earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, *27*(2), 145–153. https://doi.org/10.1109/36.20292

Sayahi, T., Butterfield, A., & Kelly, K. E. (2019). Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environmental Pollution*, *245*, 932–940. https://doi.org/10.1016/j.envpol.2018.11.065

Sayer, A. M., Hsu, N. C., Bettenhausen, C., Jeong, M.-J., Holben, B. N., & Zhang, J. (2012). Global and regional evaluation of over-land spectral aerosol optical depth retrievals from SeaWiFS. *Atmospheric Measurement Techniques*, *5*(7), 1761–1778. https://doi.org/10.5194/amt-5-1761-2012

Shi, C., Wang, C., Wang, Y., & Xiao, B. (2017). Deep Convolutional Activations-Based Features for Ground-Based Cloud Classification. *IEEE Geoscience and Remote Sensing Letters*, *14*(6), 816–820. https://doi.org/10.1109/LGRS.2017.2681658

Shuang Liu, Chunheng Wang, Baihua Xiao, Zhong Zhang, & Xiaozhong Cao. (2013). Tensor Ensemble of Ground-Based Cloud Sequences: Its Modeling, Classification, and Synthesis. *IEEE Geoscience and Remote Sensing Letters*, *10*(5), 1190–1194. https://doi.org/10.1109/LGRS.2012.2236073

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv:1409.1556 [Cs]*. http://arxiv.org/abs/1409.1556

Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., & Slutsker, I. (2000). Cloud-Screening and Quality Control Algorithms for the AERONET Database. *Remote Sensing of Environment*, *73*(3), 337–349. https://doi.org/10.1016/S0034-4257(00)00109-7

Snider, G., Weagle, C. L., Martin, R. V., van Donkelaar, A., Conrad, K., Cunningham, D., Gordon, C., Zwicker, M., Akoshile, C., Artaxo, P., Anh, N. X., Brook, J., Dong, J., Garland, R. M., Greenwald, R., Griffith, D., He, K., Holben, B. N., Kahn, R., … Liu, Y. (2015). SPARTAN: A global network to evaluate and enhance satellite-based estimates of ground-level particulate matter for global health applications. *Atmospheric Measurement Techniques*, *8*(1), 505–521. https://doi.org/10.5194/amt-8-505-2015

Sorek-Hamer, M., Franklin, M., Chau, K., Garay, M., & Kalashnikova, O. (2020). Spatiotemporal Characteristics of the Association between AOD and PM over the California Central Valley. *Remote Sensing*, *12*(4), 685. https://doi.org/10.3390/rs12040685

Suzuki, S., & be, K. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, *30*(1), 32–46. https://doi.org/10.1016/0734-189X(85)90016-7

Tammina, S. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(10), p9420. https://doi.org/10.29322/IJSRP.9.10.2019.p9420

Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 839–846. https://doi.org/10.1109/ICCV.1998.710815

Tryner, J., L'Orange, C., Mehaffy, J., Miller-Lionberg, D., Hofstetter, J. C., Wilson, A., & Volckens, J. (2020). Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers. *Atmospheric Environment*, *220*, 117067. https://doi.org/10.1016/j.atmosenv.2019.117067

UCAR/NCAR - Earth Observing Laboratory. (2016). *METCRAX-II ISS All Sky Camera Imagery. 1.0*.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., & Winker, D. M. (2016). Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science & Technology*, *50*(7), 3762–3772. https://doi.org/10.1021/acs.est.5b05833

van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J.

(2010). Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. *Environmental Health Perspectives*, *118*(6), 847–855. https://doi.org/10.1289/ehp.0901623

van Donkelaar, A., Martin, R. V., Li, C., & Burnett, R. T. (2019). Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. *Environmental Science & Technology*, *53*(5), 2595–2611. https://doi.org/10.1021/acs.est.8b06392

van Donkelaar, A., Martin, R. V., & Park, R. J. (2006). Estimating ground-level PM $_{2.5}$ using aerosol optical depth determined from satellite remote sensing. *Journal of Geophysical Research*, *111*(D21), D21201. https://doi.org/10.1029/2005JD006996

van Donkelaar, A., Martin, R. V., Pasch, A. N., Szykman, J. J., Zhang, L., Wang, Y. X., & Chen, D. (2012). Improving the Accuracy of Daily Satellite-Derived Ground-Level Fine Aerosol Concentration Estimates for North America. *Environmental Science & Technology*, *46*(21), 11971–11978. https://doi.org/10.1021/es3025319

van Donkelaar, A., Martin, R. V., Spurr, R. J. D., Drury, E., Remer, L. A., Levy, R. C., & Wang, J. (2013). Optimal estimation for global ground-level fine particulate matter concentrations: OPTIMAL ESTIMATION FOR PM2.5. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5621–5636. https://doi.org/10.1002/jgrd.50479

Vohra, K., Vodonos, A., Schwartz, J., Marais, E. A., Sulprizio, M. P., & Mickley, L. J. (2021). Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem. *Environmental Research*, *195*, 110754. https://doi.org/10.1016/j.envres.2021.110754

Volckens, J., Quinn, C., Leith, D., Mehaffy, J., Henry, C. S., & Miller-Lionberg, D. (2017). Development and evaluation of an ultrasonic personal aerosol sampler. *Indoor Air*, *27*(2), 409–416. https://doi.org/10.1111/ina.12318

Wendt, E. A., Quinn, C., L'Orange, C., Miller-Lionberg, D. D., Ford, B., Pierce, J. R., Mehaffy, J., Cheeseman, M., Jathar, S. H., Hagan, D. H., Rosen, Z., Long, M., & Volckens, J. (2021). A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth – Part 3: Automation and design improvements. *Atmospheric Measurement Techniques*, *14*(9), 6023–6038. https://doi.org/10.5194/amt-14-6023-2021

Wendt, E. A., Quinn, C. W., Miller-Lionberg, D. D., Tryner, J., L'Orange, C., Ford, B., Yalin, A. P., Pierce, J. R., Jathar, S., & Volckens, J. (2019a). A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth – Part 1: Specifications and testing. *Atmospheric Measurement Techniques*, *12*(10), 5431–5441. https://doi.org/10.5194/amt-12-5431-2019

Wendt, E. A., Quinn, C. W., Miller-Lionberg, D. D., Tryner, J., L'Orange, C., Ford, B., Yalin, A. P., Pierce, J. R., Jathar, S., & Volckens, J. (2019b). A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth – Part 1: Specifications and testing. *Atmospheric Measurement Techniques*, *12*(10), 5431–5441. https://doi.org/10.5194/amt-12-5431-2019

Xia, M., Lu, W., Yang, J., Ma, Y., Yao, W., & Zheng, Z. (2015). A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image. *Neurocomputing*, *160*, 238–249. https://doi.org/10.1016/j.neucom.2015.02.022

Xie, W., Liu, D., Yang, M., Chen, S., Wang, B., Wang, Z., Xia, Y., Liu, Y., Wang, Y., & Zhang, C. (2020). SegCloud: A novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation. *Atmospheric Measurement Techniques*, *13*(4), 1953–1961. https://doi.org/10.5194/amt-13-1953-2020

Young, A. T. (1994). Air mass and refraction. *Applied Optics*, *33*(6), 1108. https://doi.org/10.1364/AO.33.001108

Zhang, J., Liu, P., Zhang, F., & Song, Q. (2018). CloudNet: Ground-Based Cloud Classification With Deep Convolutional Neural Network. *Geophysical Research Letters*, *45*(16), 8665–8672. https://doi.org/10.1029/2018GL077787

Zheng, T., Bergin, M. H., Johnson, K. K., Tripathi, S. N., Shirodkar, S., Landis, M. S., Sutaria, R., & Carlson, D. E. (2018). Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. *Atmospheric Measurement Techniques*, *11*(8), 4823–4846. https://doi.org/10.5194/amt-11-4823-2018

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, *109*(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555

APPENDIX A

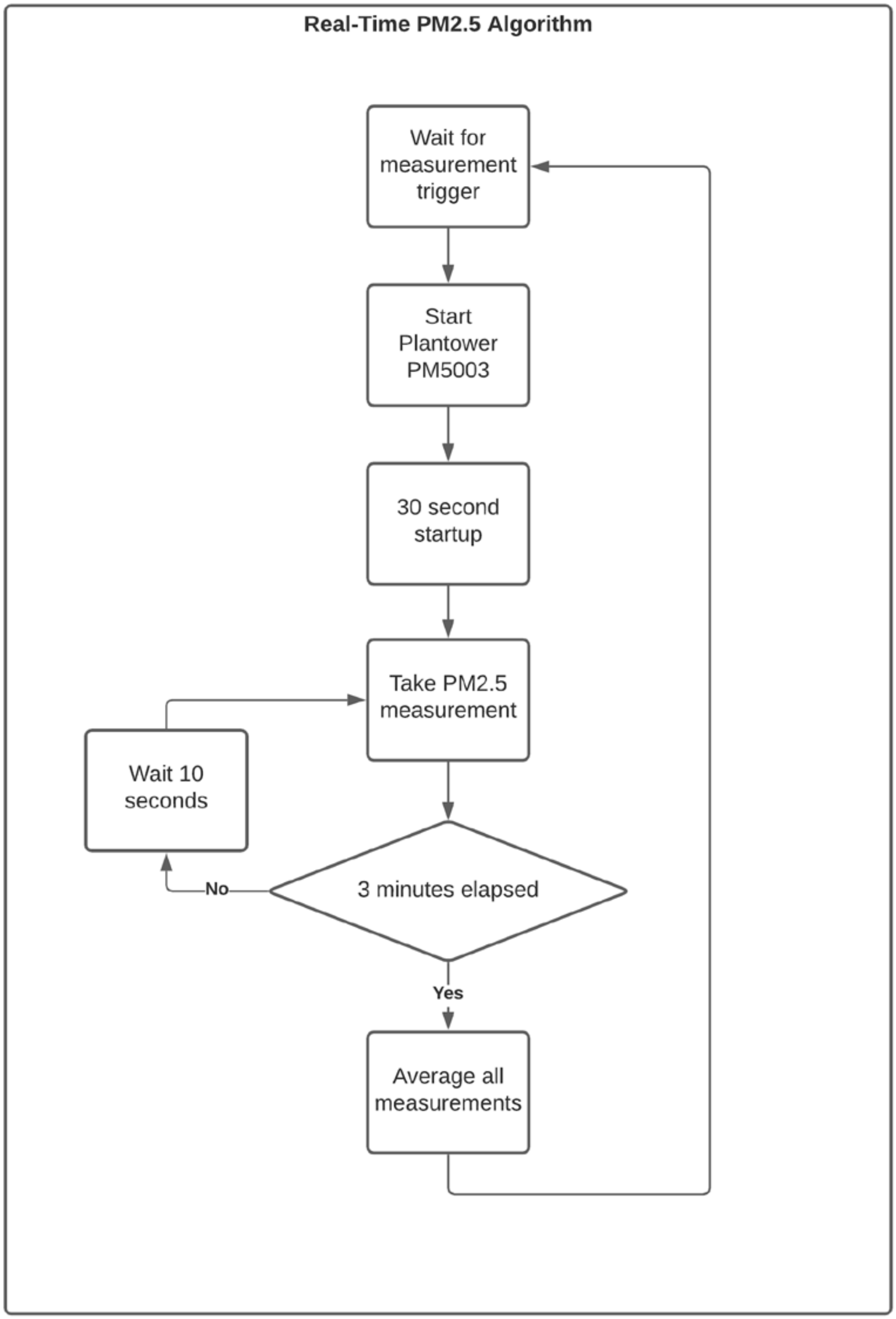

Figure A1: Overview of real-time $PM_{2.5}$ measurement protocol.

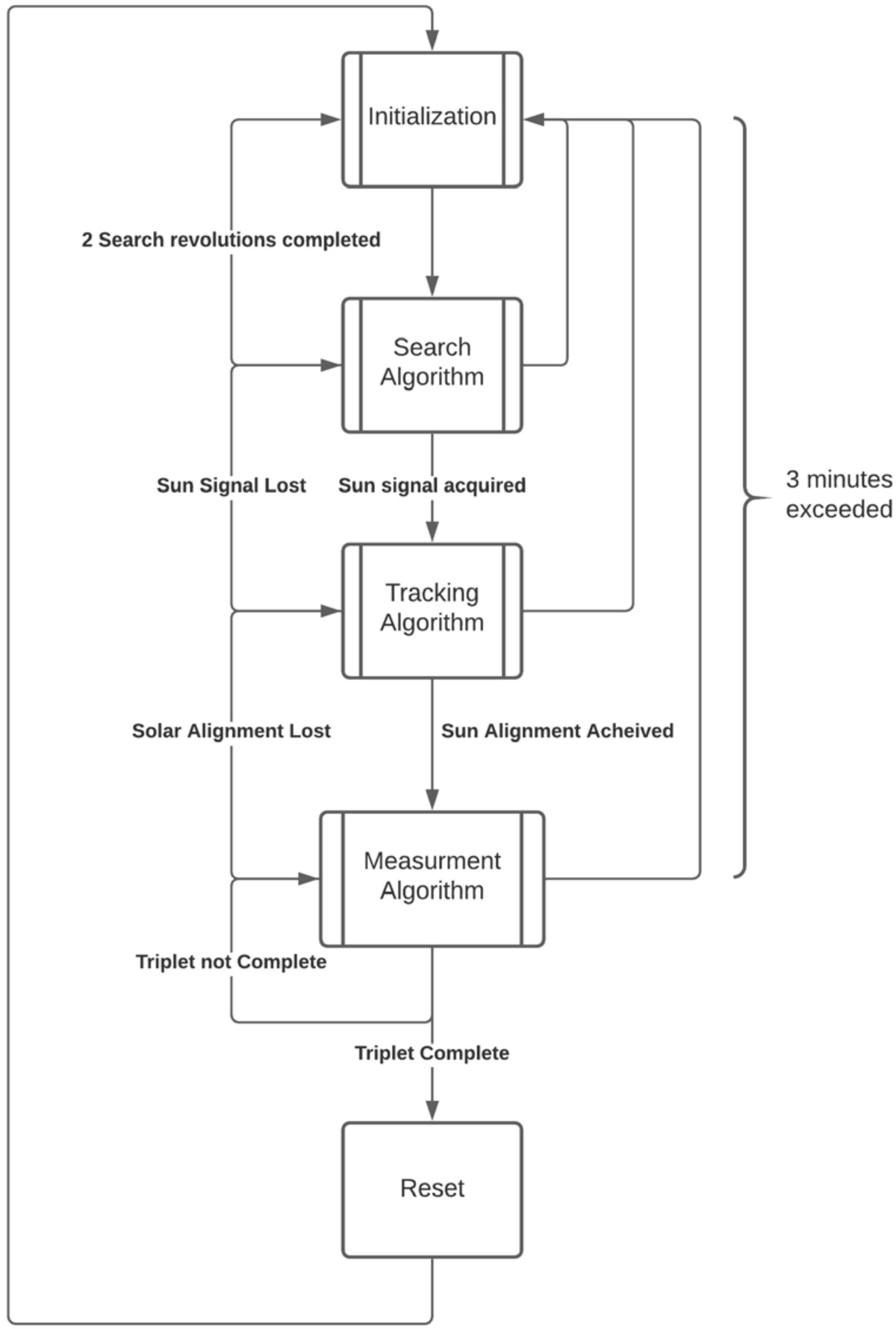

Figure A2: Overview of AOD measurement protocol. Initialization, search, tracking, and measurement algorithms are detailed in Figs. A2-A5.

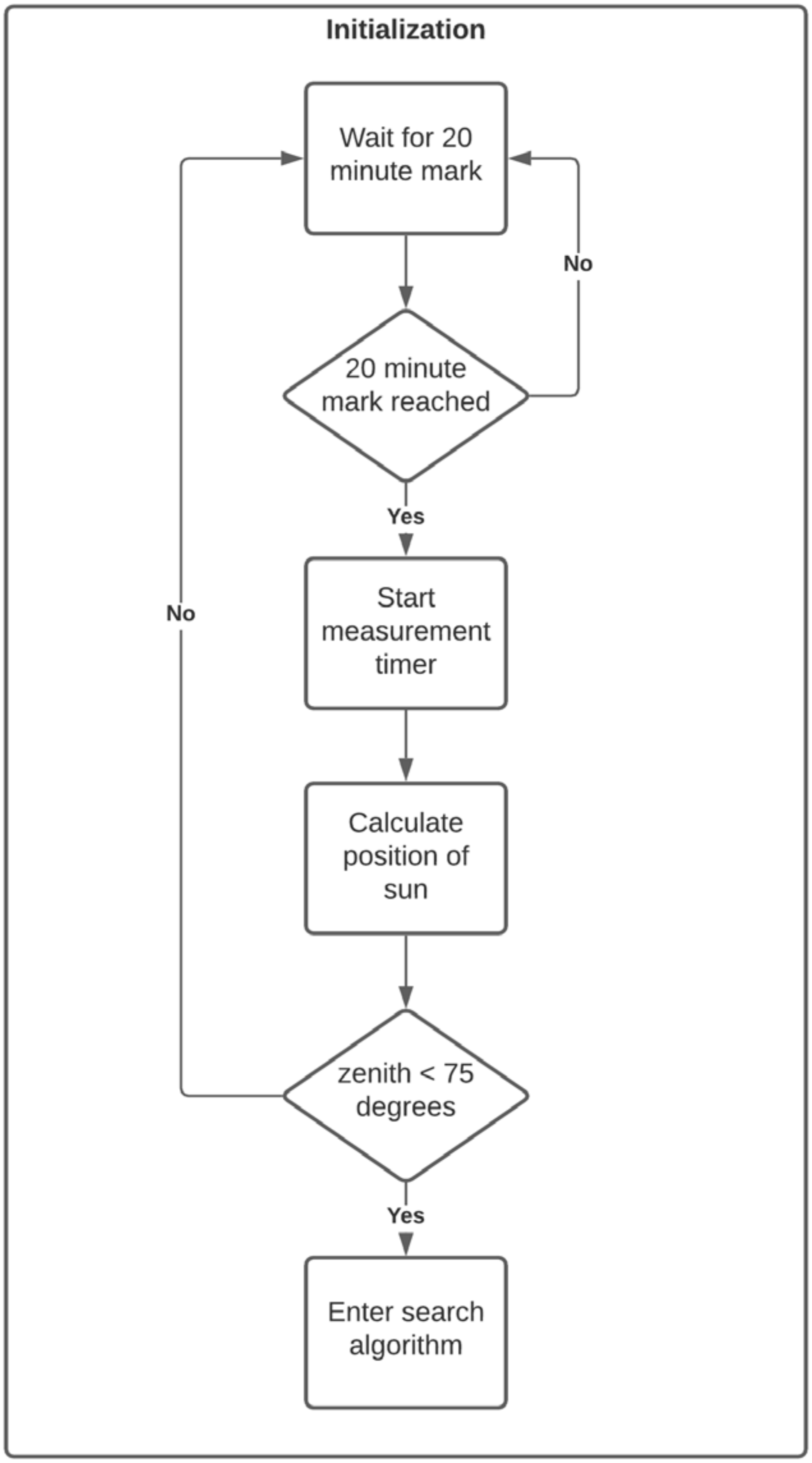

Figure A3: AOD subsystem initialization protocol.

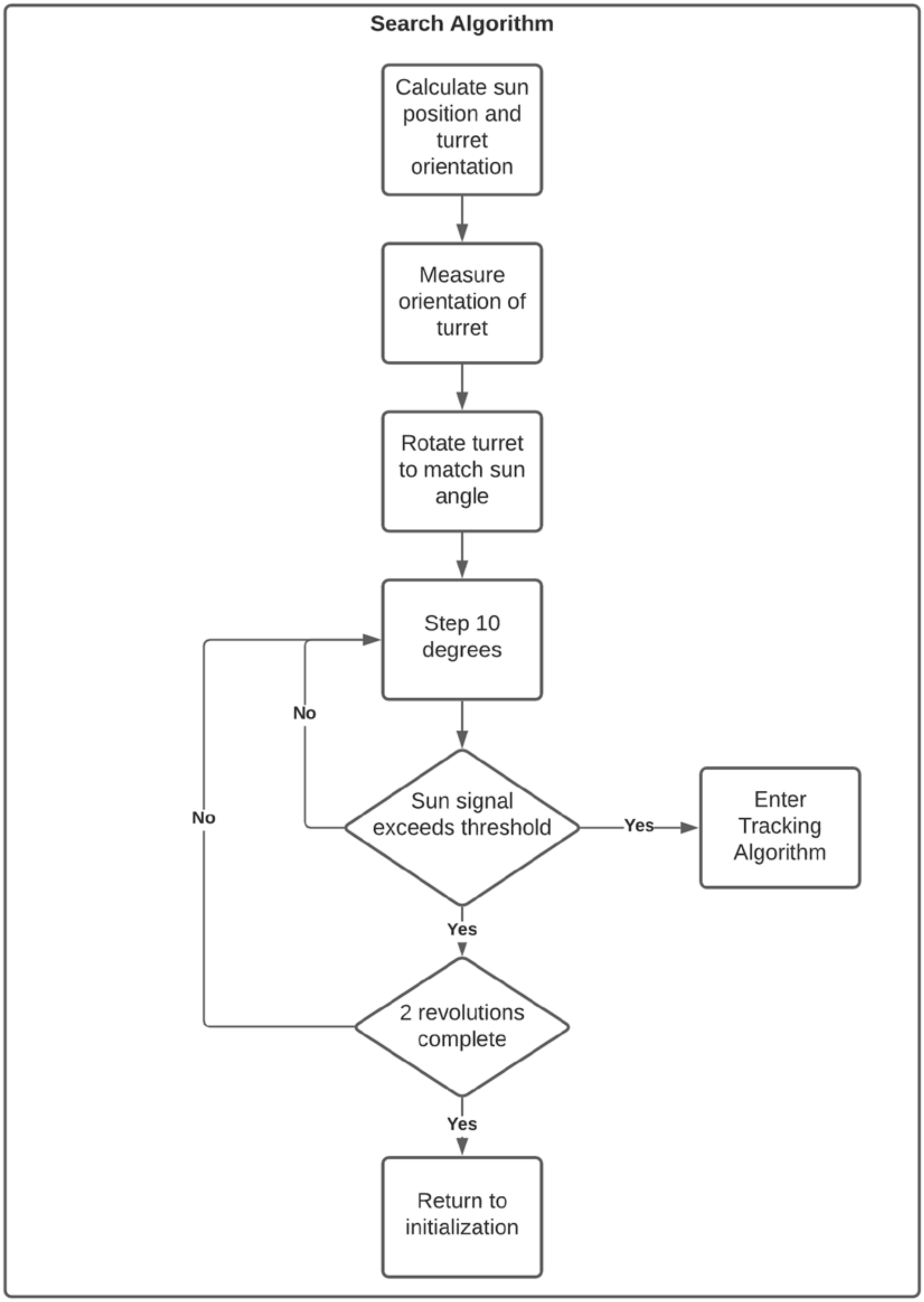

Figure A4: AOD subsystem search algorithm.

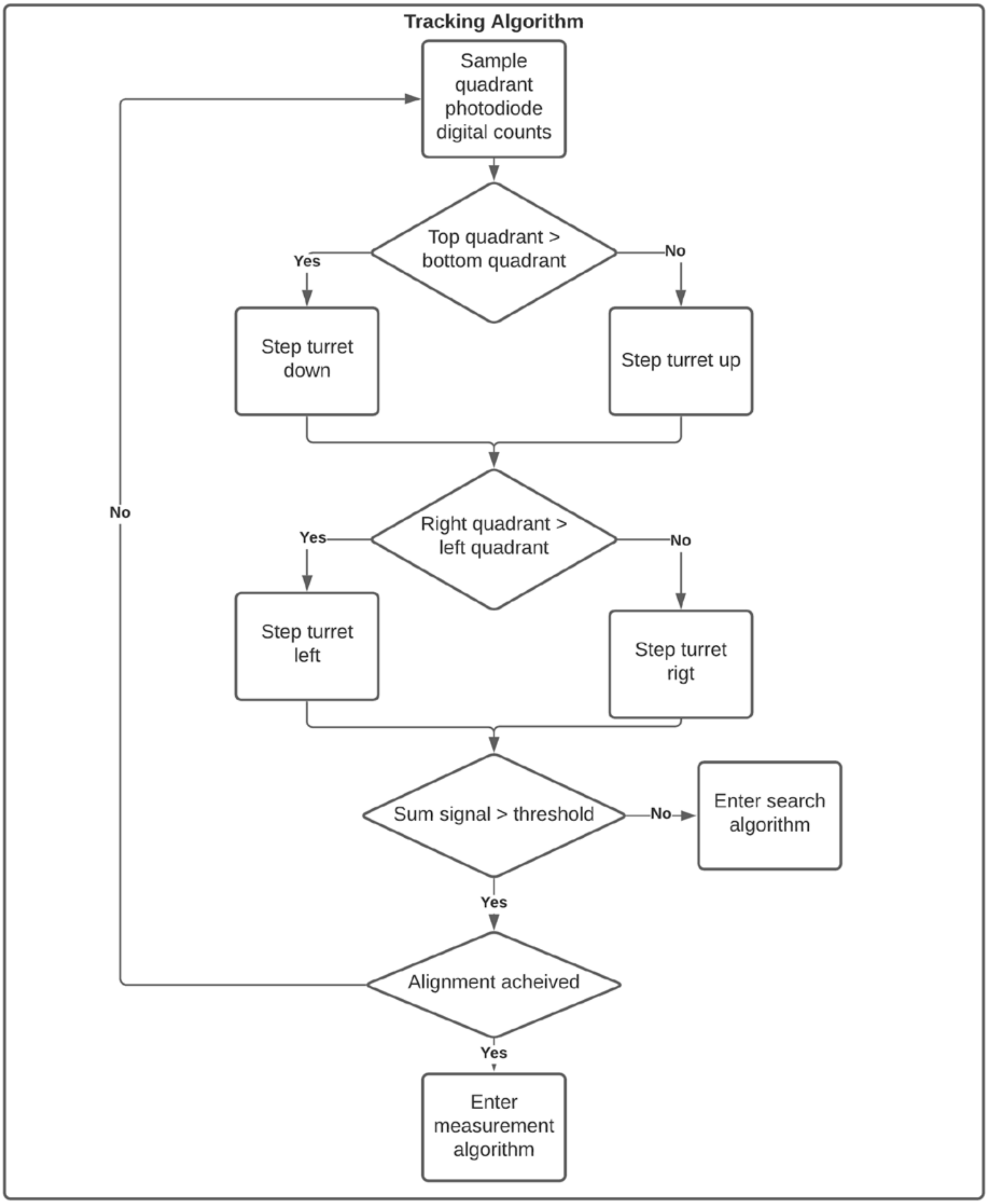

Figure A5: AOD subsystem tracking algorithm

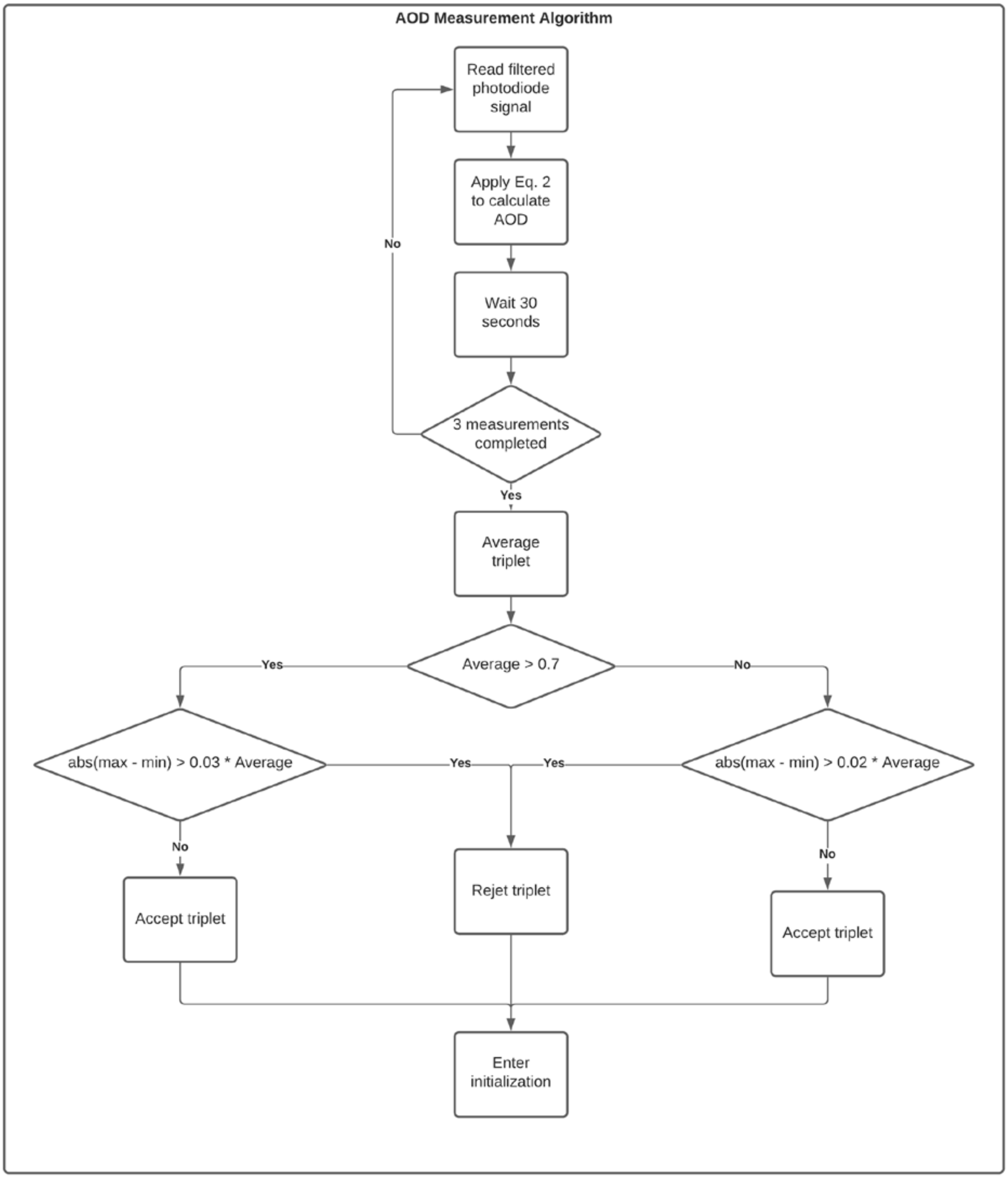

Figure A6: AOD subsystem measurement algorithm

Table A1: AMODv2 validation summary statistics calculated separately for elevated-AOD days and clear days. Elevated-AOD days were defined as days when the average AERONET AOD at 500 nm was greater than or equal to 0.15. Clear days were defined as days in which the average AOD was less than 0.15. In total, five days were identified as clear and four days were identified as elevated-AOD.

Wavelength (nm)	AERONET 500 nm AOD magnitude	Number of samples	Mean absolute error (AOD)	AOD Precision (AOD)
440	All included	426	0.04	0.02
500	All included	426	0.06	0.02
675	All included	426	0.03	0.01
870	All included	426	0.03	0.02
440	> 0.15 (elevated)	115	0.05	0.02
500	> 0.15 (elevated)	115	0.05	0.02
675	> 0.15 (elevated)	115	0.03	0.01
870	> 0.15 (elevated)	115	0.03	0.01
440	< 0.15 (clear)	311	0.04	0.02
500	< 0.15 (clear)	311	0.06	0.03
675	< 0.15 (clear)	311	0.02	0.01
870	< 0.15 (clear)	311	0.03	0.02

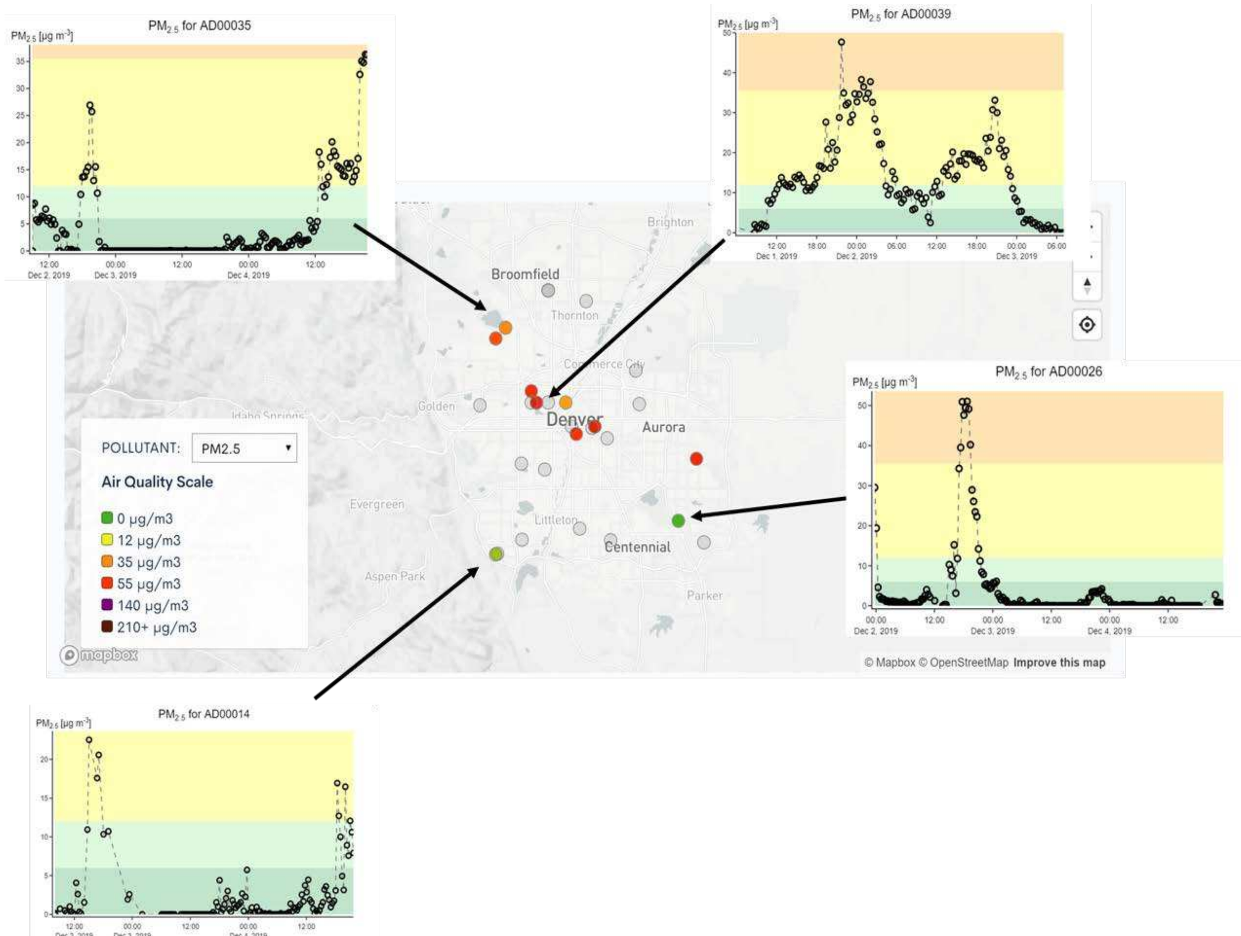

Figure A7: Example live map from project website csu-ceams.com overlaid with time series of $PM_{2.5}$ from selected units. This snapshot was taken at a time when AMODv2 units were located at different locations in Colorado for test deployments, for purposes of illustrating the web interface. Colored circles represent active AMODv2s. Grey circles represent inactive AMODv2 units. Inactive units are either charging between samples or have been sent back from the testing site. The color scale is determined by the current Air Quality Index (AQI) calculated based on the $PM_{2.5}$ measurement. The four sample $PM_{2.5}$ time series plots are linked to specific participant locations with arrows. Time series plots can be accessed by clicking on an active circle. Users may select the option to view AOD from a drop-down menu for both the map and the time series plot. Note: that this figure has been edited to show map and time series plots on the same page. On the actual website selecting a point displays only one simplified time series on the map itself. Detailed time series shown here are available on a separate page which can be accessed through selecting a unit on the map.

Here we present results from a sample deployment of 10 units. We configured the units to sample for approximately 60 hours. The 10 units were co-located and sampled simultaneously. We collected and analysed real-time PM2.5 mass concentrations, AOD, PM2.5 to AOD ratio, meteorological data, and quality control data. In Fig. A8, we provide real-time AOD at 500 nm, real-time $PM_{2.5}$, and the corresponding $PM_{2.5}$ to AOD ratios.

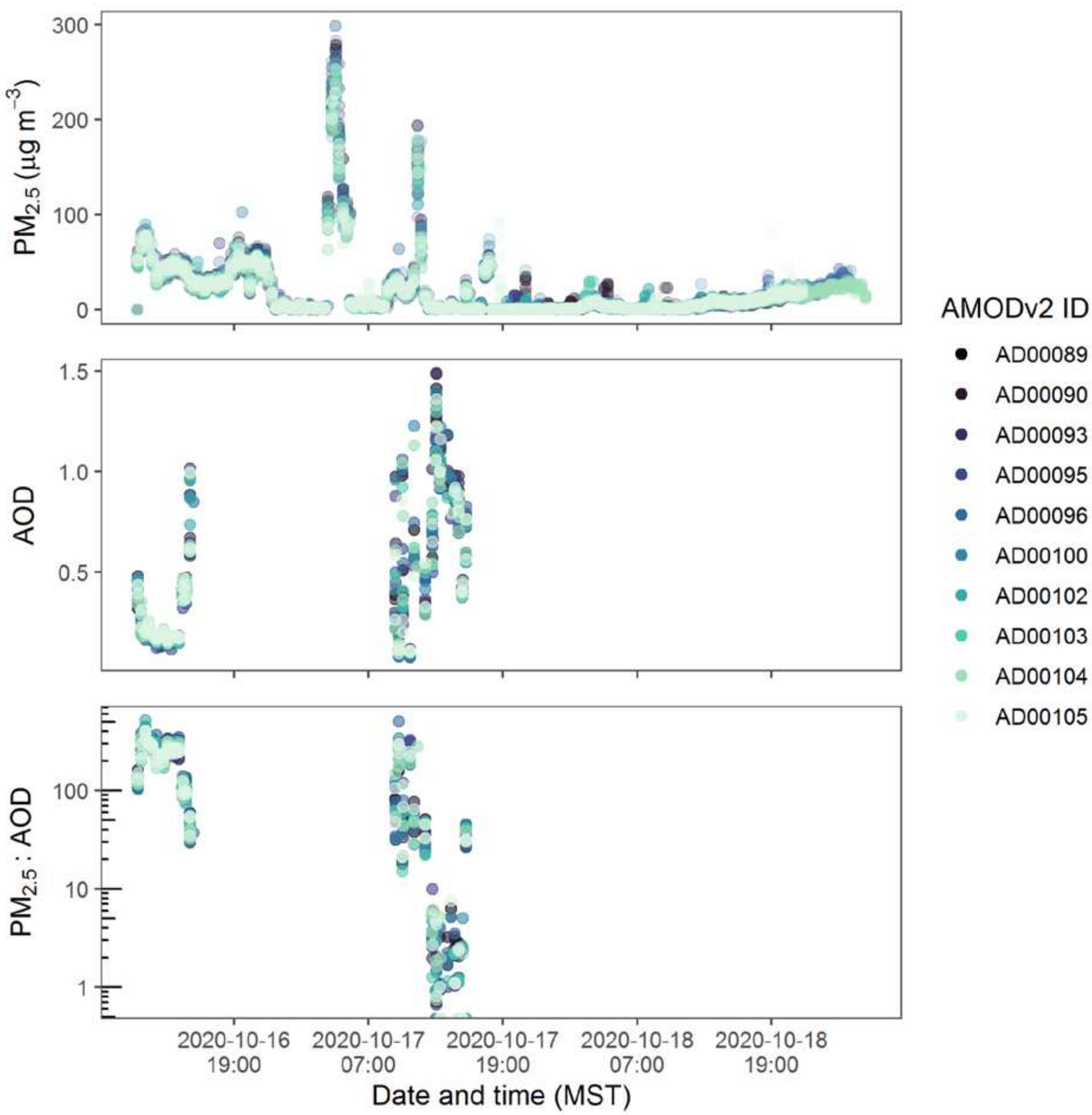

Figure A8: Time series from 10 co-located AMODv2s featuring $PM_{2.5}$ concentration, AOD at 500 nm, and $PM_{2.5}$ to AOD (at 500nm) ratio for 17-19 October 2020 in MST. $PM_{2.5}$ measurements are from the Plantower PMS5003 and are the CF = 1 values. These values have not been corrected relative to the filter mass concentrations. Points are colored according to the AMODv2 ID. Note the vertical axis for $PM_{2.5}$:AOD is provided in a logarithmic scale to clarify lower values indicative of lofted smoke.

In Fig. A9, we provide detailed results from a single unit including 4-channel AOD, $PM_{2.5}$, and meteorological data including temperature, pressure, and relative humidity. This sample deployment highlighted several important strengths of the AMODv2 relative to AMODv1 and other prior sampling approaches. The high temporal resolution of AOD and $PM_{2.5}$ measurements facilitated a more complete understanding of the air pollution event occurring during the sample. With the AMODv2, we observed moderate air pollution at the start of the sample on the afternoon of 16 October 2020, with all units reporting consistent values for AOD (>0.30 ± 0.06) and $PM_{2.5}$ (50 ± 20.0 to 100 ± 40.0 µg m^{-3}). This was followed by increases on 17 October 2020 to severe levels (AOD up to 1.5 ± 0.06 and $PM_{2.5}$ up to 300 ± 66.2 µg m^{-3}) as wildfire smoke swept over the city in the afternoon and gradually subsided over the course of 18 October 2020. We observed reductions in $PM_{2.5}$:AOD (<10) as ground level $PM_{2.5}$ decreased to moderate and mild levels (<20 ± 2.0 µg m^{-3}), while the AOD remained elevated (>0.50 ± 0.06) due to the presence of lofted smoke. We then noted the continuation of the trend at ground level with the further reduction of ground-level $PM_{2.5}$ on 19 October 2020 (5 ± 2.0 to 15 ± 2.0 µg m^{-3}). Cloud cover prevented additional AOD measurements on 19th October, which was automatically screened for using the cloud screening algorithm. The meteorological data was also consistent with cloud cover with lower temperatures and elevated relative humidity reported on that day (Fig. A9).

Data from the sample deployment were accessed from our companion website (csu-ceams.com) in real time. With AOD, $PM_{2.5}$ and $PM_{2.5}$:AOD reported every 20 minutes throughout the sample to the website, we could assess the progress of wildfire smoke in Fort Collins remotely in real time. This was not possible with AMODv1, which lacked wireless transmission capabilities. In terms of scalability, the AMODv2 was relatively easy to deploy and maintain owing to its compact design, coupled with its automated measurement protocols. In the sample test, we were able to quickly prepare and deploy units in response to wildfire activity.

We leveraged the data accessibility features of AMODv2 for real-time quality control of incoming sample data. We monitored sample flow rate and total sampled volume to detect potential errors with the gravimetric sample collection. We monitored battery temperature to detect potential overheating of the unit, allowing proper intervention (e.g. temporarily moving the unit into shade) before the instrument reaches a shutoff threshold. We used battery voltage, battery state of charge, and current draw data to identify units unlikely to complete the intended sample duration. Current draw data was also used to identify when the tracking motors were engaged, indicating an attempted AOD measurement at the expected time. Wireless signal strength data were used to identify units with relatively poor connection and move them into areas with better signal. In the sample deployment detailed here, no interventions based on quality control data were warranted. However, in general, these data can be used to remotely identify and address malfunctioning units mid-sample. This feature represents a substantial improvement compared with AMODv1, which provided no sample quality control data in real time, requiring manual data acquisition (via micro SD card) and unit inspection following a failed sample.

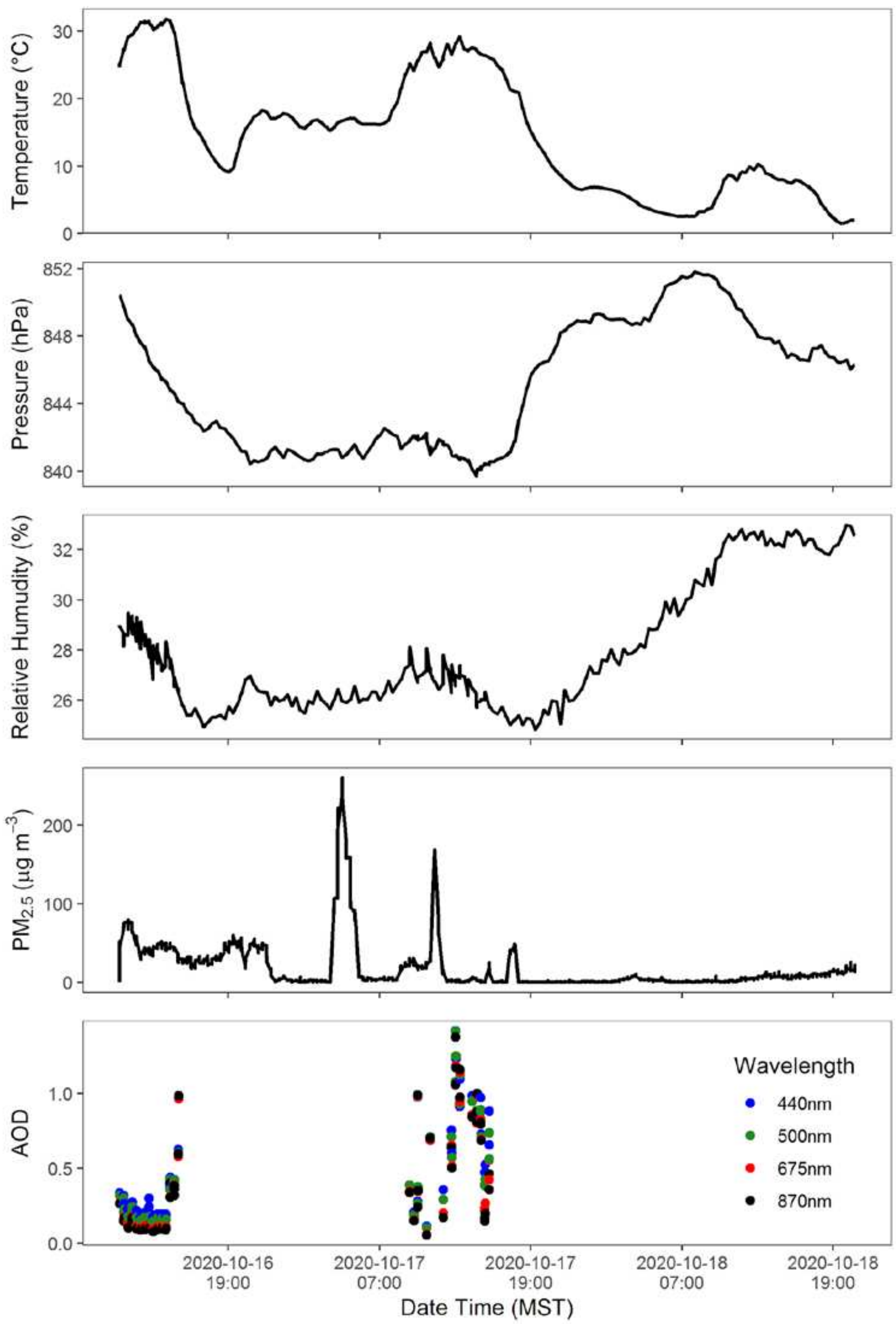

Figure A9: Sample time series from completed AMODv2 sample in MST. Temperature, pressure, relative humidity, and $PM_{2.5}$ reported at 30 second intervals are provided in the top four panels. The bottom panel gives screened AOD measurements, reported at 20 minute intervals. The presence of wildfire smoke on October 17 corresponded with increases in $PM_{2.5}$ and AOD.

Table A2: AMODv2 Cost of Goods and Assembly Summary. Costs tabulated here are for a production run of 100 units.

Component	Manufacturer	Part Number	Cost (USD)
Printed Circuit Boards	Vergent Engineering	Custom Parts	400
440 nm Filtered Photodiode	Intor	Custom Parts	28
520 nm Filtered Photodiode	Intor	Custom Parts	26
680 nm Filtered Photodiode	Intor	Custom Parts	26
870 nm Filtered Photodiode	Intor	Custom Parts	28
Light-Scattering $PM_{2.5}$ Sensor	Plantower	PMS5003	15
Solar Alignment Sensor	Solar MEMS	NANO-ISS5	45
Electrical Box	Polycase	Custom Part	55
3D Printed Fixtures	GoProto	Custom Part	67
Cyclone and Inlet	Synergy Core	Custom Part	74
Battery Pack	Dakota LithiumBatteries	12V 10AH LiFePO4	63
Auxiliary Battery Pack	Battery Space	LFH4S4R1WR-C5	68
Zenith Stepper Motor	Stepper Online	17HS10-0704S-C2	7
Azimuth Stepper Motor	Stepper Online	17HS19-1684S-C6	8

Misc. Housing Components	N/A	N/A	25
Assembly Labor	N/A	N/A	240
Total Costs			1175

APPENDIX B

Text B1: Summary of CEAMS quality control results

Of the 192 sample logs received at the conclusion of the study sample logs, 140 (72.9%) completed at least 76.8 hours (75%) of the prescribed 96-hour runtime. Of the prematurely terminated samples, 30 (57.7%) had a depleted battery, 14 (26.9%) were manually turned off by the operator via the pushbutton, and 8 (15.4%) failed due to an unknown electrical or mechanical error. In Fig. 2, we provide the distribution of CEAMS runtimes, colored by the failure mode.

Compared with our prior laboratory reliability testing (Wendt et al., 2021), fewer AMODv2s completed their full runtime. In our laboratory testing (Wendt et al., 2021), 75.0% of the samples completed their prescribed runtime, compared with 72.9% in the CEAMS network. However, relatively similar proportions of samplers failed due to battery depletion (64.0% for laboratory and 57.7% for CEAMS) and unspecified electrical or mechanical failure (9% for laboratory and 15.4% for CEAMS). The difference in overall performance can then be partially explained by user-initiated pushbutton shutdowns, which did not occur in the laboratory testing. We received some explanations for why some samples were manually shut down prematurely including concern over inclement weather, concern over theft from relatively insecure positions, and choosing to move the device to a new location mid-sample. Further study on participant interactions with the samplers is the subject of ongoing work.

We applied additional quality control analyses to specific AOD and $PM_{2.5}$ measurements. With respect to AOD, the AMODv2 reported either a triplet average of AOD measurements or a unique error code at every 20-minute mark throughout the day. Unique error codes were provided for measurements initiated at night (no execution of tracking protocol), those that never achieved alignment with the sun, thoses with incomplete triplets, and those identified as cloud-contaminated. Across all wavelengths, 981 AOD measurement attempts did not complete a full triplet of measurements. The number of triplets screened for clouds and those marked as valid varied by wavelength. There were 1,915; 1,813; 1,892; and 2,490 triplets marked as cloud-contaminated for 440 nm, 500 nm, 675 nm, and 870 nm channels respectively. There were 3,658; 3,760; 3,681; and 3,083 triplets marked as valid for 440 nm, 500 nm, 675 nm, and 870 nm channels; respectively. We accounted for discrepancies with the 870 nm channel by including triplets that were valid for the remaining three wavelengths.

Compared with our prior laboratory validation work (Wendt et al., 2021), a higher proportion of measurement triplets were marked as valid for all wavelengths in our crowdsourced CEAMS campaign. In our prior study, 33.0%, 34.0%, 35.0%, and 33.0% of triplet attempts were marked as valid for 440 nm, 500 nm, 675 nm, and 870 nm channels, respectively (Wendt et al., 2021). In the present study, 55.8%, 57.4%, 56.2%, and 47.0% of triplet attempts were marked as valid for 440 nm, 500 nm, 675 nm, and 870 nm channels, respectively.

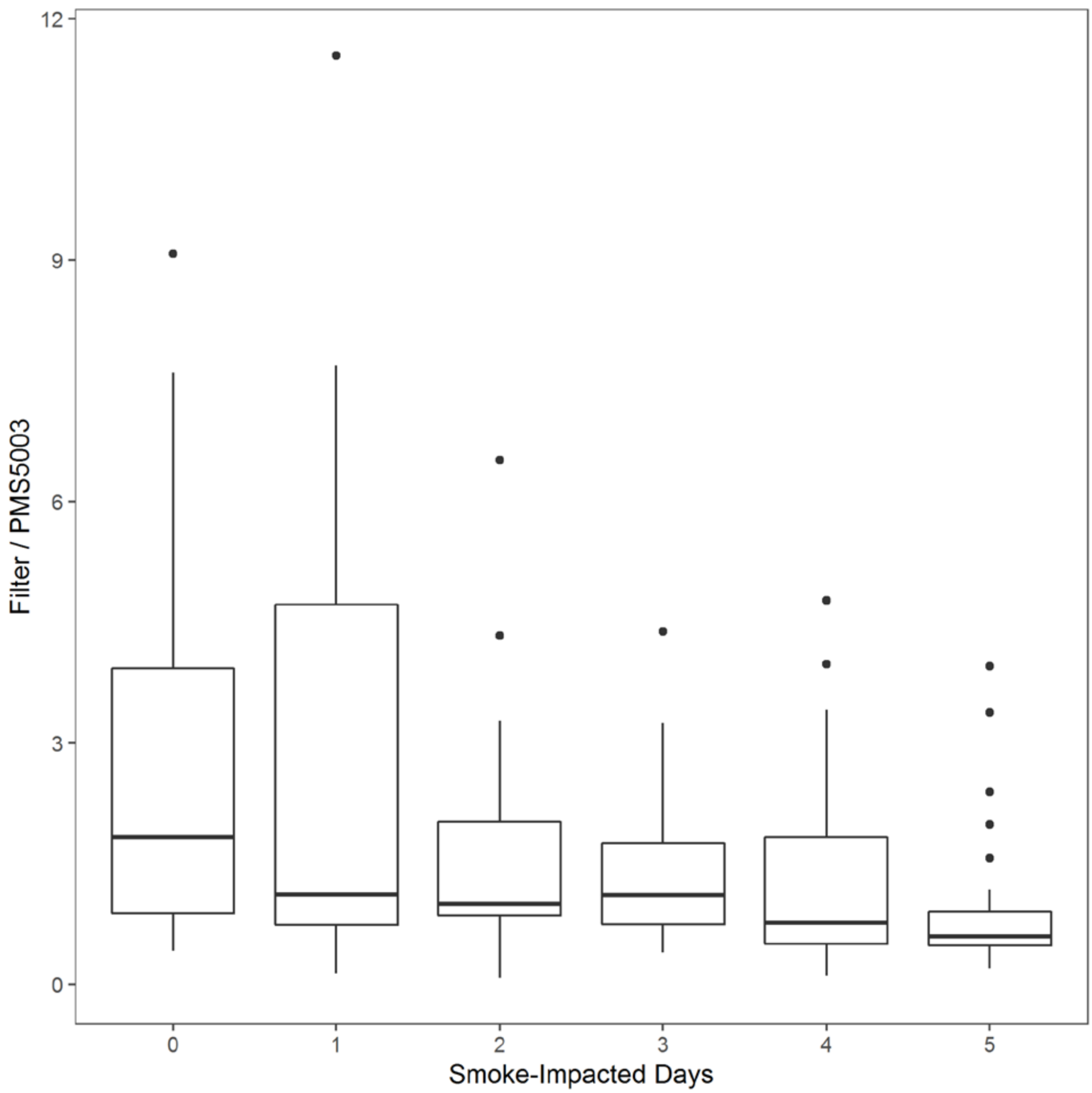

Figure B1: Box and whisker plots of how filter/PMS5003 values (i.e., Plantower scaling factors) varied with the number of smoke days for AMODv2 sampling runs. Filters with five smoke-impacted days included two partial sampling days (i.e. sample for 12 hours on the first day, and another 12 hours on the last day, for a total of 96 hours.)

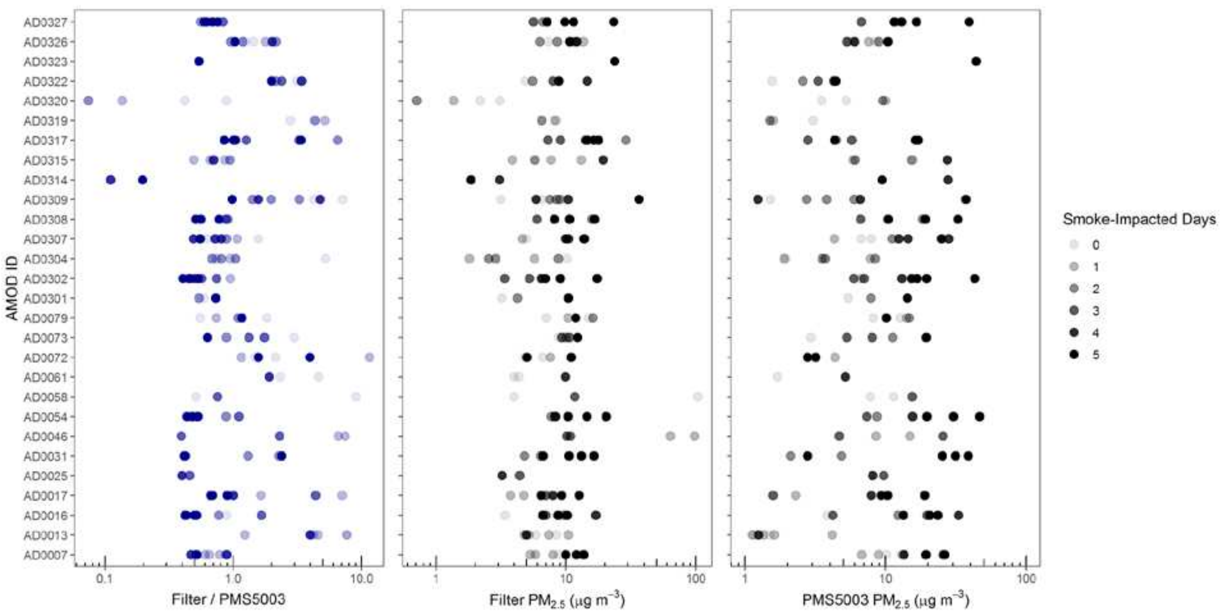

Figure B2: Distribution of filter/PMS5003 concentration ratio by number of smoke impacted days and AMODv2 identification number. Filters with five smoke-impacted days included two partial sampling days (i.e. sample for 12 hours on the first day, and another 12 hours on the last day, for a total of 96 hours.)

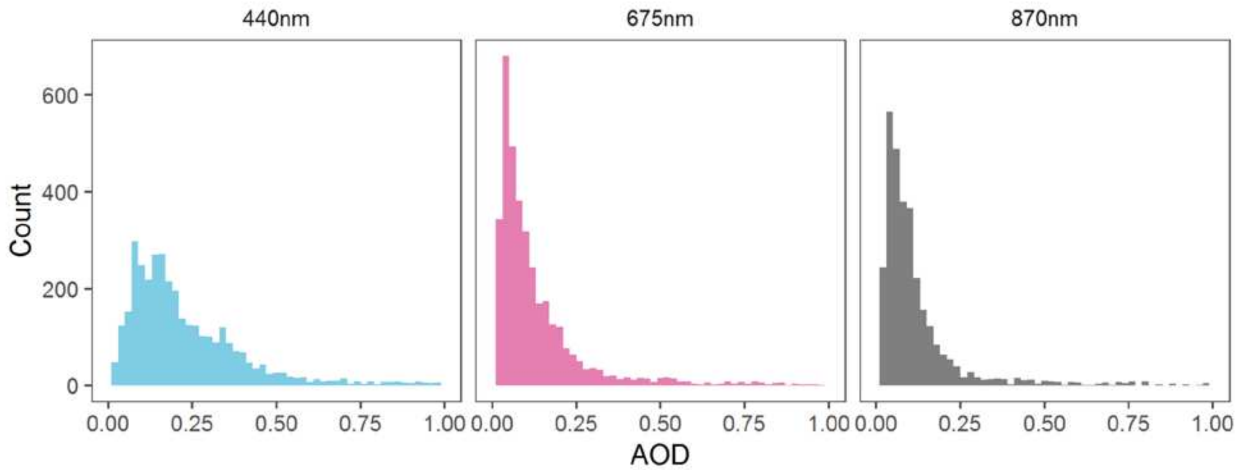

Figure B3: Distribution of valid AOD measurements in CEAMS campaign by wavelength.

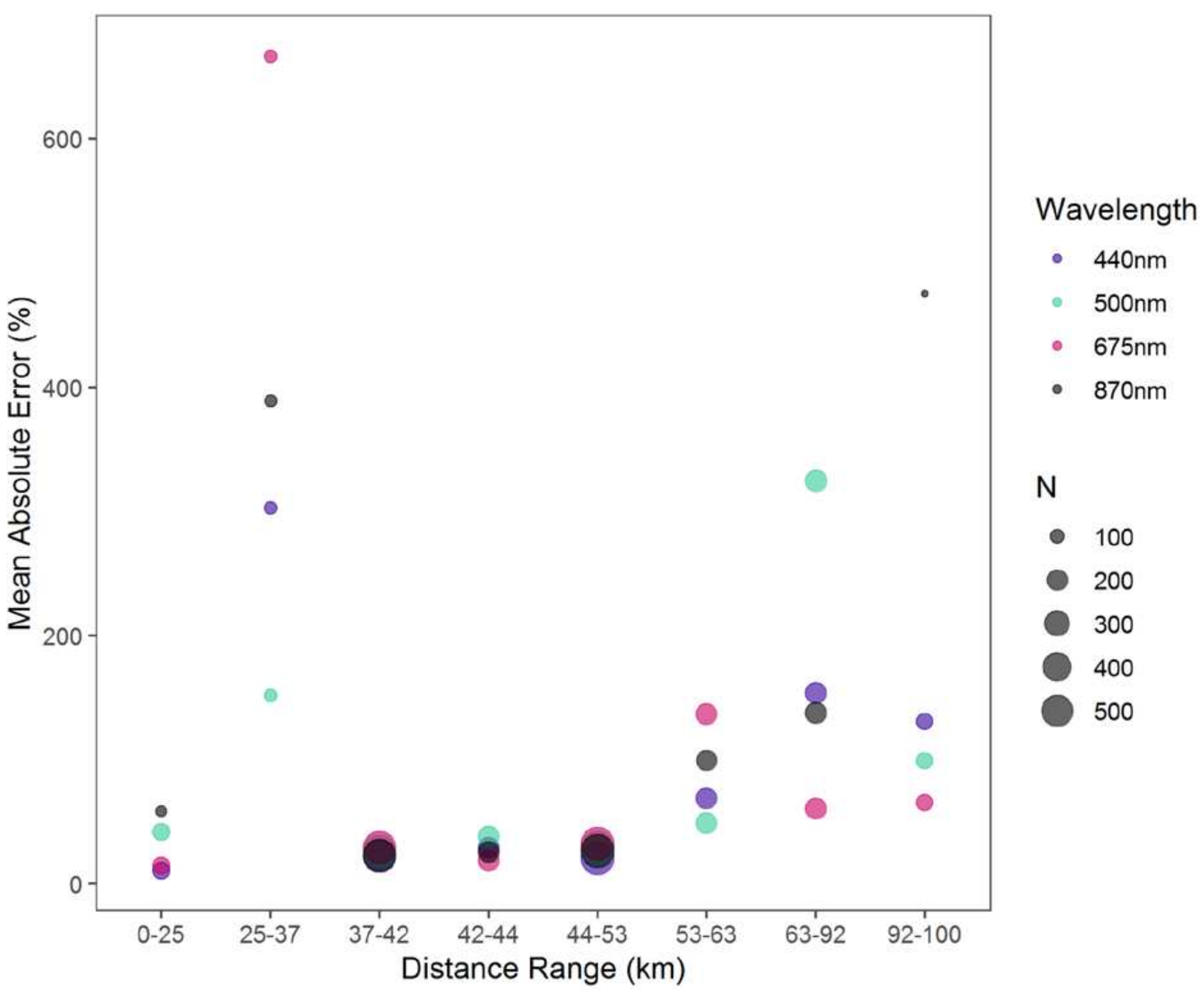

Figure B4: SARP vs. AERONET mean absolute percent error as a function of distance separating SARP and AERONET monitors.

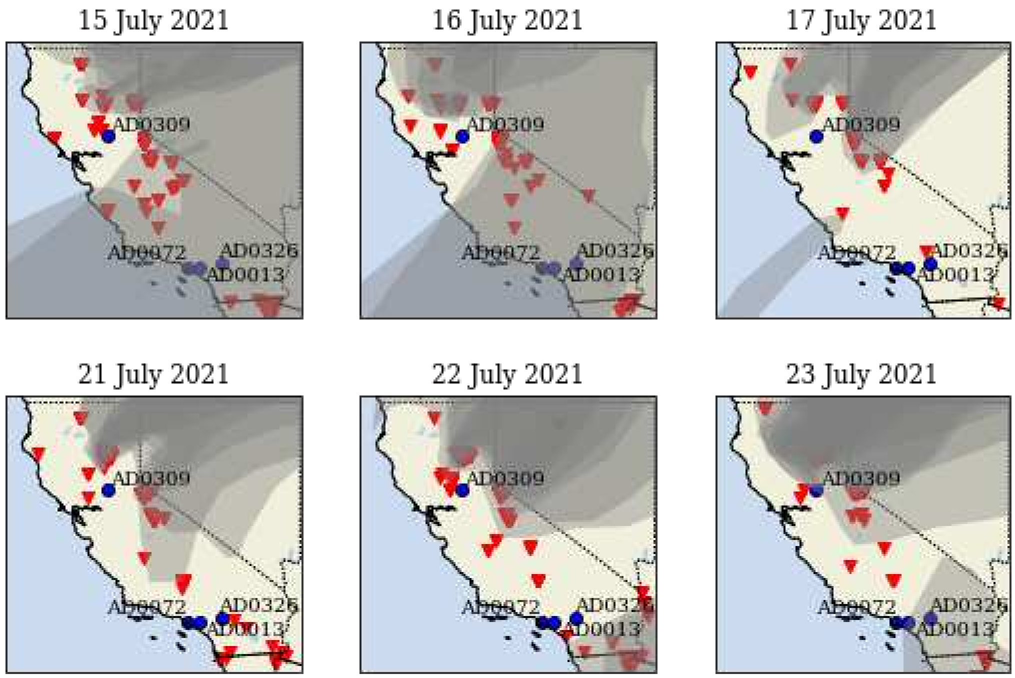

Figure B5: HMS smoke maps for selected time periods with smoke sources and AMODv2 locations labeled.

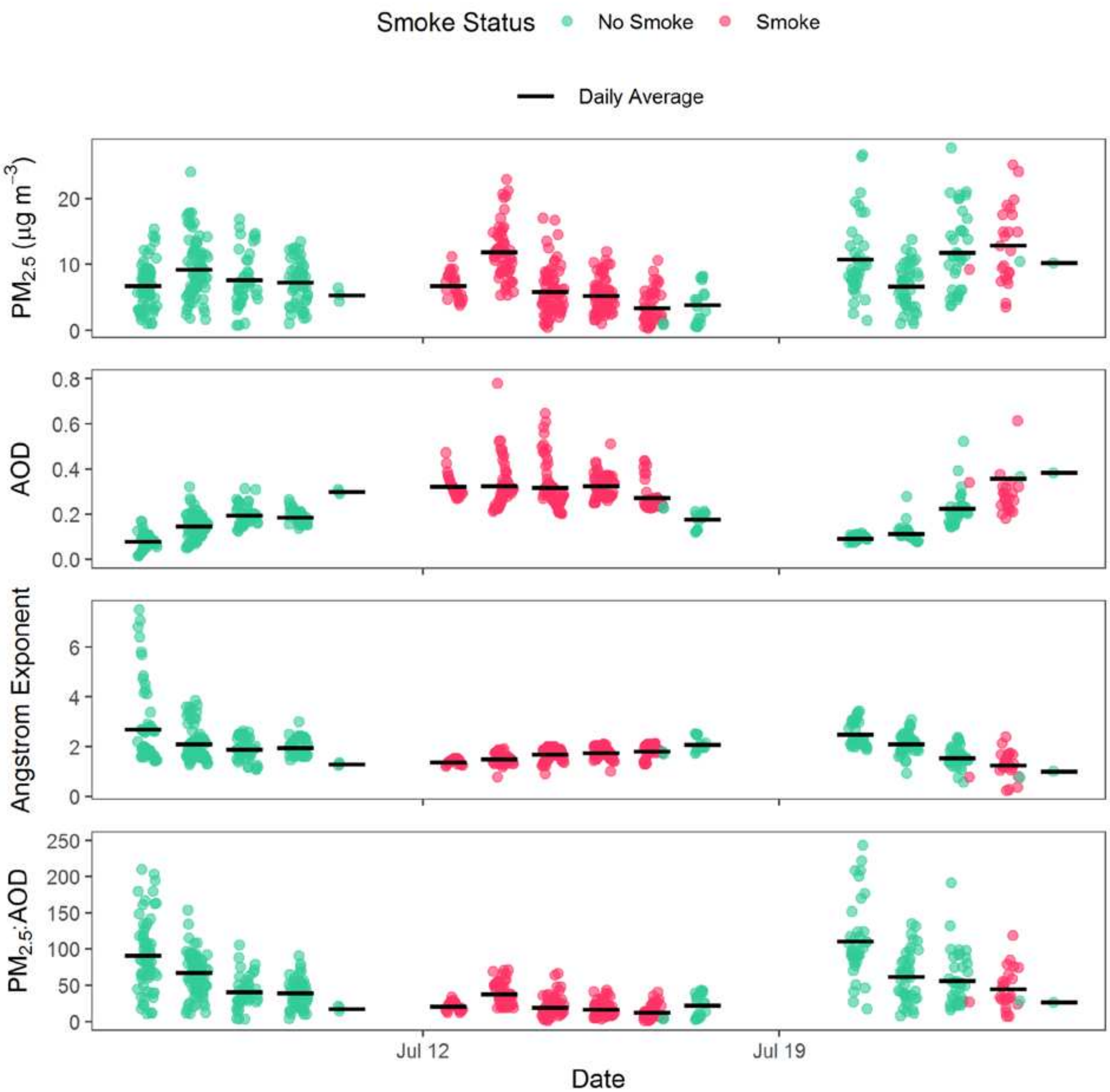

Figure B6: Time series plots of $PM_{2.5}$, AOD, Angstrom exponent, and $PM_{2.5}$:AOD from July 6, 2021 to July 24, 2021 from four CEAMS sites in Southern California. Measurements are from four AMODv2s (Fig. 1). Note that measurements on July 12, 2022 were from a single AMODv2 started early by the CEAMS participant.

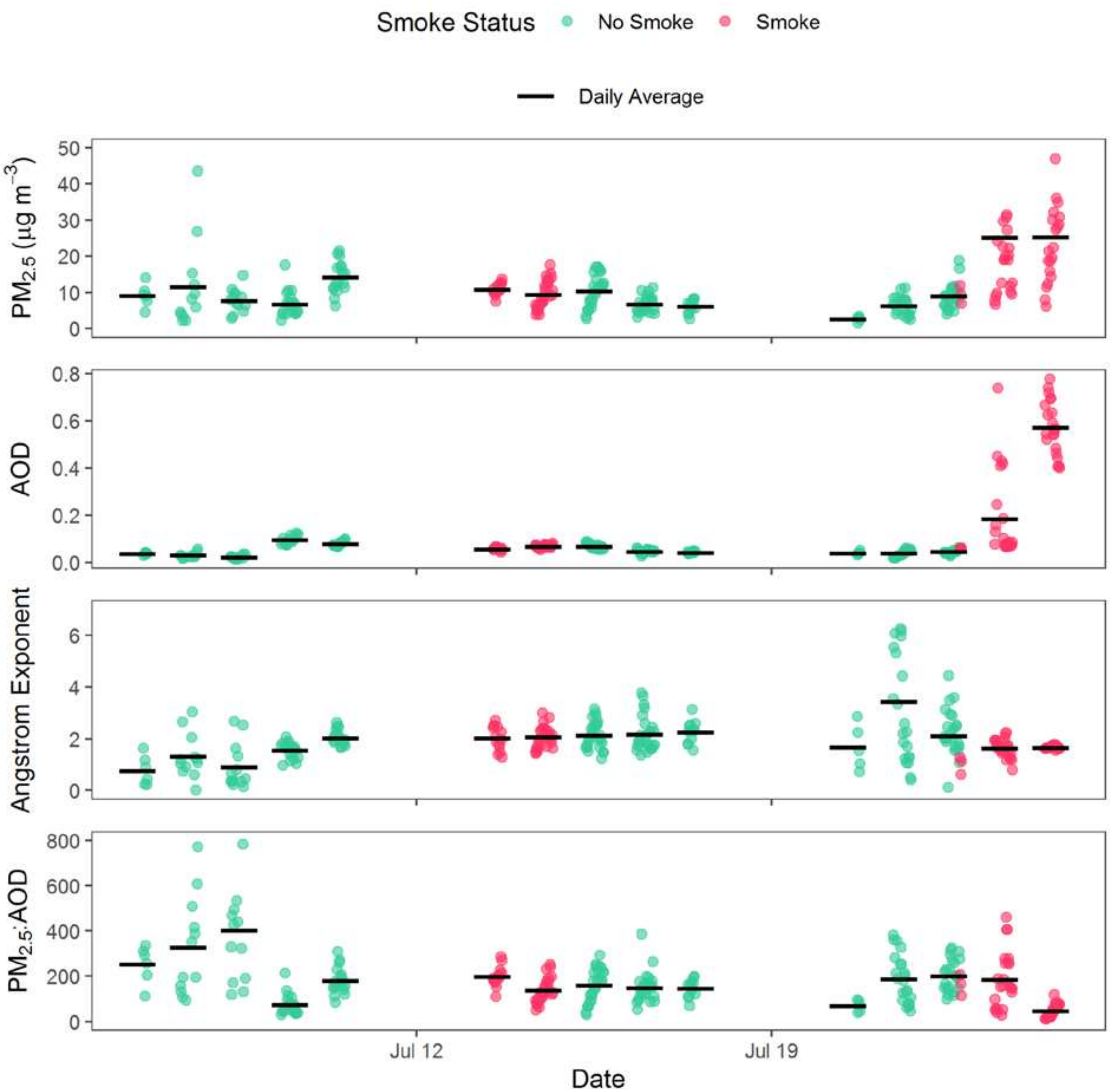

Figure B7: Time series plots of $PM_{2.5}$, AOD, Angstrom exponent, and $PM_{2.5}$:AOD from July 6, 2021 to July 24, 2021 from a CEAMS site near Sacramento, California. Note that measurements on July 12, 2022 were from a single AMODv2 started early by the CEAMS participant.

APPENDIX C

Text C1: We designed a prototype all-sky-imaging module using relatively low-cost, and commercially available components. The camera included an image sensor (Sony, IMX477-AACK-C, Minato City, Tokyo, Japan) and integrated circuitry for simple interfacing with the Raspberry Pi 4B. We achieved a 180° field of view using a fisheye lens (Arducam Technology Co., Limited, M25170H12, Hong Kong), and adequate sunlight attenuation by layering three, 3-stop, neutral density filters (Kodak, 1964741, Rochester, New York, USA) between the image sensor element and the lens. A photograph of the module and its associated componentry is provided in Fig. B1.

We used a Raspberry Pi 4B with 4 GB of RAM (Raspberry Pi Foundation, Cambridge, United Kingdom) as the processing unit and the Raspberry Pi HQ Camera Module (Raspberry Pi Foundation, SC0261, Cambridge, United Kingdom) as the imaging unit. The camera included an image sensor (Sony, IMX477-AACK-C, Minato City, Tokyo, Japan) and integrated circuitry for simple interfacing with the Raspberry Pi 4B. We achieved a 180° field of view using a fisheye lens (Arducam Technology Co., Limited, M25170H12, Hong Kong), and adequate sunlight attenuation by layering three, 3-stop, neutral density filters (Kodak, 1964741, Rochester, New York, USA) between the image sensor element and the lens. This configuration had the effect of reducing the image sensor exposure by a factor of 512. The Raspberry Pi 4B was powered with a 5 V, 4 Ah Lithium Polymer battery. We housed the electrical components in an electrical enclosure (Polycase, YQ-080804, Avon, Ohio, USA) and cut a circular hole in the lid to expose the image sensor to the sky. The cost of goods to produce a single prototype was 282.48 USD. A summary of component costs is provided in Table C1.

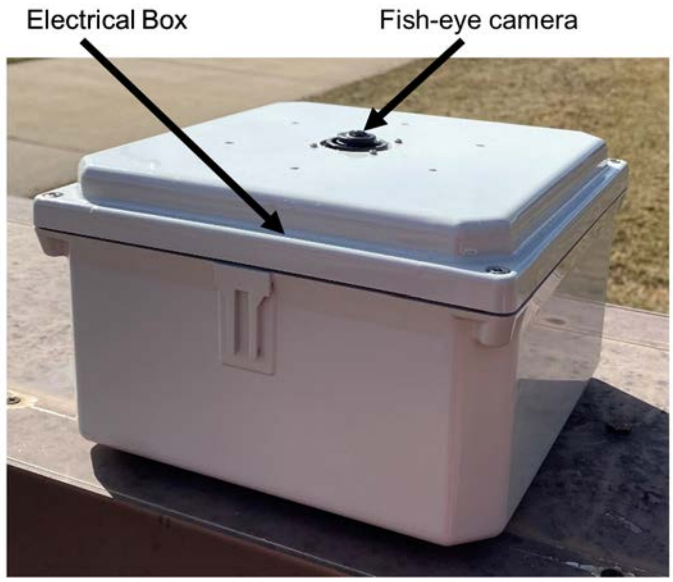

Figure C1: Photograph of NCASI imaging module.

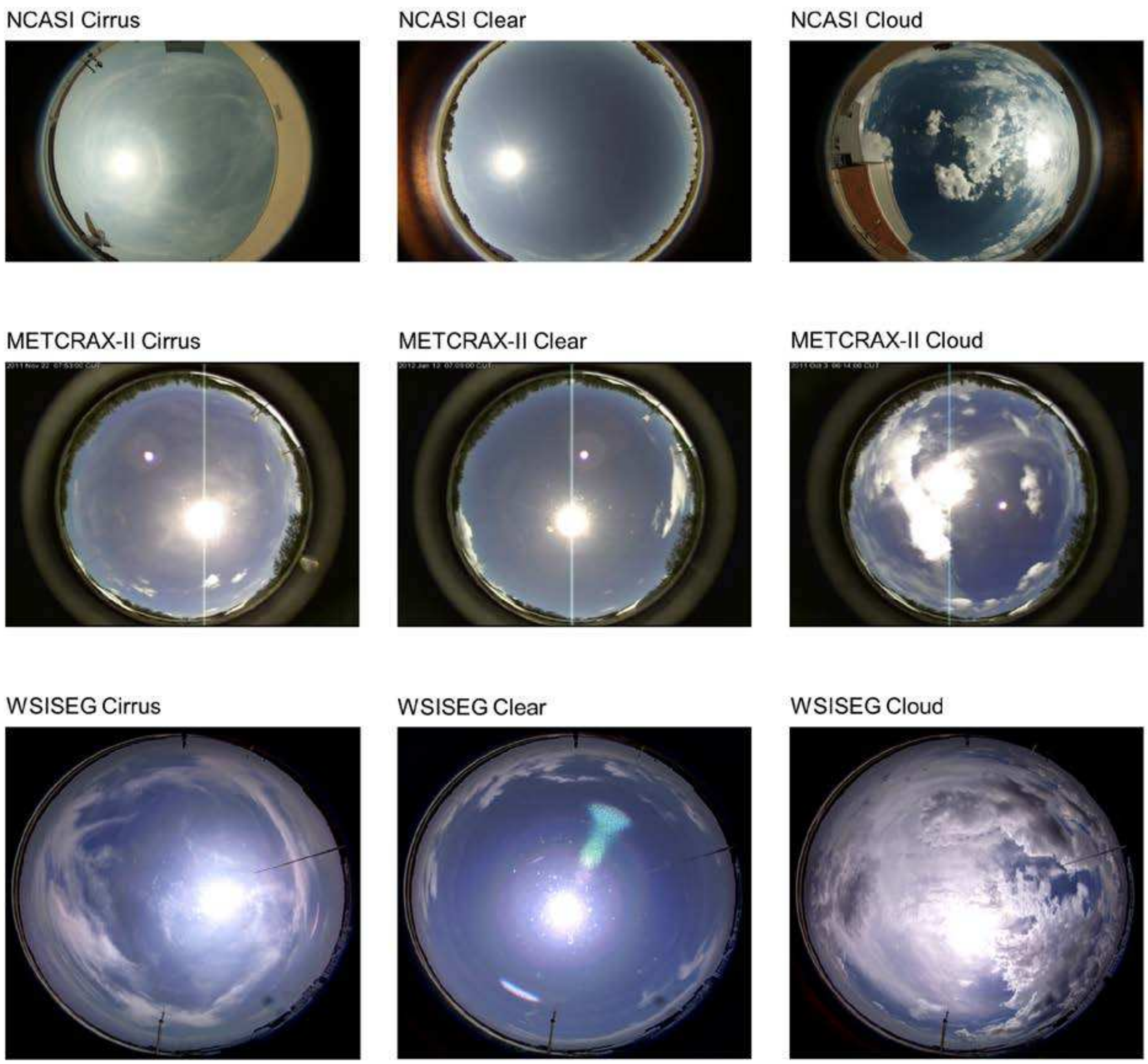

Figure C2: Example images from NCASI, METACRAX-II, and WSISEG for cirrus, clear, and cloud designations

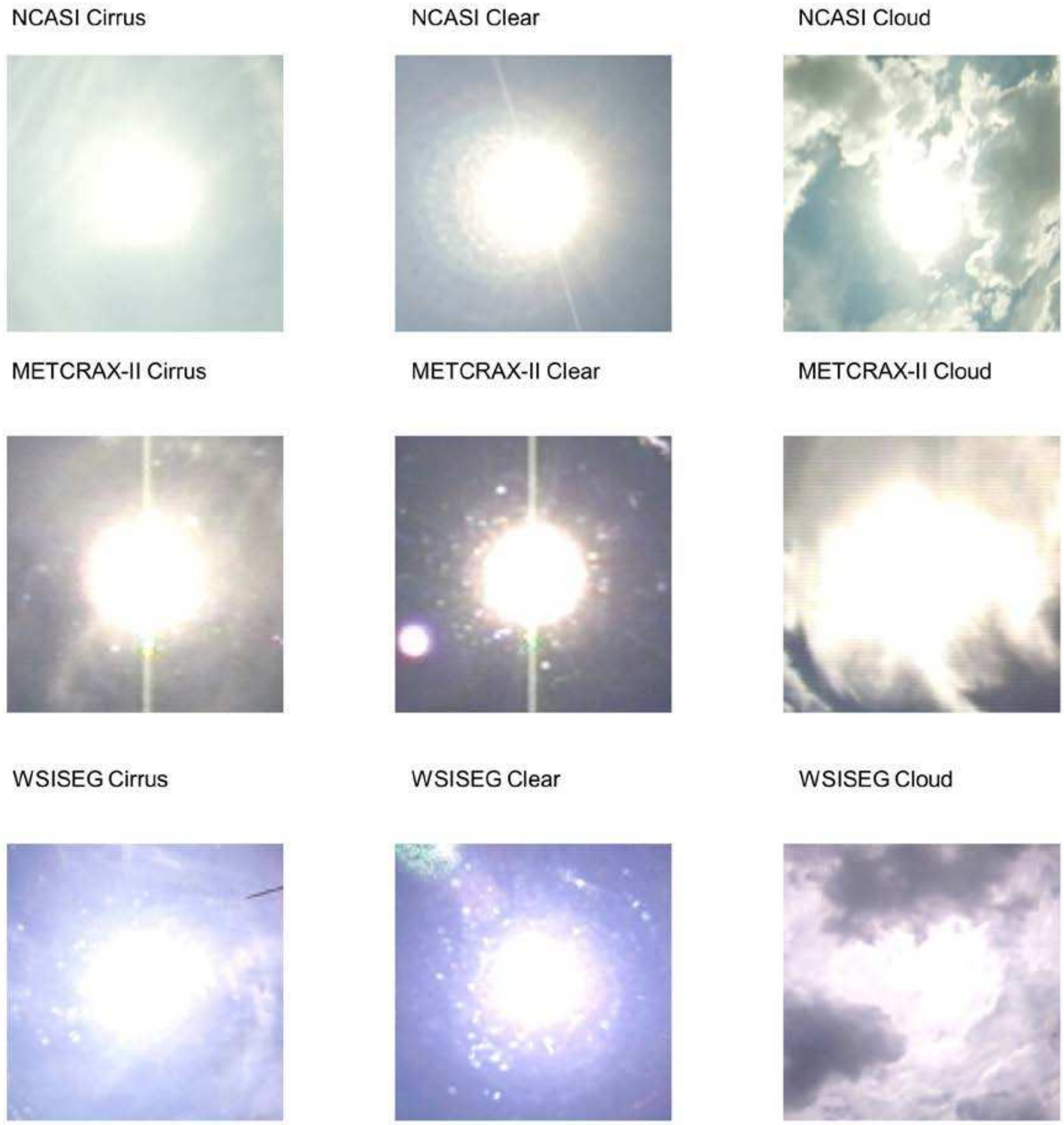

Figure C3: Example pre-processed images from NCASI, METACRAX-II, and WSISEG for cirrus, clear, and cloud designations.

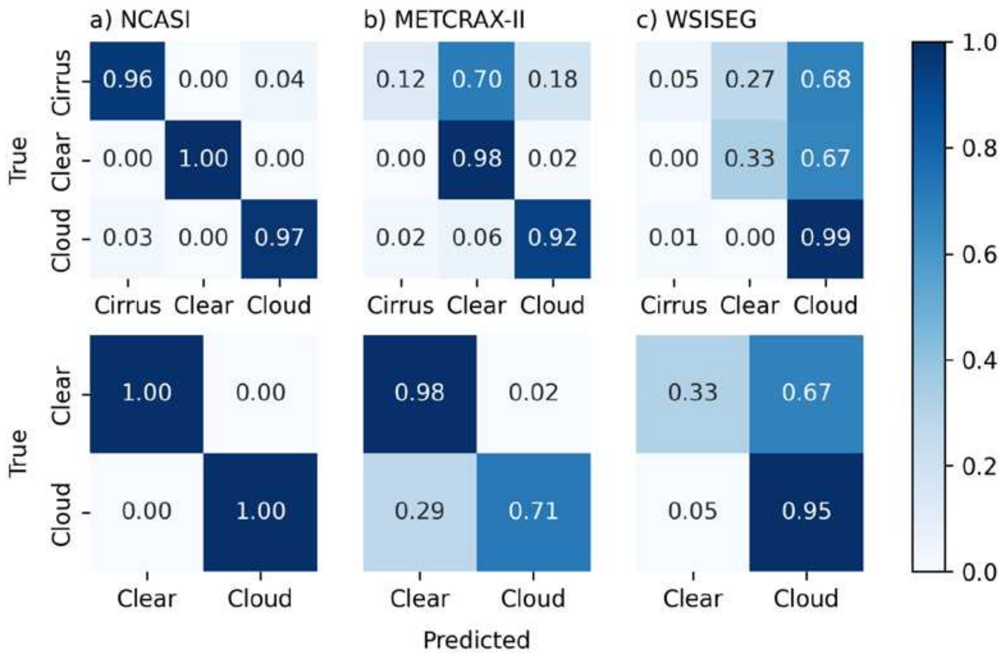

Figure C4: Top row: three-class confusion matrices for transfer learning model trained on NCASI training data set (model 1). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. Bottom row: two-class confusion matrices for transfer learning model trained on NCASI training data set (model 1). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set.

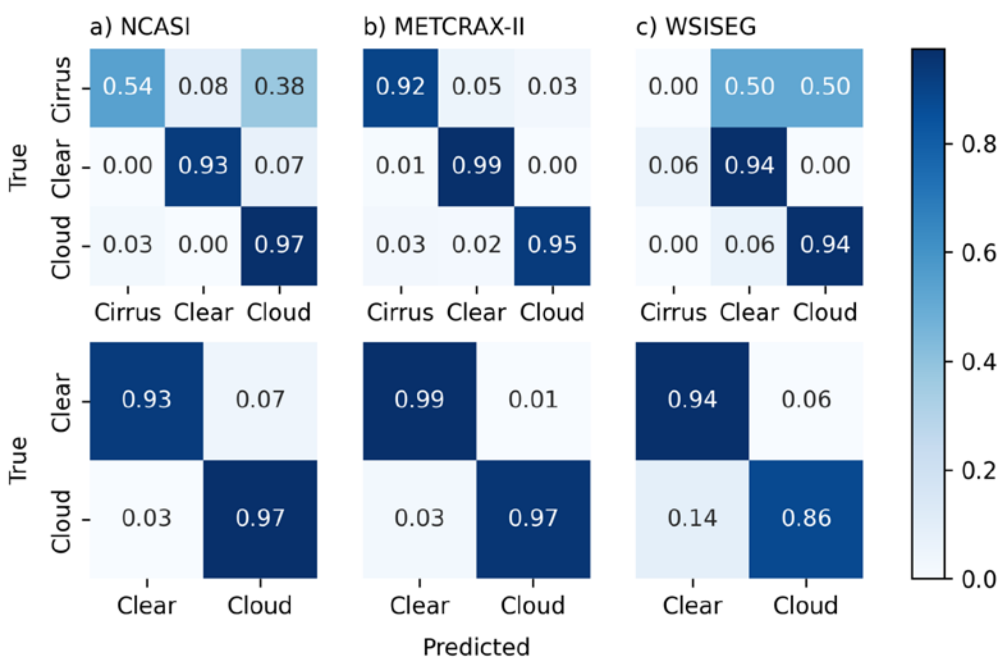

Figure B5: Top row: three-class confusion matrices for transfer learning model trained on METCRAX-II training data set (model 2). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. Bottom row: two-class confusion matrices for transfer learning model trained on METCRAX-II training data set (model 2). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set.

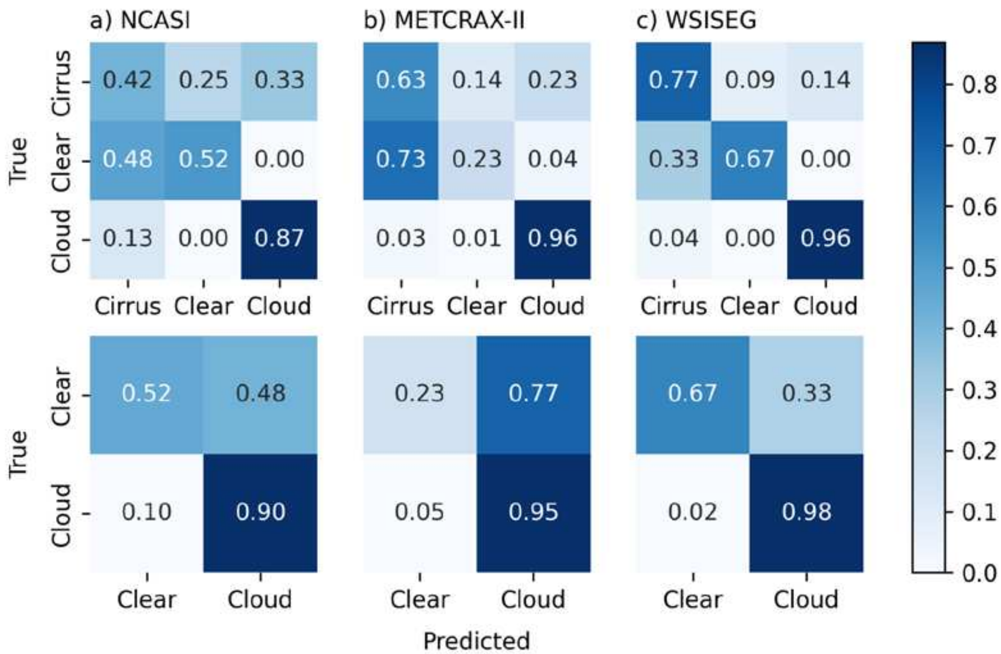

Figure C6: Top row: three-class confusion matrices for transfer learning model trained on WSISEG training data set (model 3). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. Bottom row: two-class confusion matrices for transfer learning model trained on WSISEG training data set (model 3). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set.

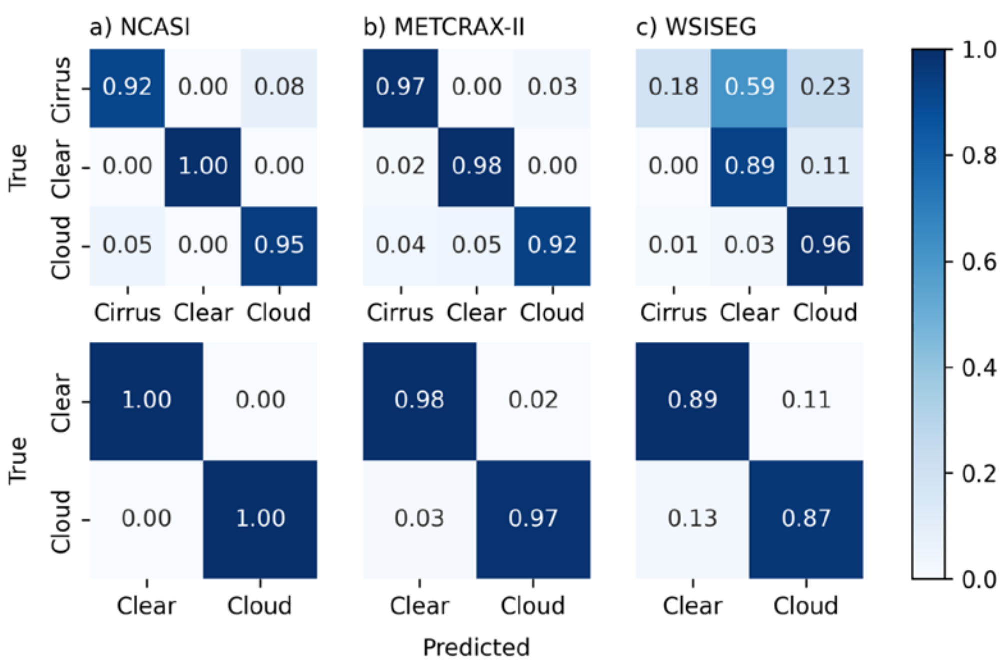

Figure C7: Top row: three-class confusion matrices for transfer learning model trained on NCASI and METCRAX-II training data sets (model 4). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. Bottom row: two-class confusion matrices for transfer learning model trained on NCASI and METCRAX-II training data sets (model 4). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set.

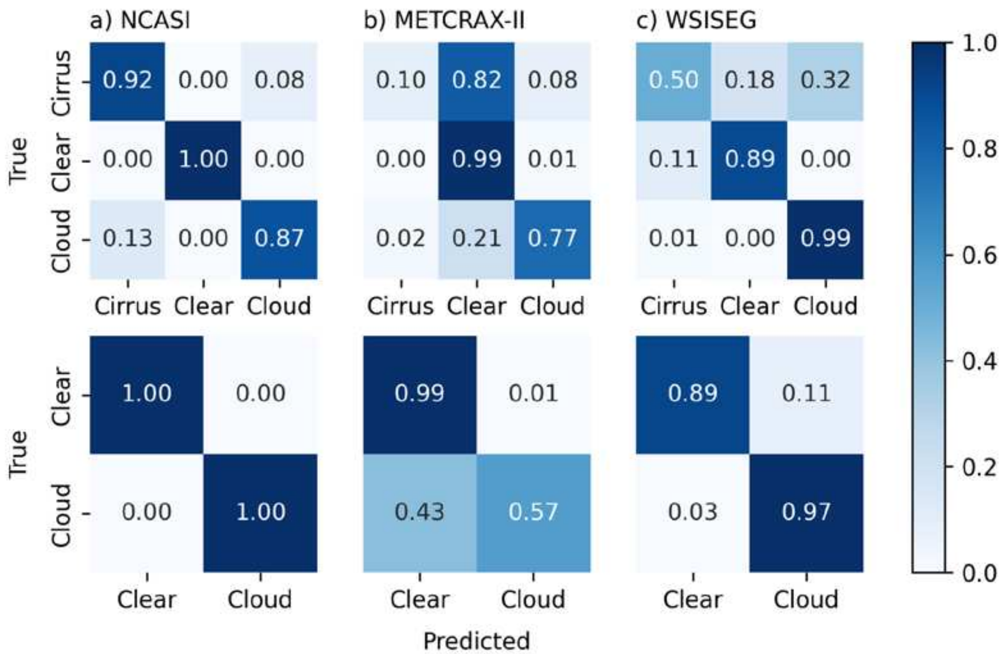

Figure C8: Top row: three-class confusion matrices for transfer learning model trained on NCASI and WSISEG training data sets (model 5). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. Bottom row: two-class confusion matrices for transfer learning model trained on NCASI and WSISEG training data sets (model 5). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set.

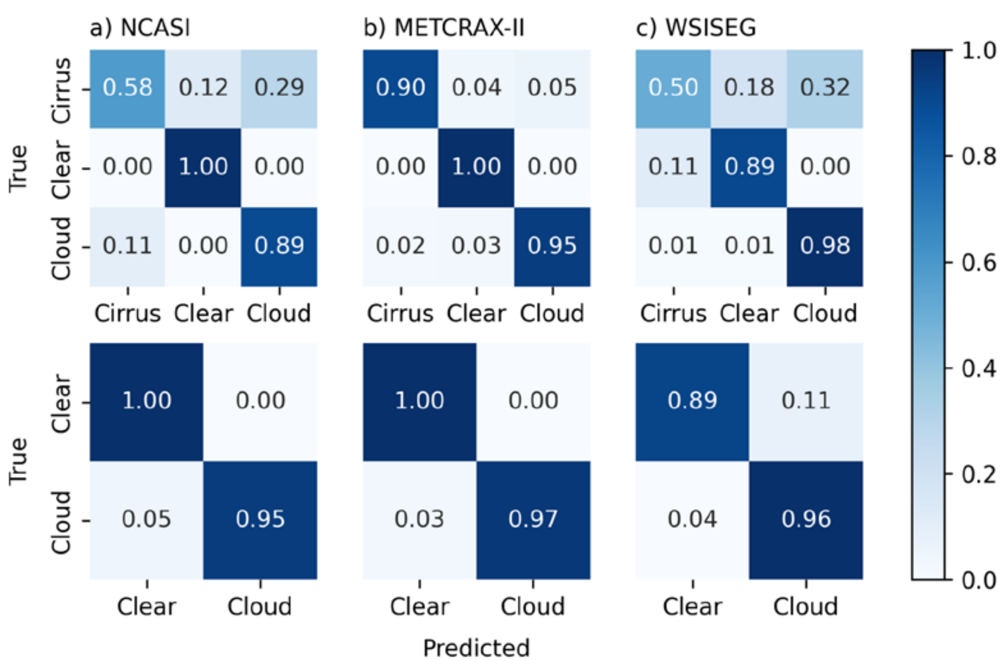

Figure C9: Top row: three-class confusion matrices for transfer learning model trained on METCRAX-II and WSISEG training data sets (model 6). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. Bottom row: two-class confusion matrices for transfer learning model trained on METCRAX-II and WSISEG training data sets (model 6). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set.

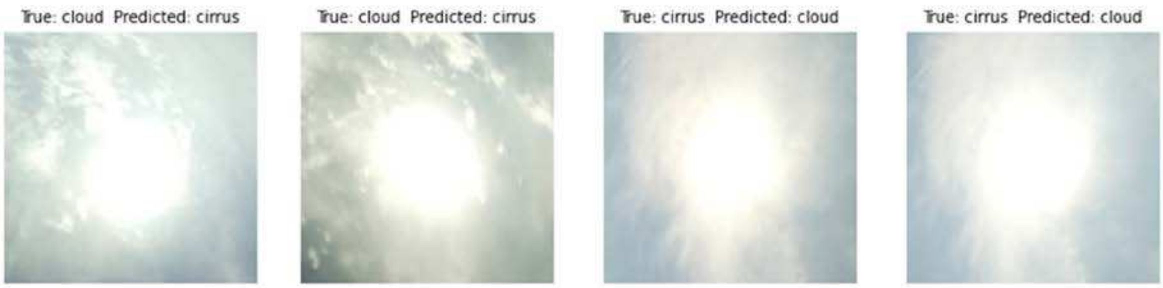

Figure C10: Misclassified images on NCASI testing data set for the model trained on NCASI, METCRAX-II, and WSISEG data sets (model 7).

Figure C11: Misclassified images on METCRAX-II testing data set for the model trained on NCASI, METCRAX-II, and WSISEG data sets (model 7).

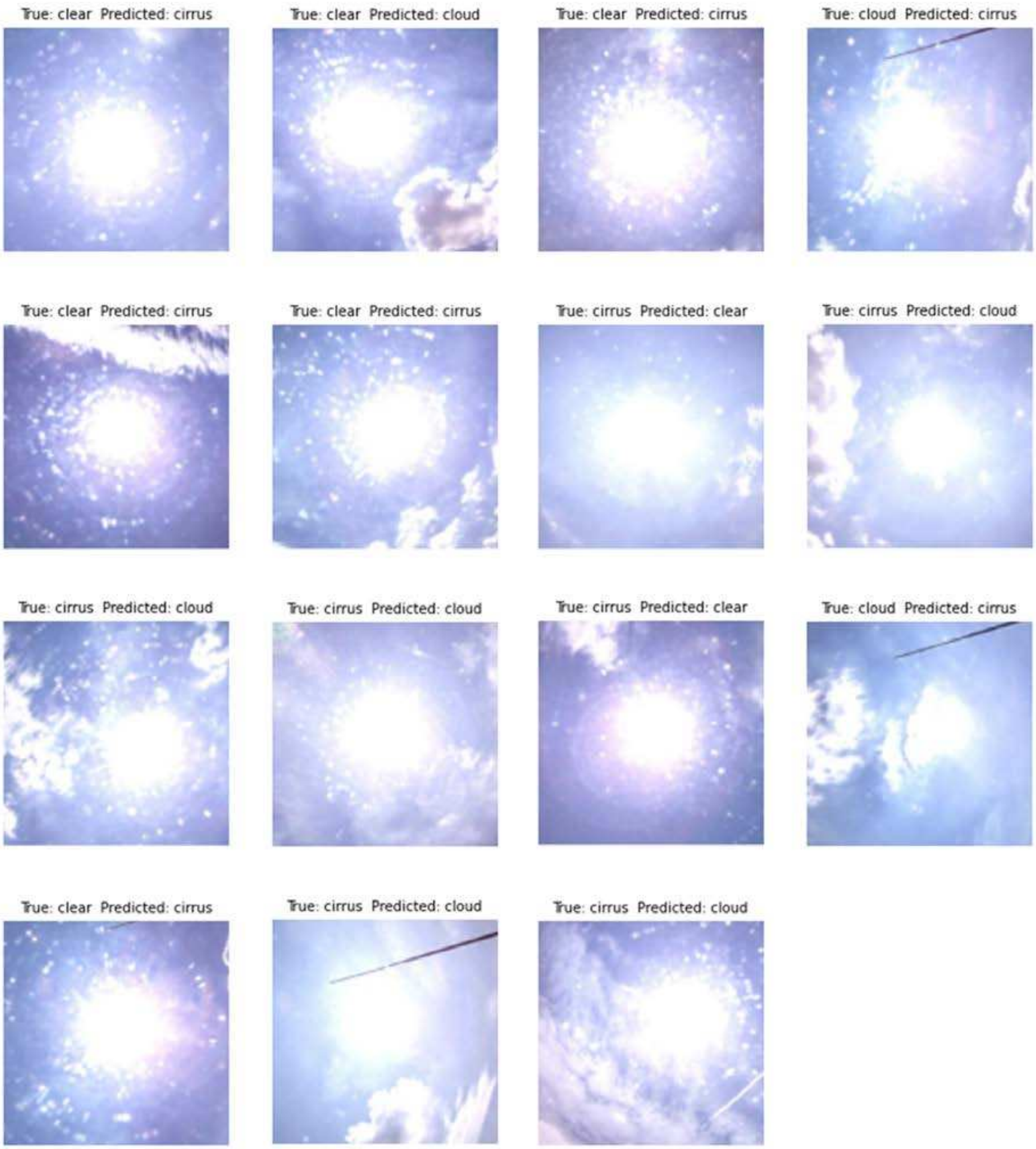

Figure C12: Misclassified images on WSISEG testing data set for the model trained on NCASI, METCRAX-II, and WSISEG data sets (model 7).

Table C1: Component costs of imaging module

Component	Manufacturer	Part Number	Cost (USD)
Raspberry Pi 4 Model B 4 GB	Raspberry Pi Foundation	Raspberry Pi 4B/4GB	55.00
Raspberry Pi HQ Camera Module	Raspberry Pi Foundation	SC0261	50.00
Arducam M12 Fisheye Lens	Arducam Technology Co., Limited	M25170H12	9.99
Neutral Density Optical Gelatin Wratten Filter (75 mm ×75 mm)	Kodak	1964741	97.50
Electrical Box	Polycase	YQ-0808804-13	55.00
Battery Pack for Raspberry Pi, 4000mAh	Yapears	5647469919	14.99

Total Cost: 282.48 USD