United States Environmental Protection Agency

EPA/600/R-20/280 | February 2021

Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors

USE IN AMBIENT, OUTDOOR, FIXED SITE, NON-REGULATORY SUPPLEMENTAL AND INFORMATIONAL MONITORING APPLICATIONS

Office of Research and Development
Center for Environmental Measurement and Modeling

EPA/600/R-20/280
February 2021

Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors

Use in Ambient, Outdoor, Fixed Site, Non-Regulatory Supplemental and Informational Monitoring Applications

By

Rachelle M. Duvall, Andrea L. Clements, Gayle Hagler, Ali Kamal, and Vasu Kilaru
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Laura Goodman and Samuel Frederick
National Student Services Contract
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Karoline K. (Johnson) Barkjohn and Ian VonWald
Oak Ridge Institute for Science and Education
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Danny Greene
Eastern Research Group, Inc.
Morrisville, NC 27560

Tim Dye
TD Environmental Services, LLC
Petaluma, CA 94952

Disclaimer

This technical report presents work performed by the U.S. Environmental Protection Agency's (U.S. EPA) Office of Research and Development (ORD) with technical support provided by Eastern Research Group through a task order (Task Order 68HERC19F0071, EPA Contract #EP-C-16-015). The effort represents a collaboration between ORD and the U.S. EPA Office of Air Quality Planning and Standards (OAQPS). Any mention of trade names, manufacturers, commercial products, or research institutions does not constitute endorsement. This report has been reviewed in accordance with U.S. EPA policy and approved for publication. Testing results do not constitute certification or endorsement by the U.S. EPA.

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

AMTIC	Ambient Monitoring Technology Information Center
AQI	Air Quality Index
AQ-SPEC	Air Quality Sensor Performance Evaluation Center
AQS	Air Quality System
b	intercept
CEMM	Center for Environmental Measurement and Modeling
CFR	Code of Federal Regulations
CSN	Chemical Speciation Network
CV	coefficient of variation
°C	degrees Celsius
DL	detection limit
DP	dew point
Eq	equation
FEM	Federal Equivalent Method
FRM	Federal Reference Method
m	slope
MEE	Ministry of Ecology and Environment (People's Republic of China)
NAAQS	National Ambient Air Quality Standards
NCore	National Core Network
NIST	National Institute of Standards and Technology
NRMSE	normalized root mean square error
NSIM	non-regulatory supplemental and informational monitoring
OAQPS	Office of Air Quality Planning and Standards
ORD	Office of Research and Development
PM	particulate matter
$PM_{2.5}$	fine particulate matter; particles with aerodynamic diameters less than 2.5 micrometers
QA	quality assurance
QAPP	Quality Assurance Project Plan
QC	quality control
r	Pearson correlation coefficient
R^2	coefficient of determination
RH	relative humidity
RMSE	root mean square error
SD	standard deviation
SLAMS	State and Local Air Monitoring Station
T	temperature
$\mu g/m^3$	microgram per cubic meter
U.S.	United States
U.S. EPA	United States Environmental Protection Agency

Acknowledgments

The authors acknowledge the Eastern Research Group technical staff associated with Task Order 68HERC19F0071, EPA Contract #EP-C-16-015 for their research efforts summarized here. This effort was led by the United States Environmental Protection Agency (U.S. EPA), Office of Research and Development (ORD), Center for Environmental Measurement and Modeling (CEMM) with support from the Sensor Performance Targets Deliberation Workgroup that consisted of staff across the U.S. EPA. We acknowledge the Subject Matter Experts that participated in the U.S. EPA's 2018 workshop on "Deliberating Performance Targets for Air Quality Sensors", whose feedback helped inform this report. Libby Nessley and Christina Alvarez (U.S. EPA/ORD/CEMM Quality Assurance) are recognized for quality assurance support in developing this report. We acknowledge U.S. EPA internal reviewers Kristen Benedict, Trisha Curran, Ron Evans, Tim Hanley, Doug Jager, Sue Kimbrough (retired), and Robert Vanderpool. We also acknowledge the following external reviewers: Pratim Biswas (Washington University); Jason Treutel (Wisconsin Department of Natural Resources); and Vasileios Papapostolou (South Coast Air Quality Management District).

Title page photo by South Coast Air Quality Management District as part of work under the U.S. EPA Contract #EP-16-W-000117. Photo shows the custom-built Citizen Science Air Monitors at the Rubidoux Air Monitoring site in Riverside, California.

Executive Summary

Air sensors have become more accessible nationwide and their development continues to expand and evolve at a rapid pace. There has been a dramatic increase in the use of air sensors for a variety of air monitoring applications and data sets have become more available to the public. While air sensors have encouraged innovation in air monitoring approaches, it is widely known that the data quality from these technologies is highly variable. The variability in data quality makes it challenging to understand the performance of any given sensor device and if a sensor will appropriately fit an application of interest. Additionally, organizations that manage air quality face challenges in responding to air sensor data provided by the public as there is a lack of knowledge on how air sensor technologies perform which can make it more difficult to trust or interpret the data.

While programs such as the United States Environmental Protection Agency's (U.S. EPA) Federal Reference Method and Federal Equivalent Method (FRM/FEM) Program [Code of Federal Regulations, Title 40 (40 CFR) Parts 50, 53, and 58] contain standards and performance certification processes for air quality instruments used for regulatory monitoring purposes, it is recognized that air sensors will likely not meet those stringent requirements. However, sensors could be useful in many non-regulatory applications such as understanding local air quality trends, identifying hot spots, supplemental monitoring, and promoting educational/environmental awareness. Currently, there are no standard testing protocols or targets for air sensors.

The objective of this report is to provide a consistent set of testing protocols, metrics, and target values to evaluate the performance of fine particulate matter ($PM_{2.5}$) air sensors specifically for non-regulatory supplemental and informational monitoring (NSIM) applications in ambient, outdoor, fixed site environments. Two testing protocols, base testing and enhanced testing, are recommended (Table ES-1).

Table ES-1. Recommended Testing Protocols for Understanding $PM_{2.5}$ Air Sensor Performance

Test Type	Setting	Description	Purpose
Base Testing	Field	Consists of field deployments of at least three replicate $PM_{2.5}$ air sensors with collocated FRM/FEM monitors for a minimum of 30 days each, at two test sites within different climate regions.	Provides information on sensor performance that is relevant to real-world, ambient, outdoor conditions. Allows consumers to predict how a sensor might perform in similar conditions.
Enhanced Testing	Laboratory	Consists of testing at least three replicate $PM_{2.5}$ air sensors in controlled laboratory conditions to understand the effect of temperature and relative humidity; drift; and accuracy at higher concentration levels.	Allows for evaluation of sensors over a range of conditions that may be challenging to capture in the field. Characterizes certain performance parameters that are difficult to test in the field.

All testers are encouraged to conduct base testing at a minimum. Enhanced testing is also encouraged although it calls for a controlled laboratory exposure chamber. For NSIM applications where high $PM_{2.5}$ concentrations are expected (e.g., wildfire smoke applications), it is recommended that testers conduct base testing in more than two locations and include sites impacted by wildfire smoke and higher $PM_{2.5}$ concentrations.

Performance metrics and corresponding target values have been identified based on the current state-of- the-science, literature reviews, findings from other organizations that conduct routine sensor evaluations, sensor standards/certification programs in development by other organizations, and the U.S. EPA expertise in sensor evaluation research. A summary of the performance metrics and target values for the base and enhanced testing protocols are shown in Table ES-2. For base testing, an additional data visualization called 'exploring meteorological effects' is recommended which includes graphing meteorological data to understand its influences on sensor performance. Further for base testing, it is recommended that at the base test sites, at least one day of the testing period has a 24-hour average $PM_{2.5}$ concentration of at least 25 microgram per cubic meter ($\mu g/m^3$) or greater. Additional performance metrics and test conditions for the enhanced testing protocols are shown in Table ES-3. This report provides details on how to calculate the performance metrics for $PM_{2.5}$ sensors (see Section 3.0) and templates for base and enhanced testing reports for consistent reporting of testing results (see Appendix F and H).

Table ES-2. Base and Enhanced Testing – Recommended Performance Metrics and Target Values for $PM_{2.5}$ Air Sensors

Performance Metric		**Target Value**	
		Base Testing	**Enhanced Testing***
Precision	Standard Deviation (SD)	$\leq 5\ \mu g/m^3$	No target values recommended; report results
	-OR-		
	Coefficient of Variation (CV)	$\leq 30\%$	
Bias	Slope	1.0 ± 0.35	
	Intercept (b)	$-5 \leq b \leq 5\ \mu g/m^3$	
Linearity	Coefficient of Determination (R^2)	≥ 0.70	
Error	Root Mean Square Error (RMSE) or Normalized Root Mean Square Error (NRMSE)	RMSE $\leq 7\ \mu g/m^3$ or NRMSE $\leq 30\%$†	

*No specific target values are recommended due to limited feasibility, lack of consensus regarding testing protocols, and inconsistency in sensor evaluation results that can result from the limited amount of data that will be collected and variation in the tester's choice of PM surrogate. See Appendix D for further discussion.

†A sensor will meet this target if either the RMSE or NRMSE meet this criterion. See Appendix D for further discussion.

Table ES-3. Enhanced Testing – Additional Recommended Performance Metrics and Test Conditions for $PM_{2.5}$ Air Sensors

Performance Metric	Test Conditions
Effect of Relative Humidity (RH)	Moderate RH: 40% ± 5%
	Elevated RH: 85% ± 5%
Effect of Temperature (T)	Moderate T: 20°C ± 1°C
	Elevated T: 40°C ± 1°C
Drift	Low concentration: 10 $\mu g/m^3$ ± 10%
	Mid concentration: 35 $\mu g/m^3$ ± 5%
Accuracy at High Concentrations	High concentration: 150 $\mu g/m^3$ ± 5%
	Higher concentration: 250 $\mu g/m^3$ ± 5%

The performance metrics and target values for base and enhanced testing are recommended based on the current knowledge of $PM_{2.5}$ air sensors at the time this report was released. Target values for enhanced testing are not included at this time due to limited feasibility, lack of consensus regarding testing protocols, and inconsistency in sensor evaluation results that can result due to the limited amount of data that will be collected and variation in the tester's choice of PM surrogate.

It is recognized that $PM_{2.5}$ sensor technologies will likely continue to develop and improve over time. The U.S. EPA anticipates updating Tables ES-2 and ES-3 as well as other information in this report, as feasible, to reflect advances in $PM_{2.5}$ sensor technologies and knowledge gained from sensor evaluation results. Updates will likely be shared as an addendum to this report.

The intended audience for this report includes potential testing organizations, sensor manufacturers, and sensor developers. It is anticipated that a variety of consumers (e.g., state/local/tribal agencies, federal government agencies, community groups, citizen scientists, academia) will benefit from the consistent presentation of testing results to identify sensor technologies that would be best suited for their NSIM application and understand the performance of the air sensor technologies. Consumers may also choose to conduct these testing protocols.

Testing results do not constitute certification or endorsement by the U.S. EPA. It is recommended that testers make the testing reports available on their respective websites to inform consumers on the testing results.

1.0 Introduction

1.1 Background

The term ‘air sensor’ refers to a class of non-regulatory technology that are lower in cost, portable, and generally easier to operate than regulatory monitors. Air sensors often provide relatively quick or instant air pollution concentrations (both gas-based and particulate matter) and allow air quality to be measured in more locations. The term ‘air sensor’ often describes an integrated set of hardware and software that uses one or more sensing elements (also sometimes called sensors) to detect or measure pollutant concentrations. Other commonly used terms for air sensors include “low-cost air sensors”, “lower cost air sensors”, “air sensor devices”, “air sensor pods”, and “air quality sensors”. Advancements in microprocessors and miniaturization have led to a rapid expansion in the availability of air sensors to measure a variety of air pollutants. As air sensors have become more accessible nationwide, there has been a dramatic increase in their use for non-regulatory air quality monitoring purposes and greater access to publicly available sensor data sets (e.g., Zamora et al., 2020; Li et al., 2020; Feenstra et al., 2019; Bulot et al., 2019; Badura et al., 2018; Crilley et al., 2018; Zheng et al., 2018; Nakayama et al, 2018; Kelly et al., 2017; Zikova et al., 2017; Mukerjee et al., 2017).

Since 2012, the United States Environmental Protection Agency (U.S. EPA) has been involved in many activities related to air sensors including, but not limited to, hosting workshops and webinars, evaluating new technologies and applications, developing tools to analyze and visualize data, and disseminating information. More details on these efforts can be found on the U.S. EPA’s Air Sensor Toolbox website (https://www.epa.gov/air-sensor-toolbox, *last accessed 07/25/2020*).

A variety of options for air sensors are available and development continues to expand and evolve at a rapid pace. However, it is widely known that the data quality from these technologies is highly variable (Williams et al., 2019). Some of the key challenges with $PM_{2.5}$ air sensor technologies include:

- Determining whether the sensor will measure the target pollutant accurately and reliably within the expected concentration range for the application;
- Determining how different parameters including relative humidity (RH), temperature (T), and variations in particle composition or size can impact measurements;
- Estimating how the sensor’s response changes over time and at what point in time the sensor reading becomes inaccurate or unreliable; and
- Understanding how sensors perform out-of-the-box and if correction or adjustments are needed to provide more accurate data.

While programs such as the U.S. EPA’s Federal Reference Method and Federal Equivalent Method (FRM/FEM) Program [Code of Federal Regulations, Title 40 (CFR 40) Parts 50, 53, and 58] contain standards and performance certification processes for air quality instruments used for regulatory monitoring purposes, it is recognized that air sensors will not meet those stringent requirements for several reasons. Monitors designated as FRM/FEMs are specifically designed and manufactured to produce reliable, high quality measurements for use in compliance monitoring that meet all acceptance criteria for laboratory and field tests as outlined 40 CFR Parts 50 and 53. Sensors are typically not designed with these criteria in mind. However, some testing requirements and acceptance criteria in Parts 50 and 53 may be adaptable to evaluate sensor performance.

Currently, there is an absence of testing protocols and performance targets that air sensor manufacturers/developers can use to evaluate their devices. The comparability of sensors with FRM/FEMs is highly variable and the ability of sensors to provide consistent, accurate, and precise measurement data under real-world conditions is not well understood. Nevertheless, there is ongoing interest in using air sensors in non-regulatory air monitoring applications. Testing protocols and targets for air sensors would increase confidence in data quality and help consumers in selecting sensors that appropriately suit an application of interest.

1.2 Motivation

Around 2012, when the availability of air sensors began to expand rapidly, questions related to using sensors and interpreting sensor data began to increase significantly among the user community. The U.S. EPA responded by developing the Air Sensor Guidebook (U.S. EPA, 2014). The guidebook was designed to provide basic foundational knowledge to help those interested in using sensors for air quality measurements with a focus on: 1) background information on common air pollutants and air quality, 2) selecting appropriate sensors for different applications, 3) data quality considerations, and 4) sensor performance for different applications. The target audience for the Air Sensor Guidebook was citizen scientists and sensor manufacturers/developers. Since then, the user community has grown to include individuals, communities, schools, air quality and health agencies, medical professionals, and more.

New air sensor technologies continue to flood the market. Despite ongoing research to evaluate these technologies, variability in data quality persists. While several organizations are in the process of developing performance standards or guidance for evaluating air sensors, currently there are no consistent testing protocols that can be used for uniform evaluation and comparison of different technologies. Furthermore, recommended and testable performance metrics that can guide technology improvement, i.e., performance targets, do not exist for air sensors. The lack of testing protocols and targets can lead to confusion in the marketplace for both sensor manufacturers/developers and consumers. Without proper guidance, sensor manufacturers/developers may not know which procedures are needed to appropriately test the performance of a sensor for a given application. Consumers may struggle to understand the performance of a sensor and which sensors will appropriately fit their desired application. Additionally, organizations that manage air quality (e.g., air or health agencies) may have difficulty responding to air sensor data that is provided by the public, especially when there is interest in using those data to bring attention to air pollution issues and to influence policy decisions. Without knowledge of how air sensor technologies perform, it is hard to understand the comparability of air sensor data with data from regulatory monitors.

While air sensor technologies are creating significant opportunities to monitor air quality, the variability in data quality creates challenges in understanding sensor performance. Having a consistent approach for evaluating the performance of air sensors benefits all stakeholders as it will provide confidence in data quality and help consumers identify appropriate air sensors for their intended application, encourage innovation and product improvement in the marketplace, and reduce uncertainty about the performance of a given technology. A priority for the U.S. EPA is to support technology development toward data that are of known quality and help establish best practices for the use of air sensors and their data.

1.3 Objectives

The purpose of this report is to provide a standard, streamlined, unbiased approach to testing the performance of fine particulate matter ($PM_{2.5}$) air sensors for non-regulatory supplemental and informational monitoring (NSIM) applications in ambient, outdoor, fixed site environments. NSIM applications (summarized in Table 1-1) are the focus of this report as these areas have been identified as the primary use of air sensors in the U.S.

Table 1-1. NSIM Categories and Specific Examples (adapted from U.S. EPA, 2018)

Category	Definition	Examples
Spatiotemporal Variability	Characterizing a pollutant concentration over a geographic area and/or time	Daily trends, gradient studies, air quality forecasting, citizen science, education
Comparison	Analysis of differences and/or similarities in air pollution characteristics against a threshold value or between different networks, locations, regions, time periods, etc.	Hot-spot detection, data fusion, emergency response, supplemental monitoring
Long-term Trend	Change in a pollutant concentration over a period of (typically) years	Long-term changes, epidemiological studies, model verification

This report provides specific guidance on testing protocols, performance metrics, and target values for those metrics for $PM_{2.5}$ air sensors used in NSIM applications. This guidance combines U.S. EPA expertise in sensor evaluation and application research, expertise of other organizations who administer routine sensor evaluation programs, as well as findings from organizations that are developing similar guidance on sensors. Additionally, this guidance utilizes information gathered from two literature reviews conducted by the U.S. EPA that informed the development of sensor performance targets and testing protocols for NSIM applications. The first review identified the most important performance attributes to characterize instruments used to monitor air pollutants and quantitative performance metrics describing those performance attributes (U.S. EPA, 2018). The second review had a similar objective but examined more recent literature as well as results from field and laboratory sensor performance evaluations (U.S. EPA, 2020).

The specific objectives of this report are as follows:

- Provide a consistent set of testing protocols, metrics, and target values to systematically evaluate the performance of air sensors;
- Provide a consistent framework for communicating performance evaluation results; and
- Help consumers make informed decisions on choosing sensors that might best suit a NSIM application of interest.

Collectively, these objectives will help provide a streamlined framework to understand air sensor performance for NSIM applications. It should be noted that other applications (e.g., mobile monitoring, indoor monitoring, personal exposure monitoring) may require different testing protocols which are not covered in this report.

The intended audience for this report includes potential testing organizations, sensor manufacturers, and sensor developers. It is anticipated that a variety of consumers (e.g., state/local/tribal agencies, federal government agencies, community groups, citizen scientists, academia) will benefit from the consistent presentation of testing results to identify sensor technologies that would be best suited for their NSIM application and understand the performance of the air sensor technologies. Consumers may also choose to conduct these testing protocols.

Results from these testing protocols do not constitute certification or endorsement by the U.S. EPA. It is recommended that testers make the testing reports available on their respective websites to inform consumers on the testing results.

2.0 Performance Testing Protocols for $PM_{2.5}$ Air Sensors

The procedures outlined in this section provide standardized test protocols for evaluating the performance of $PM_{2.5}$ air sensors (also called 'air sensor' and 'sensor' in this report). These procedures only apply to sensors used in NSIM applications in ambient, outdoor, fixed site environments. Two testing procedures are summarized: 1) base testing which involves field evaluations, and 2) enhanced testing which involves laboratory evaluations. Base testing at an ambient, outdoor evaluation site provides information on air sensor performance that is relevant to real-world conditions and allows consumers to predict how the sensor might perform in similar conditions. For more comprehensive sensor performance information, enhanced testing in a controlled laboratory environment allows air sensors to be evaluated over a range of conditions that may be challenging to capture in an ambient, outdoor environment. Additionally, enhanced testing characterizes some parameters that are difficult to test under ambient, outdoor conditions. All testers are encouraged to conduct base testing at a minimum. Enhanced testing is also encouraged although it calls for a controlled laboratory exposure chamber.

For both the base and enhanced testing, at least three (3) identical air sensors should be tested to help consumers understand the out-of-the-box performance and variation that may be present among identical sensors. As a caution, sensor performance can change over time and during the testing procedures. It may be informative to test sensors from multiple production batches provided that the sensors are the same make, model, and firmware version. A separate set of at least three (3) air sensors can be used to conduct base and enhanced testing if tests will be conducted simultaneously, but the sensors should all have the same make, model, and firmware version. If conducting both base and enhanced testing with a single set of sensors, an example approach is shown in Figure 2-1. To make the most effective use of time, the second field deployment (i.e., Field Deployment 2) in the base testing and the aging period between drift evaluation [i.e., Drift (Day1) and Drift (Day 60)] in the enhanced testing can be conducted simultaneously.

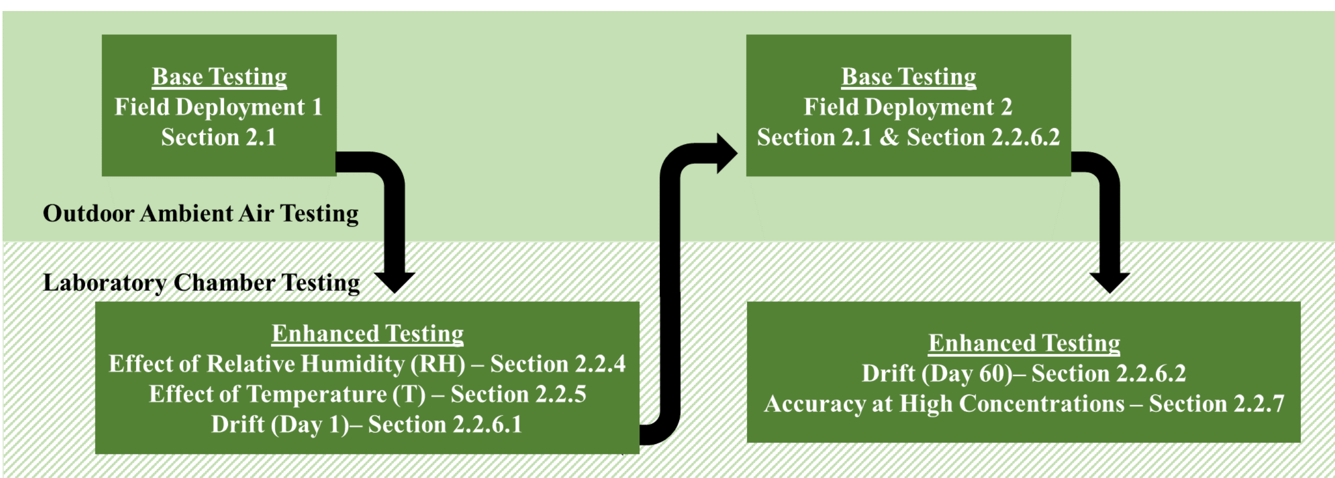


Figure 2-1. Example Approach for Conducting Base and Enhanced Testing of a Single Set of Sensors

2.1 Base Testing

Base testing consists of two (2) field deployments of $PM_{2.5}$ air sensors with collocated FRM/FEM monitors for at least 30 days for each deployment. Testers may set up their own FRM/FEM monitors using guidance and information on ambient air monitoring and monitoring methods available on the U.S. EPA's Ambient Monitoring Technology Information Center (AMTIC) webpage (https://www.epa.gov/amtic, *last accessed 07/25/2020*). FRM/FEM monitors should be calibrated using transfer standards that are certified and NIST traceable. Alternatively, testers may wish to develop relationships with state, local, or tribal air quality agencies to collocate sensors near regulatory FRM/FEM monitors located at existing air quality monitoring sites around the U.S. These sites can be found on the following website: https://www.epa.gov/outdoor-air-quality-data/interactive-map-air-quality-monitors (*last accessed 07/25/2020*).

Base testing should occur at two (2) test sites to give the greatest possible variety in $PM_{2.5}$ variables including $PM_{2.5}$ concentrations and particle sources, types, and size distributions. The combination of field tests should also demonstrate sensor performance over a range of T, RH, and weather conditions that will inform on sensor use across the U.S. For base testing, the sensor and FRM/FEM data will be compared at 24-hour averages which will allow testers to use a variety of technology options. FRM/FEM monitors provide near-equivalent measurements at this time averaging interval which will lead to greater comparability in the reported results. For NSIM applications where high $PM_{2.5}$ concentrations are expected (e.g., wildfire smoke applications), it is recommended that testers conduct base testing in more than two (2) locations and include sites impacted by wildfire smoke and higher $PM_{2.5}$ concentrations.

The procedure in this section outlines the materials and equipment needed, site selection, set up, testing procedure, and data reporting needs to evaluate air sensor performance. To assist testers in ensuring that the requested data before and during the base testing procedure is documented, a checklist is provided in Appendix E. All information for this testing procedure should be recorded in the base testing report (template available in Appendix F). As mentioned previously, it is recommended that testers make the testing report(s) available on their respective websites to inform consumers on the testing results.

2.1.1 Materials and Equipment

The following materials and equipment are needed for this testing procedure:

- Three (3) or more $PM_{2.5}$ air sensors having the same make, model, and firmware version
- Calibrated $PM_{2.5}$ FRM/FEM monitor[*]
- Calibrated RH monitor[†]
- Calibrated T monitor[†]
- Support structures and/or enclosures for air sensors (as recommended by the manufacturer)

[*]The FRM/FEM monitor must be calibrated on site prior to conducting base testing. Additional materials and equipment may be needed to accomplish the calibration. Calibration procedures are outlined in the Quality Assurance Guidance Document 2.12 (U.S. EPA, 2016; herein referred to as QA Document 2.12) and 40 CFR Parts 50, 53, and Appendix A of Part 58. Calibration procedures for continuous FEMs are detailed in the manufacturer's user manual which are approved as part of the FEM designation process. If testing is conducted

at an established regulatory air quality monitoring station with established calibration and quality control procedures, attach or cite the site Quality Assurance Project Plan (QAPP) to the base testing report (Appendix F). Additionally, if testing at an established site, it is recommended to confirm with the site operators (i.e., state/local agency) whether or not the FRM/FEM monitor(s) passed the monthly checks before and after testing the sensors and include this information in the base testing report.

[†]Meteorological monitors should be certified by the manufacturer or calibrated, installed, maintained, and audited according to quality assurance procedures outlined in U.S. EPA's Quality Assurance Handbook for Air Pollution Measurement Systems Volume IV: Meteorological Measurements (U.S. EPA, 2008).

Some additional measurements may be useful in interpreting base testing results. These measurements include: 1) particle size distribution, 2) particle chemical composition (e.g., carbon, nitrogen), and 3) refractive index. Testers may be responsible for collecting these measurements if they choose to include them. The measurements could be particularly useful in understanding the physical reasons behind variations in sensor performance.

Preferably, measurements should be logged internally on each instrument or through a central data acquisition system. If possible, sensors should not be connected to the internet. The main reasons for this are as follows (based on Schneider et al., 2019 and experience):

- Relying solely on internet capabilities may lead to data loss in the event of network outages;
- It is difficult to verify the integrity of the sensor data if the sensor is connected to the internet (e.g., firmware could update during testing). Many consumers want the ability to trace and verify how data is transformed from a raw format to a final format. This can be especially problematic for sensors that rely on machine learning approaches;
- Some consumers may want to use sensors where internet or cellular connections are not available. Consumers may need to know how sensor devices may work in those situations; and
- If a sensor uses a nearby measurement (e.g., FRM/FEM, meteorological, other sensor data) to verify proper operation or correct the data, a consumer may not know how the sensor performs when these data are not available.

It is recognized that not all sensors can log internally or be disconnected from the internet and may stream data to a cloud platform or manufacturer server. If an internet or cellular connection is needed to operate the sensor, this information should be reported, and testers should attest that no data from collocated or nearby FRM/FEMs will be used to manipulate sensor data throughout the data processing procedure for primary testing and reporting. It is recommended that, testers issue a second report with the connectivity, enhanced data processing description, and test results if they believe that many consumers will choose to operate the sensors in such a manner.

In order to properly compare the measurements (FRM/FEM, sensor, RH, T), it is important that the data streams are time aligned. This can be done by adjusting instrument times to a common standard clock [e.g., National Institute of Standards and Technology (NIST) time], carefully checking time stamps when devices are started and stopped, and/or using a common data logger. If data from any instrument is reported as an average, it is also important to understand if the data average is 'time ending' or 'time beginning'. For example, when logging hourly averages, the 07:00 time stamp may reflect data collected between 06:01-7:00 (time ending) or 7:00-7:59 (time beginning). This information should be considered when time aligning data.

For base testing, the sensor and FRM/FEM data will be compared at 24-hour averages which will allow testers to use a variety of technology options. FRM/FEM monitors provide near-equivalent measurements at this time averaging interval which will lead to greater comparability in the reported results. It is recommended that testers issue a supplemental test report of 1-hour data but should note that there may be unquantified variation between FEM instruments at this time averaging interval.

2.1.2 Selecting and Setting Up a Test Site

Potential consumers need information on how well they might expect sensors to perform in the area in which they intend to make measurements. Therefore, testing a sensor's performance over a range of conditions (e.g., T, RH, pollutant concentrations) would be most informative to the widest variety of consumers. Table 2-1 provides recommended criteria for the test sites.

Table 2-1. Test Site Selection Criteria

Base Testing Plan	Location(s)	Season	Goal 24-Hour Average $PM_{2.5}$ Concentration (for at least one day)
Two test sites	Site 1	Climate Region 1	≥ 25 μg/m^3
	Site 2	Climate Region 2	≥ 25 μg/m^3

As shown in Table 2-1, it is recommended that base testing be conducted in a minimum of two (2) locations in two (2) different climate regions across the U.S. (Figure 2-2). Although all test results are valuable, testing under a range of $PM_{2.5}$ concentrations is most informative. Therefore, a goal of at least one day of the 30-day testing period with a 24-hour average $PM_{2.5}$ concentration of at least 25 μg/m^3 is suggested. It is acknowledged that $PM_{2.5}$ concentrations across the U.S. may be unpredictable and even sites that are expected to exceed 25 μg/m^3 may not achieve this level. A level of 25 μg/m^3 is recommended as a goal as this will help ensure statistics are comparable across sites and that testing does not result in a low R^2 due to low $PM_{2.5}$ concentration ranges. For NSIM applications where high $PM_{2.5}$ concentrations are expected (e.g., wildfire smoke applications), it is recommended that testers conduct base testing in more than two (2) locations and include sites impacted by wildfire smoke and higher concentrations.

$PM_{2.5}$ concentrations and particle sources, types, and size distributions can vary across the U.S. as does climate itself; thus, tests conducted in separate climate regions (Figure 2-2) may offer performance information over a range of $PM_{2.5}$ variables and meteorological conditions. When choosing test sites, it is recommended to select sites located in climate regions that are not adjacent for the greatest possible variation in $PM_{2.5}$ variables.

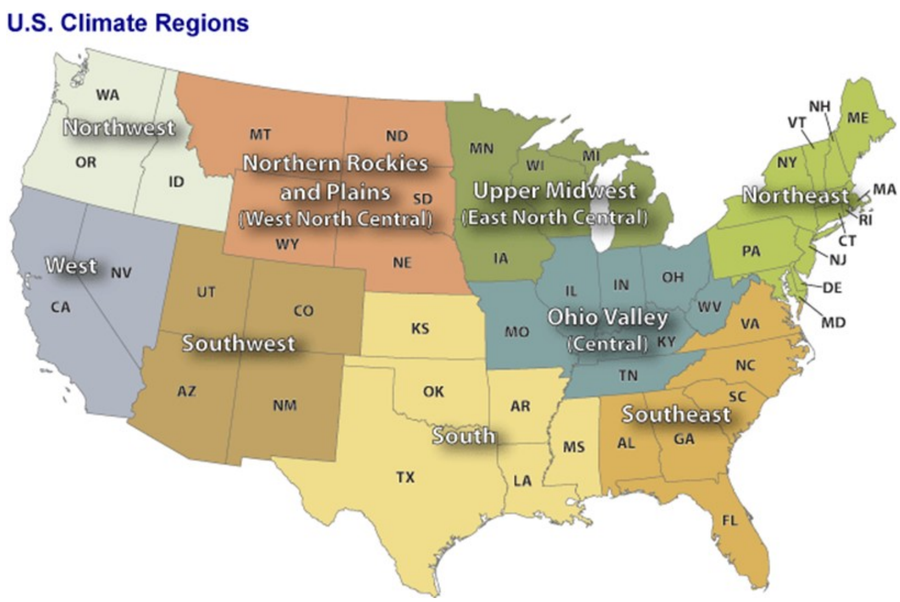


Figure 2-2. U.S. Climate Regions (<u>https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php</u>, *last accessed 07/29/2020*)

Based on historical data, there are a number of sites across the U.S. that should offer the conditions shown in Table 2-1 during certain times of the year. If using an existing ambient air monitoring network site [e.g., National Core Network (NCore), Chemical Speciation Network (CSN), State and Local Air Monitoring Station (SLAMS)], a tester can examine historical air quality data, found on the U.S. EPA AirData website (<u>https://www.epa.gov/outdoor-air-quality-data</u>, *last accessed 07/25/2020*), to see if/when a site is likely to meet the criteria. If using another site that does not have historical air quality data, consult data from the nearest regulatory monitoring site to determine if the site(s) is likely to meet the criteria in Table 2-1. Additional details outlining how to identify sites that will likely meet this criterion and how this criterion was determined are detailed in Appendix B.

Take the following steps when selecting and setting up a test site:

1. Select a test site(s) that meets the criteria in Table 2-1. If using an existing ambient air monitoring network site, record the Air Quality System (AQS) site ID.
2. Record the calibration or certification date for the T and RH monitors and attach a copy of the calibration certificate(s) to the base testing report (Appendix F).
3. If not already set up at a test site, install the FRM/FEM, T, and RH monitors at the test site such that the sampling probe inlet or monitoring path meets the siting criteria in Table 2-2.

Table 2-2. Sampling Probes or Monitoring Path Siting Criteria

Description	Distance (meters)
Height from ground	2 to 15
Horizontal and vertical distance from supporting structures	> 1
Distance from trees	> 10*
Distance from roadways	> 10 to 250†

*Should be greater than 20 meters from the tree(s) dripline and must be 10 meters from the dripline when the tree(s) act as an obstruction (see 40 CFR Part 58, Table E-4 of Appendix E).
†The roadway average daily traffic, vehicles per day determines the minimum distance (see 40 CFR Part 58, Table E-1 of Appendix E).

2.1.3 Setting Up the Air Sensors

Take the following steps when setting up the air sensors for base testing:

1. Verify that there are at least three (3) $PM_{2.5}$ air sensors of the same make, model, and firmware version. The firmware version should not be updated during the testing. Use sensors in the same condition as they were received from the manufacturer and do not modify any manufacturer calibration(s).
2. Disconnect the sensors from internet access. Ideally, data should be stored locally on the sensors (such as on a local data card) or on a common datalogger. If an internet or cellular connection is necessary for sensor operation, data from either collocated or nearby FRM/FEM monitors should not be used by the sensors during this testing procedure.
3. In the base testing report (Appendix F), record information about the equipment and set-up, to the extent possible, including the following:

    - Parameters measured (e.g., pollutant(s), T, RH, dew point) and units
    - Sampling time interval (e.g., 1-minute, 15-minute, 1-hour, 24-hour)
    - Data storage and transmission method(s), including:
        - Where the data are stored (e.g., local data card, transmitted to cloud system)
        - If applicable, where the data are transmitted (e.g., manufacturer's cloud server)
        - Form of data stored (e.g., raw data, corrected or cleaned data)
    - Data correction approach (if applicable), including:
        - Procedure used to correct the data including: [a] how the data are corrected (e.g., manufacturer derived multilinear correction), [b] variables used to correct the data (e.g., RH, T), [c] where the correction variable(s) comes from (e.g., on-board RH sensor), and [d] how the data are validated or calibrated (e.g., RH sensor is calibrated by the manufacturer)
        - If the way data are corrected does not change and is static, record this information and any mathematical approaches used

  - If the way data are corrected changes or is a dynamic process, record the following: (a) when the process changes, (b) why the process changes, (c) how/where changes are recorded, and (d) how the correction method is validated
- Data analysis/data correction scripts (e.g., Jupyter Notebook, R Markdown)
- Location of final reported data and its format (e.g., website shows raw data and corrected data on user interface, data provided as .csv, expanded definitions of data headers)

4. Install air sensors at the test site using the ideal setup guidance summarized in Table 2-3.
5. Include photographs that clearly show the entire equipment setup at the test site, and document distances, in the base testing report (Appendix F).

Table 2-3. Guidance on Air Sensor Setup at Testing Site

Recommendations	Cautions
• Mount sensors within 20 meters horizontal of the FRM/FEM monitor • Mount sensors in a location where they are exposed to unrestricted air flow • Ensure the air sampling inlet for the sensors are within a height of ± 1 meter vertically of the air sampling inlet of the FRM/FEM monitor • Mount identical sensors ~1 meter apart from each other • If necessary, install sensors within a weather-protective shelter/enclosure that maintains ample air flow around the sensor (as recommended by manufacturer)	• Do not place sensors near structures/objects that can affect air flow to the sensor OR block the sensor air intake (e.g., against a wall, near a vent, or on the ground blocking the inlet) • Do not place sensors near structures/objects that can alter T or RH near the sensor (e.g., vents, exhausts) • Do not place sensors near sources/sinks that can alter pollutant concentrations (e.g., idling cars, smoking) • Do not place sensors in locations with risk of vibration, electrical shock, or other potential hazards

2.1.4 Conduct Base Testing

The step-by-step procedure for conducting the base testing is as follows:

1. Record the calibration date of the FRM/FEM monitor. Calibration should be conducted after the monitor is in place at the test site, not before. If the FRM/FEM monitor requires calibration, follow the procedures as outlined in QA Document 2.12 and 40 CFR Parts 50, 53, and Appendix A of Part 58. Calibration procedures for continuous FEMs are detailed in the manufacturer's user manual which are approved as part of the FEM designation process.
2. Verify that the system(s) for data logging and data storage will collect all equipment data and store it in a way that can be accessed later. Make sure that there is enough storage capacity available to prevent older data from being overwritten and allow new data to be stored.
3. Use sensors in the same condition as they were received from the manufacturer and do not modify any manufacturer calibration(s). The firmware version should not be updated during testing.
4. Provide a warm-up and stabilization period for all equipment as specified by the manufacturer.
5. Verify that all equipment is reporting measurements.

6. Conduct a one-point flow rate verification check on the FRM/FEM monitor (see procedures outlined in QA Document 2.12 and the manufacturer's user manual) and record the date of the check.
7. Allow all equipment to run for at least 30 consecutive days. All equipment should be running during the same time period to allow for comparable results.
8. Follow the manufacturer's maintenance recommendations, as applicable, for all equipment (e.g., sensors, FRM/FEM) throughout testing. Record and report all maintenance or troubleshooting performed, including dates/times, on the instruments (e.g., power cycling, FRM/FEM flow rate verification check).
9. Record and report the rationale for missing or invalidated data. For a full 30 consecutive day run, at least 75% uptime, with all instruments reporting is ideal. This corresponds to all equipment reporting at least 23 valid 24-hour pairs of time-matched data points over the course of the 30-day deployment.
    a. If the sensor fails irreparably before the 30-day deployment is complete, another sensor should not be substituted. In addition, the sensor should not be sent back to the manufacturer for repairs without restarting the testing procedure. A preliminary report could present results with documentation as to why the sensors failed as these details may be useful to potential consumers. Testing can be restarted with three (3) sensors.
    b. Occasionally, low uptime or a deployment period of less than 30 days might occur, for example, due to an unplanned electrical outage or weather event (e.g., hurricane, tornado). In those instances, the dates and reasons for missing data should be recorded. In these scenarios, ideally testing would continue/resume until at least 23 valid 24-hour pairs of time-matched data points are collected.
    c. If data from any piece of equipment is not available during each 24-hour sampling period, record and report the reason (e.g., outage, maintenance).
    d. Additionally, if any of the data are invalidated due to QC criteria, record the reason and criteria used. FRM/FEM instruments have more established QC criteria (visit the AMTIC webpage at https://www.epa.gov/amtic, *last accessed 07/25/2020*). QC criteria for the sensor may be available from the manufacturer or may be developed as part of these tests. General information on how the U.S. EPA manages data quality can be found at https://www.epa.gov/quality (*last accessed 12/07/2020*). Reporting QC criteria for the sensor is strongly recommended as this information is beneficial for consumers.
10. Select a test site for the second field deployment based on the test site criteria outlined in Table 2-1.
11. Repeat Sections 2.1.2 to 2.1.4 for the second field deployment using the sensors from the first field deployment, if possible. A separate base testing report should be generated for the second field deployment.

2.2 Enhanced Testing

Enhanced testing consists of testing the sensors in a controlled laboratory environment to understand the effects of RH and T, and other important parameters including drift and measurement accuracy at higher concentration levels. Such tests are particularly valuable in controlling conditions so that results can be repeatable and reproducible. Further, enhanced testing allows sensors to be tested at concentrations that are rarely encountered in the field yet important to understand (e.g., performance during wildfire smoke conditions). An overview of enhanced testing procedures is shown in Figure 2-3. The procedure in this section outlines the materials and equipment needed, set up, testing procedure, and data reporting needs to evaluate air sensor performance.

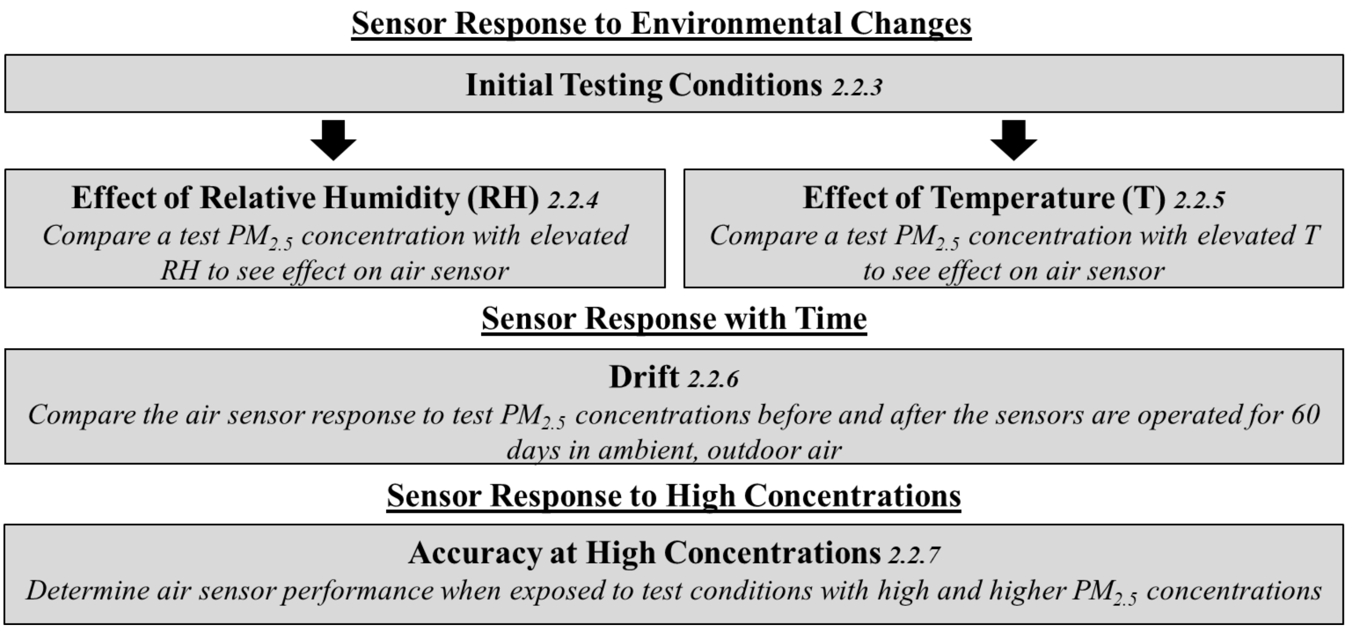


Figure 2-3. Overview of the Enhanced Testing Procedures

In the ambient, outdoor environment, $PM_{2.5}$ concentration and particle composition can vary substantially. Particles have different chemical composition, a variety of shapes, and particle size is a distribution of aerodynamic diameters which vary over several orders of magnitude. Moreover, the specific properties of the particles (e.g., optical characteristics, aerodynamic behavior, electrical properties, potential health effects) can also vary substantially. As a result, it is impossible to generate an aerosol in the laboratory which will perfectly simulate those encountered under field conditions. Therefore, testers may prefer to conduct additional base testing (field deployments) rather than enhanced testing (laboratory testing).

The design of an experimental setup for aerosol generation is a balance of several factors including capital and operating cost, space, system complexity, level of effort, and system maintenance. At a minimum, methods are necessary for generating particles of known size distribution, composition, and concentration. The ability to measure the size distribution, composition, and concentrations inside the chamber is also ideal for verifying these traits of the generated aerosol. The ability to maintain these parameters over extended testing periods is also important. Lastly, the ability to vary testing conditions (specifically T and RH) is also important.

Maintaining stable particle delivery (including particle size distribution, composition, and concentration) for long periods of time can be difficult. This test protocol allows for higher time resolution measurements (1-hour, 10-minute, 1-minute averages) to be compared. This will require the use of an FEM instrument. However, there are often unquantified variations between FEM instruments at higher time resolution and some instruments may be more stable than others, especially at low concentration. Assessments of ambient network FRM/FEM monitors can be found here: https://www.epa.gov/amtic/amtic-ambient-air-monitoring-assessments (*last accessed, 12/09/2020*). Collecting an integrated filter sample may be helpful in verifying the high-time resolution mass concentration FEM measurement.

At this time, there is no consensus among organizations about which test aerosols would be most informative for use in the enhanced testing, so no recommendations have been made at this time. Test aerosols that may be reasonable proxies for ambient aerosols would be most informative for consumers.

Potential proxies include:

- Ammonium sulfate;
- Ammonium nitrate;
- Oleic acid or sucrose; or
- Smoke from smoldering wood chips or other biomass material.

Other proxies currently in use include polystyrene latex spheres and dust. The type of particles generated for the tests and their particle size distribution, chemical composition, and refractive index, if known, should be recorded on the enhanced testing report. It should be noted that the type of particles generated can influence the results of the enhanced tests. For example, polystyrene latex spheres are not hygroscopic and will tend to underestimate the sensitivity of sensor measurements to high RH, as latex spheres have lower water uptake than many ambient aerosols. Additionally, sensors may not be able to detect all particle sizes, therefore testers may want to evaluate this by testing different particle sizes.

To assist testers in ensuring that the requested data before and during the enhanced testing procedure is documented, a checklist is provided in Appendix G. All information for this testing procedure should be recorded in an enhanced testing report (Appendix H). As mentioned previously, it is recommended that testers make the testing report(s) available on their respective websites to inform consumers.

2.2.1 Materials and Equipment

The following materials and equipment are needed for this testing procedure:

- Three (3) or more $PM_{2.5}$ air sensors having the same make, model, and firmware version[*]
- Calibrated $PM_{2.5}$ FEM monitor[†, ‡]
- Exposure chamber that can control environmental conditions
- $PM_{2.5}$ test aerosol generator system[§]
- Zero air generator[#]
- Dynamic calibration system
- Calibrated RH monitor[**]
- Calibrated T monitor[**]

[*]Sensors can be the same ones used in the base testing procedure.

[†]The FEM monitor should be calibrated on-site prior to conducting enhanced testing and additional materials and equipment may be needed to accomplish the calibration. Calibration procedures are outlined in QA Document 2.12 and are detailed in the manufacturer's user manual which are approved as part of the FEM designation process. If testing is conducted at an established sensor testing facility with established calibration and QC procedures, attach or cite the QAPP to the enhanced testing report (Appendix H).

‡Limitations on maintaining stable particle delivery require FEM measurements at a 1-hour time resolution or higher. Measurements will be most stable with an optical-based FEM monitor. It is recommended to also collect a filter sample to validate the FEM measurement.

§The $PM_{2.5}$ test aerosol generator system should be capable of producing stable levels of $PM_{2.5}$ over the range of concentration values and durations specified in the enhanced testing.

#Zero air generator should be calibrated using transfer standards that are certified and NIST traceable and include expiration dates.

**Meteorological monitors should be certified by the manufacturer or calibrated, installed, maintained, and audited according to quality assurance procedures outlined in U.S. EPA's Quality Assurance Handbook for Air Pollution Measurement Systems Volume IV: Meteorological Measurements (U.S. EPA, 2008).

As with base testing, a number of additional measurements may be useful in understanding the physical reasons behind variations in sensor performance, including: 1) particle size distribution, 2) particle chemical composition (e.g., carbon, nitrogen), and 3) refractive index. Collection of an integrated filter sample may be helpful in verifying the high-time resolution mass concentration measurement from the FEM. Testers may be responsible for collecting these measurements if they choose to include these as part of testing.

Preferably, measurements should be logged internally on each instrument or through a central data acquisition system. If possible, sensors should not be connected to the internet; please see Section 2.1.1 which describes the reasoning for this. It is recognized that not all sensors can log internally or be disconnected from the internet and may stream data to a cloud platform or manufacturer server. If an internet or cellular connection is needed to operate the sensor, this information should be reported.

In order to properly compare the measurements (FEM, sensor, RH, T), it is important that the data streams are time aligned. This can be done by adjusting instrument times to a common standard clock (e.g., NIST), carefully checking time stamps, and/or using a common data logger. If data from any instrument is reported as an average, it is also important to understand if the data average is 'time ending' or 'time beginning'. For example, when logging hourly averages, the 07:00 time stamp may reflect data collected between 06:01-7:00 (time ending) or 7:00-7:59 (time beginning). This information should be considered when time aligning data.

The exposure chamber should meet the following criteria:

- Ability to control, maintain, and monitor T, RH, and $PM_{2.5}$ concentrations. Approximate recommended ranges based on testing conditions outlined in this report: T – 19 to 41°C; RH – 35 to 90%; $PM_{2.5}$ – 5 to 280 μg/m^3.
- Ability to maintain the particle size distribution, composition, and concentration.
- Ability to maintain atmospheric pressure by balancing the incoming flow with the sampling and vent flow.
- Allows for air to be well-mixed.
- Capable of accommodating three (3) or more air sensors.
- Sampling ports should not be obstructed and allow for sufficient sampling flow.

- The particle generation system should be positioned above the exposure chamber and connected with as few bends as possible to prevent particle loss and build-up before particles enter the chamber.

- The $PM_{2.5}$ FEM monitor should be mounted directly below the sampling chamber to reduce particle loss in the sampling line due to bends.

If possible, provide documentation on the chamber specifications, characterization, and any laboratory intercomparison.

2.2.2 Equipment Set Up in Exposure Chamber

To properly set up equipment in the exposure chamber, take the following steps:

1. Check that all equipment is properly calibrated. Record the calibration date for each piece of equipment (as applicable). Calibration procedures for each continuous FEM are detailed in the manufacturer's user manual which are approved as part of the FEM designation process.
2. Conduct a one-point flow rate verification check on the FEM monitors (see procedures outlined in QA Document 2.12 and the manufacturer's user manual) and record the date of the check.
3. Verify that there are at least three (3) $PM_{2.5}$ air sensors of the same make, model, and firmware version. The firmware version should not be updated during the testing. Use sensors in the same condition as they were received from the manufacturer and do not modify manufacturer calibration(s).
4. Disconnect the sensors from internet access (if possible). Ideally, data should be stored locally on the sensors (such as on a local data card). If an internet or cellular connection is necessary for sensor operation, data from either collocated or nearby FEMs should not be used during this testing procedure.
5. In the enhanced testing report (Appendix H), record information about the equipment and set-up, to the extent possible, including the following:
    - Parameters measured (e.g., pollutant(s), T, RH, dew point) and units
    - Sampling time interval (e.g., 1-minute, 15-minute, 1-hour)
    - Data storage and transmission method(s), including:
        - Where the data are stored (e.g., local data card, transmitted to cloud system)
        - If applicable, where the data are transmitted (e.g., manufacturer's cloud server)
        - Form of data stored (e.g., raw data, corrected or cleaned data)
    - Data correction approach (if applicable), including:
        - Procedure used to correct the data including: [a] how the data are corrected (e.g., manufacturer derived multilinear correction), [b] variables used to correct the data (e.g., RH, T), [c] where the correction variable(s) comes from (e.g., on-board RH sensor), and [d] how the data are validated or calibrated (e.g., RH sensor is calibrated by the manufacturer)
        - If the way data are corrected does not change and is static, record this information and any mathematical approaches used
        - If the way data are corrected changes, or is a dynamic process, record the following: (a) when the process changes, (b) why the process changes, (c) how/where changes are recorded, and (d) how the correction method is validated
    - Data analysis/data correction scripts (e.g., Jupyter Notebook, R Markdown)

- Location of the final reported data and its format (e.g., website shows raw data and corrected data on user interface, data provided as .csv, expanded definitions of data headers)

6. Provide a warm-up and stabilization period for all equipment as specified by the manufacturer.
7. Verify that all equipment is reporting measurements.
8. Document the particle size distribution and chemical composition of the particles used in the aerosol generator system.
9. Throughout testing, follow the manufacturer's maintenance recommendations, as applicable, for all equipment (e.g., sensors, FEM). Record and report all maintenance or troubleshooting performed, including dates/times, on the instruments (e.g., power cycling, FEM flow rate verification check).

2.2.3 Initial Testing Conditions

Take the following steps to begin the enhanced testing procedure:

1. Supply the exposure chamber with the conditions shown in Table 2-4.
2. Allow all measurements to stabilize within the tolerances shown in Table 2-4.
3. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on enhanced testing duration and data time averaging is provided in Appendix B.
   a. $PM_{2.5}$ concentration from each sensor (µg/m^3)
   b. FEM $PM_{2.5}$ concentration (µg/m^3)
   c. RH (%)
   d. T (°C)

Table 2-4. Initial Testing Conditions

Parameter	Reference Setpoint
$PM_{2.5}$ Concentration	35 µg/m^3 ± 5%
T	20°C ± 1°C
RH	40% ± 5%

2.2.4 Effect of Relative Humidity (RH)

To determine the effect of elevated RH on sensor performance, take the following steps:

1. Repeat the procedure outlined in Section 2.2.3.
2. Supply the exposure chamber with the conditions in Table 2-5.
3. Allow the measurements to stabilize within the tolerances shown in Table 2-5.
4. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on enhanced testing duration and data time averaging is provided in Appendix B.
   a. $PM_{2.5}$ concentration from each sensor (µg/m^3)
   b. FEM $PM_{2.5}$ concentration (µg/m^3)

c. RH (%)
d. T (°C)

Table 2-5. Elevated RH Test Conditions

Parameter	Reference Setpoint
$PM_{2.5}$ Concentration	35 µg/m^3 ± 5%
T	20°C ± 1°C
RH	85% ± 5%

2.2.5 Effect of Temperature (T)

To determine the effect of elevated T on sensor performance, take the following steps.:

1. Repeat the procedure outlined in Section 2.2.3.
2. Supply the exposure chamber with the conditions in Table 2-6.
3. Allow the measurements to stabilize within the tolerances shown in Table 2-6.
4. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on enhanced testing duration and data time averaging is provided in Appendix B.
   a. $PM_{2.5}$ concentration from each sensor (µg/m^3)
   b. FEM $PM_{2.5}$ concentration (µg/m^3)
   c. RH (%)
   d. T (°C)

Table 2-6. Elevated T Test Conditions

Parameter	Reference Setpoint
$PM_{2.5}$ Concentration	35 µg/m^3 ± 5%
T	40°C ± 1°C
RH	40% ± 5%

2.2.6 Drift

A summary of the entire drift testing procedure is shown in Figure 2-4.

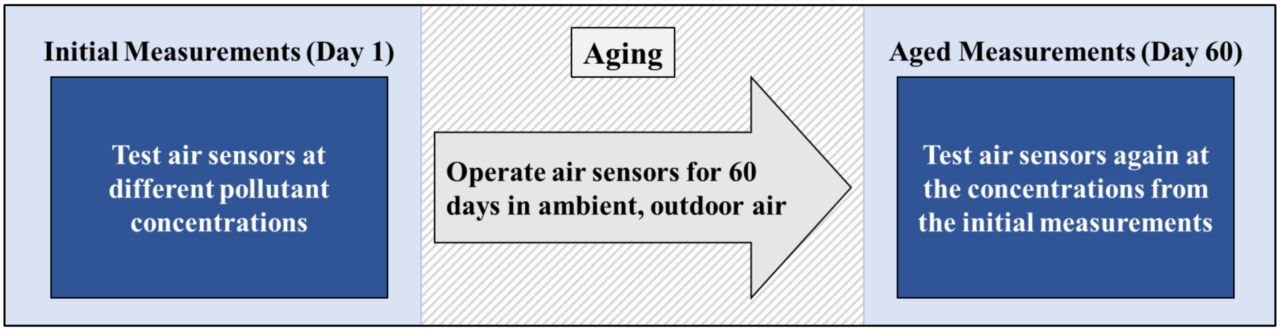


Figure 2-4. Drift Testing to Determine Changes After 60 days or More of Continuous Operation

2.2.6.1 Drift (Day 1) – Low and Mid Concentration Drift

To assess the drift, begin by testing at low and mid level $PM_{2.5}$ concentrations to assess sensor performance on Day 1. To do so, take the following steps:

1. Supply the exposure chamber with the conditions shown in Table 2-7.
2. Allow the measurements to stabilize within the tolerances shown in Table 2-7.
3. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on enhanced testing duration and data time averaging is provided in Appendix B.
    a. $PM_{2.5}$ concentration from each sensor (µg/m^3)
    b. FEM $PM_{2.5}$ concentration (µg/m^3)
    c. RH (%)
    d. T(°C)
4. Supply the exposure chamber with the conditions shown in Table 2-8.
5. Allow the measurements to stabilize within the tolerances shown in Table 2-8.
6. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on enhanced testing duration and data time averaging is provided in Appendix B.
    a. $PM_{2.5}$ concentration from each sensor (µg/m^3)
    b. FEM $PM_{2.5}$ concentration (µg/m^3)
    c. RH (%)
    d. T (°C)

Table 2-7. Low Concentration Drift Test Conditions

Parameter	Reference Setpoint
$PM_{2.5}$ Concentration	10 µg/m^3 ± 10%
T	20°C ± 1°C
RH	40% ± 5%

Table 2-8. Mid Concentration Drift Test Conditions

Parameter	Reference Setpoint
$PM_{2.5}$ Concentration	35 µg/m^3 ± 5%
T	20°C ± 1°C
RH	40% ± 5%

2.2.6.2 Drift (Day 60) to Evaluate Sensor Aging

To assess sensor drift over a 60-day period, take the following steps:

1. Operate the sensors in ambient, outdoor air for at least a consecutive 60-day period.
2. Following the 60-day period*, repeat the procedure in Section 2.2.6.1 with the aged sensors.

*The 60-day drift was chosen to balance the needs for a sufficient length of time in order to measure potential drift with the need to be unduly burdensome. It may be informative to repeat the drift test as sensors age providing additional data points at periodic intervals up to the expected lifespan of the sensor.

2.2.7 Accuracy at High Concentrations

To evaluate sensor accuracy at high $PM_{2.5}$ concentrations, take the following steps:

1. Supply the exposure chamber with the conditions in Table 2-9, with the high $PM_{2.5}$ concentration of 150 µg/m^3.
2. Allow the measurements to stabilize within the tolerances shown in Table 2-9.
3. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on enhanced testing duration and data time averaging is provided in Appendix B.
    a. $PM_{2.5}$ concentration from each sensor (µg/m^3)
    b. FEM $PM_{2.5}$ concentration (µg/m^3)
    c. RH (%)
    d. T (°C)
4. Repeat Steps 1 to 3 with the higher $PM_{2.5}$ concentration of 250 µg/m^3.

Table 2-9. High and Higher $PM_{2.5}$ Concentration Test Conditions

Parameter	Reference Setpoint
$PM_{2.5}$ Concentration (High)*	150 µg/m^3 ± 5%
$PM_{2.5}$ Concentration (Higher)*	250 µg/m^3 ± 5%
T	20°C ± 1°C
RH	40% ± 5%

*The high concentration value (150 µg/m^3) represents 1-hour average concentrations often measured under ambient conditions in some U.S. locations. The higher concentration value (250 µg/m^3) represents 1-hour average concentrations often measured in wildfire smoke impacted areas. Testing at the higher concentration value is optional but may be particularly useful for consumers interested in measurements during smoke condition.

3.0 Performance Metrics and Supporting Calculations for Evaluating $PM_{2.5}$ Air Sensors

Performance metrics are parameters used to describe data quality. There are a number of metrics that can aid in understanding the performance of a sensor device. For the base and enhanced testing protocols (outlined in Section 2.0), this section presents recommended performance metrics along with supporting calculations to evaluate the performance of $PM_{2.5}$ air sensors. The recommended metrics are deemed highly informative to understanding sensor performance and data quality. Some of these metrics are defined in multiple ways in the current sensor literature, so it is important to use the equations outlined here for comparability. Any deviations from these calculation methods should be clearly documented. Table 3-1 provides an abbreviated summary of the performance metrics. Full definitions of these metrics can be found in Appendix A; additional supporting information detailing how these metrics and descriptions were developed can be found in Appendix C.

The performance metrics were selected based on:

- Discussions during the 2018 workshop on "Deliberating Performance Targets for Air Quality Sensors" (Williams et al., 2019);
- Performance specifications for FRM/FEM monitors (40 CFR Part 53, Table B-1 to Subpart B);
- The U.S. EPA findings on air sensor evaluations (https://www.epa.gov/air-sensor-toolbox/evaluation-emerging-air-sensor-performance, *last accessed 09/19/2020*; U.S. EPA, 2015, 2020a, and 2020b);
- South Coast Air Quality Management District Air Quality Sensor Performance Evaluation Center (AQ-SPEC) sensor field evaluations (http://www.aqmd.gov/aq-spec/evaluations/summary-pm, *last accessed 12/08/2020*; SCAQMD, 2016);
- Reviews of data quality levels published in peer-reviewed literature (U.S. EPA, 2018; U.S. EPA 2020); and
- Comparison to other organizations developing sensor standards/certification programs [e.g., People's Republic of China Ministry of Environment and Ecology (MEE)].

It should be noted that the detection limit (DL) is often an important performance metric to ensure that a device can obtain measurements at the low end of the concentration range anticipated at a monitoring location. Based on literature reviews and reviews of sensor evaluation programs, the U.S. EPA considered several approaches to measure DL. However, at this time, we are not confident in a single methodology that will yield consistent and reproducible results for a variety of sensor devices; therefore, DL was not included as a performance metric. However, testers are still encouraged to provide the DL specified by the manufacturer as part of the test report. Additional discussion on this topic is available in Appendix B.

This section further discusses each recommended performance metric and presents details on how each should be calculated.

Table 3-1. Summary of Recommended Performance Metrics for $PM_{2.5}$ Air Sensors

Test Type	Metric	Description
Base Testing	Precision	Variation around the mean of a set of measurements reported concurrently by three or more sensors of the same type collocated under the same sampling conditions. Precision is measured here using the standard deviation (SD) and coefficient of variation (CV).
	Bias	The systematic (non-random) or persistent disagreement between the concentrations reported by the sensor and reference instruments. Bias is determined here using the linear regression slope and intercept.
	Linearity	A measure of the extent to which the measurements reported by a sensor are able to explain the concentrations reported by the reference instrument. Linearity is determined here using the coefficient of determination (R^2).
	Error	A measure of the disagreement between the pollutant concentrations reported by the sensor and the reference instrument. Error is measured here using the root mean square error (RMSE) and normalized root mean square error (NRMSE).
	Exploring Meteorological Effects	A graphical exploration to look for a positive or negative measurement response caused by variations in ambient temperature, relative humidity, or dew point, and not by changes in the concentration of the target pollutant.
Enhanced Testing	Precision	See definition above.
	Bias	See definition above.
	Linearity	See definition above.
	Error	See definition above.
	Effect of Relative Humidity (RH)	A positive or negative measurement response caused by variations in RH and not by changes in the concentration of the target pollutant.
	Effect of Temperature (T)	A positive or negative measurement response caused by variations in ambient T and not by changes in the concentration of the target pollutant.
	Drift	A change in the response or concentration reported by a sensor when challenged by the same pollutant concentration over a period of time during which the sensor is operated continuously.
	Accuracy at High Concentrations	A measure of the agreement between the pollutant concentrations reported by the sensor and the reference instrument during high concentration levels.

3.1 Base Testing Calculations

As a reminder, in order to properly compare the measurements (FRM/FEM, sensor, RH, T), it is important that the data streams are time aligned. This can be done by adjusting instrument times to a common standard clock (e.g., NIST time), carefully checking time stamps, and/or using a common data logger.

If data from any instrument is reported as an average, it is also important to understand if the data average is 'time ending' or 'time beginning'. For example, when logging hourly averages, the 07:00 time stamp may reflect data collected between 06:01-7:00 (time ending) or 7:00-7:59 (time beginning). This information should be considered when time aligning data.

Additionally, FRMs typically run on local standard time from midnight to midnight for the scheduled day, which may not be daily. FEMs operate every hour of every day except during periods of maintenance. This should also be considered when time aligning data.

3.1.1 Daily Averages

For base testing, performance metrics are calculated from daily (24-hour) averaged data. Any FRM/FEM, sensor, RH, and/or T data collected as sub-daily time intervals will first need to be averaged up to daily averages (Eq. 1). In calculating these averages, a 75% data completeness requirement for each 24-hour interval should be imposed. For example, a $PM_{2.5}$ sensor recording concentration measurements every hour would require a minimum of 18 valid measurements in order to calculate a valid 24-hour averaged concentration [i.e., (18/24) * 100% = 75%].

$$x_{kdj} = \frac{1}{n}\sum_{i=1}^{n} c_{ij} \qquad \text{Eq. 1}$$

where:

x_{kdj} = 24-hour averaged measurement *k* for day *d* and instrument *j* (μg/m^3, °C, % RH)

n = number of instrument measurements per 24-hour period

c_{ij} = measurement from instrument *j* for time *i* of the 24-hour period (μg/m^3, °C, % RH)

As a reminder, x_{kdj} is considered a valid 24-hour average if at least 75% of the expected data points over a 24-hour period are reported.

3.1.2 Deployment Averages

The average concentrations and meteorological parameters for the entire 30-day deployment should be reported. Deployment averaged measurements should be calculated from valid 24-hour averaged data (Eq. 2) for each field test.

$$\overline{x_k} = \frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{N}\sum_{d=1}^{N} x_{dj}\right] \qquad \text{Eq. 2}$$

where:

$\overline{x_k}$= deployment averaged measurement *k* for a field test (μg/m^3, °C, % RH)

M = number of identical instruments operated simultaneously during a field test

N = number of 24-hour periods during which all identical instruments are operating and returning valid averages over the duration of the field test

x_{dj} = valid 24-hour averaged measurement for day *d* and instrument *j* (μg/m^3, °C, % RH)

3.1.3 Precision

Precision between identical sensors should be characterized by two metrics: standard deviation (SD) between measurements (Eq. 3) and coefficient of variation (CV; Eq. 4). These metrics should be calculated for the base testing field deployments using data during which all identical sensors are operating and returning valid 24-hour averaged measurements.

$$\text{SD} = \sqrt{\frac{1}{(N \times M) - 1}\sum_{j=1}^{M}\left[\sum_{d=1}^{N}(x_{dj} - \overline{x_d})^2\right]} \qquad \text{Eq. 3}$$

where:

SD = standard deviation of 24-hour averaged sensor $PM_{2.5}$ concentration measurements (μg/m^3)

M = number of identical sensors operated simultaneously during a field test

N = number of 24-hour periods during which all identical instruments are operating and returning valid averages over the duration of the field test

x_{dj} = 24-hour averaged sensor $PM_{2.5}$ concentration for day *d* and sensor *j* (μg/m^3)

$\overline{x_d}$ = 24-hour averaged sensor $PM_{2.5}$ concentration for day *d* (μg/m^3)

$$\text{CV} = \frac{\text{SD}}{\overline{x}} \times 100 \qquad \text{Eq. 4}$$

where:

CV = coefficient of variation (%)

SD = standard deviation of 24-hour averaged sensor $PM_{2.5}$ concentration measurements (μg/m^3)

$\overline{x}$ = deployment averaged sensor $PM_{2.5}$ concentration for a field test (μg/m^3)

3.1.4 Bias and Linearity

A simple linear regression model can demonstrate the relationship between paired 24-hour averaged sensor and FRM/FEM $PM_{2.5}$ measurements. Using a simple linear regression model (y = mx + b) with the sensor $PM_{2.5}$ measurements as the dependent variable (y) and the FRM/FEM $PM_{2.5}$ measurements as the independent variable (x), calculate the slope (m), intercept (b), and the coefficient of determination (R^2).

A simple linear regression model for each identical sensor (with corresponding graphical figures) are recommended. Comparison of the figures and these metrics across identical sensors can be helpful in further visualizing sensor precision (Section 3.1.3). Sensors with very similar regression models and higher R^2 values are typically more precise than those with different regression models and lower R^2 values.

A function for determining a simple linear regression model is well established in many software packages (e.g., Excel, R) and readily available using the U.S. EPA Excel-based Macro Analysis Tool (https://www.epa.gov/air-sensor-toolbox/air-sensor-collocation-macro-analysis-tool, *last accessed 07/25/2020*), thus the equations are not presented here. Caution should be taken to appropriately select the FRM/FEM measurements as the independent (x) variable and sensor measurements as the dependent (y) variable when using these tools.

3.1.5 Error

The root mean square error (RMSE) is one metric that can be used to help understand the error associated with sensor $PM_{2.5}$ concentration measurements. The interpretation of this value is slightly more straightforward because it is calculated in concentration units. Using data during which all sensors are reporting valid 24-hour averaged measurements, the sensor and FRM/FEM $PM_{2.5}$ measurement calculations are compared (Eq. 5). This equation assumes only one FRM/FEM instrument will be running. If multiple FRM/FEM instruments are running, separate testing reports can be generated for each.

$$\text{RMSE} = \sqrt{\frac{1}{N \times M}\sum_{j=1}^{M}\left[\sum_{d=1}^{N}(x_{dj} - R_d)^2\right]} \qquad \text{Eq. 5}$$

where:

RMSE = root mean square error (μg/m^3)

N = number of 24-hour periods during which all identical instruments are operating and returning valid averages over the duration of the field test

M = number of identical sensors operated simultaneously during a field test

x_{dj} = valid 24-hour averaged sensor $PM_{2.5}$ concentration for day d and instrument j (µg/m^3)

R_d = valid 24-hour averaged FRM/FEM $PM_{2.5}$ concentration for day d (µg/m^3)

As a caution, RMSE is not defined in a consistent way throughout available resources. It has commonly been defined in two ways: 1) describing the difference between a measurement and the true value, and 2) describing the difference between a measurement and a linear regression best fit line of a measurement and a corresponding true value. In this report, RMSE is defined as the error between the sensor measurements and the reference instrument measurements or true values (see Eq. 5). This approach is presumed to provide the best indication of out-of-the-box sensor performance and the error that can be expected prior to any data corrections. Further, this approach is how RMSE is commonly calculated in air sensor literature to date.

The normalized root mean square error (NRMSE) is also included as a metric to account for testing conditions where the ambient concentrations may be much higher than typical U.S. ambient levels (e.g., wildfires). The RMSE calculated value (Eq. 5) is normalized using the average of the valid 24-hour averaged FRM/FEM $PM_{2.5}$ concentrations over the testing period (Eq. 6).

$$\text{NRMSE} = \frac{\text{RMSE}}{\overline{R_d}} \times 100 \qquad \text{Eq. 6}$$

Where:

NRMSE = normalized root mean square error (%)

RMSE = root mean square error as calculated in Eq. 5 (µg/m^3)

$\overline{R_d}$ = valid 24-hour averaged FRM/FEM $PM_{2.5}$ concentration over the entire testing period (µg/m^3)

3.1.6 Exploring Effect of Meteorology

Research suggests that meteorology [specifically T, RH, and dew point (DP)] can influence the performance of currently available $PM_{2.5}$ sensor technologies (U.S. EPA, 2015; Jayaratne et al., 2018; Zheng et al., 2018; Feenstra et al., 2019; AQ-SPEC PM sensor evaluations). There are several ways to investigate the potential influence using data from the field tests but, no single plot has proven useful in visualizing these effects for all sensor types. Here, several graphical ways to plot the data are suggested to try to understand the effect of meteorology. Additional ways may exist. Testers are encouraged to illustrate the effects of meteorology using one or more graphs that show the most profound or consistent effects for each field deployment. Graphing and plotting tools are well established in many software packages (e.g., Excel, R, SigmaPlot, Matlab, Python) and testers can choose their preferred package to create plots. It is recommended that testers attach information about the software and/or the code used for this exploratory analysis to the base testing report as part of the data analysis and correction script information.

3.1.6.1 Potential Scatter Plots

Sensor measurements should be plotted on the y-axis (dependent variable) with the meteorological parameter measurements (as measured by the T and RH monitors, rather than on-board T and RH sensor measurements) on the x-axis (independent variable). Normalized concentration (in other words, the ratio of sensor to FRM/FEM concentration), concentration difference, absolute concentration difference, and DP calculations are discussed in the list below. It is recommended that testers choose plots from this list.

- 24-hour averaged normalized sensor $PM_{2.5}$ concentration vs. 24-hour averaged DP
- 24-hour averaged normalized sensor $PM_{2.5}$ concentration vs. 24-hour averaged RH
- 24-hour averaged normalized sensor $PM_{2.5}$ concentration vs. 24-hour averaged T
- 24-hour averaged concentration difference between the sensor and FRM/FEM $PM_{2.5}$ concentration vs. 24-hour averaged DP
- 24-hour averaged concentration difference between the sensor and FRM/FEM $PM_{2.5}$ concentration vs. 24-hour averaged RH
- 24-hour averaged concentration difference between the sensor and FRM/FEM $PM_{2.5}$ concentration vs. 24-hour averaged T
- 24-hour averaged absolute concentration difference between the sensor and FRM/FEM $PM_{2.5}$ concentration vs. 24-hour averaged DP
- 24-hour averaged absolute concentration difference between the sensor and FRM/FEM $PM_{2.5}$ concentration vs. 24-hour averaged RH
- 24-hour averaged absolute concentration difference between the sensor and FRM/FEM $PM_{2.5}$ concentration vs. 24-hour averaged T

3.1.6.2 Normalized Concentration

Normalized 24-hour averaged sensor $PM_{2.5}$ concentrations are derived by dividing the 24-hour averaged sensor $PM_{2.5}$ concentration by the paired 24-hour averaged FRM/FEM $PM_{2.5}$ concentration (Eq. 7). This equation assumes only one FRM/FEM instrument will be running. If multiple FRM/FEM instruments are running, separate testing reports can be generated for each.

$$NormC_{dj} = \frac{x_{dj}}{R_d} \qquad \text{Eq. 7}$$

where:

$NormC_{dj}$ = normalized 24-hour averaged sensor $PM_{2.5}$ concentration for day d and instrument j (unitless)

x_{dj} = valid 24-hour averaged sensor $PM_{2.5}$ concentration for day d and instrument j (µg/m^3)

R_d = valid 24-hour averaged FRM/FEM $PM_{2.5}$ concentration for day d (µg/m^3)

3.1.6.3 Concentration Difference and Absolute Concentration Difference

The 24-hour averaged concentration difference is derived by subtracting the 24-hour averaged FRM/FEM $PM_{2.5}$ concentration from the 24-hour averaged sensor $PM_{2.5}$ concentration (Eq. 8a).

$$\Delta C_{dj} = x_{dj} - R_d \qquad \text{Eq. 8a}$$

where:

ΔC_{dj} = concentration difference between valid 24-hour averaged sensor and FRM/FEM $PM_{2.5}$ concentration values for day d and sensor j (µg/m^3)

x_{dj} = valid 24-hour averaged sensor $PM_{2.5}$ concentration for day d and instrument j (µg/m^3)

R_d = valid 24-hour averaged FRM/FEM $PM_{2.5}$ concentration for day d (µg/m^3)

The 24-hour averaged absolute concentration difference for sensor $PM_{2.5}$ concentrations is derived by taking the absolute value of the difference between the 24-hour averaged sensor $PM_{2.5}$ concentration and the 24-hour averaged FRM/FEM $PM_{2.5}$ concentration (Eq. 8b). Equations 8a and 8b assume only one FRM/FEM instrument will be running. If multiple FRM/FEM instruments are running, separate testing reports can be generated for each.

$$Abs\Delta C_{dj} = \left| x_{dj} - R_d \right| \qquad \text{Eq. 8b}$$

where:

$Abs\Delta C_{dj}$ = absolute concentration difference between valid 24-hour averaged sensor and FRM/FEM $PM_{2.5}$ concentration values for day d and sensor j (µg/m^3)

x_{dj} = valid 24-hour averaged sensor $PM_{2.5}$ concentration for day d and instrument j (µg/m^3)

R_d = valid 24-hour averaged FRM/FEM $PM_{2.5}$ concentration for day d (µg/m^3)

3.1.6.4 Dew Point (DP)

The 24-hour averaged ambient DP is derived from the ambient T and RH measurements made by the independent T and RH monitors running alongside the sensors and FRM/FEM instrument (Eq. 9). DP should not be calculated using on-board T and RH sensor measurements (if applicable), as these measurements may not accurately represent ambient T and RH conditions.

$$\mathrm{DP}_d = 243.04 \times \left[\frac{\ln\left(\frac{\mathrm{RH}_d}{100}\right) + \frac{(17.625 \times \mathrm{T}_d)}{(243.04 + \mathrm{T}_d)}}{17.625 - \ln\left(\frac{\mathrm{RH}_d}{100}\right) - \frac{(17.625 \times \mathrm{T}_d)}{(243.04) + \mathrm{T}_d}} \right] \quad \text{Eq. 9}$$

where:

DP_d = valid 24-hour averaged ambient DP for day *d* (°C)

RH_d = valid 24-hour averaged ambient RH for day *d* (%)

T_d = valid 24-hour averaged ambient T for day *d* (°C)

3.2 Enhanced Testing Calculations

As a reminder, in order to properly compare the measurements (FEM, sensor, RH, T), it is important that the data streams are time aligned. This can be done by adjusting instrument times to a common standard clock (e.g., NIST time), carefully checking time stamps, and/or using a common data logger.

If data from any instrument is reported as an average, it is important to understand if the data average is ‘time ending’ or ‘time beginning’. For example, when logging hourly averages, the 07:00 time stamp may reflect data collected between 06:01-7:00 (time ending) or 7:00-7:59 (time beginning). This information should be considered when time aligning data.

3.2.1 Data Averages

Because it is difficult to maintain stable particle delivery for long periods of time, enhanced testing allows for the comparison of higher time resolution data (1-hour, 10-minute, 1-minute averages). The time interval to which all data should be averaged may be variable depending on the FEM, sensor, RH, and/or T instruments used and will be defined by the instrument with the lowest time resolution. For example, if the sensor, RH, and T are all recorded at a 1-minute time resolution, but the FEM is recorded at a 10-minute time resolution, all data should be averaged to the 10-minute time resolution. In Equation 10 (Eq. 10), this time interval is defined as *t*.

Consistent with base testing, a 75% data completeness requirement should be used for all time-averaged data collected in the enhanced testing procedure. For example, a $PM_{2.5}$ sensor recording concentration measurements every minute would require a minimum of 8 valid measurements in order to calculate a 10-minute averaged concentration (8/10 * 100% = 80% data completeness, which is greater than 75%).

$$x_{ktj} = \frac{1}{n}\sum_{i=1}^{n} c_{ij} \qquad \text{Eq. 10}$$

where:

x_{ktj} = averaged measurement *k* for time interval *t* and instrument *j* (μg/m^3, °C, % RH)

n = number of instrument measurements during time interval *t*

c_{ij} = measurement from instrument *j* for time *i* of time interval *t* (μg/m^3, °C, % RH)

As a reminder, x_{ktj} is considered valid if 75% of the time interval is represented by the c_{ij} measurements.

3.2.2 Test Averages

Test averaged measurements should be calculated from valid averaged data (Eq. 11) collected during the steady state period for each test.

$$\overline{x_k} = \frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{N}\sum_{h=1}^{N} \mathrm{x}_{ktj}\right] \qquad \text{Eq. 11}$$

where:

$\overline{x_k}$ = test averaged measurement *k* for the chamber test (μg/m^3, °C, % RH)

M = number of identical instruments operated simultaneously during the chamber test

N = number of valid time intervals during which all identical instruments are operating and returning valid averages over the duration of the chamber test

x_{ktj} = valid averaged measurement for time interval *t* and instrument *j* (μg/m^3, °C, % RH)

3.2.3 Precision

Precision between identical sensors can be characterized by two metrics: standard deviation (SD) between measurements (Eq. 12) and coefficient of variation (CV; Eq. 13). This metric should be calculated from valid averaged data collected during the mid concentration test condition during the post-aging (Day 60) drift test (Section 2.2.6).

$$\mathrm{SD} = \sqrt{\frac{1}{(N \times M) - 1}\sum_{j=1}^{M}\left[\sum_{t=1}^{N}(x_{tj} - \overline{x_t})^2\right]} \qquad \text{Eq. 12}$$

where:

SD = standard deviation of test averaged sensor $PM_{2.5}$ concentration measurements ($\mu g/m^3$)

N = number of valid time intervals during which all identical instruments are operating and returning valid averages over the duration of the chamber test

M = number of identical sensors operated simultaneously during the chamber test

x_{tj} = averaged sensor $PM_{2.5}$ concentration for time interval t and sensor j ($\mu g/m^3$)

$\overline{x_t}$ = test averaged sensor $PM_{2.5}$ concentration for time interval t ($\mu g/m^3$)

$$CV_{Enhanced} = \frac{\text{SD}}{\overline{x}} \times 100 \qquad \text{Eq. 13}$$

where:

$CV_{Enhanced}$ = coefficient of variation (%)

SD = standard deviation of test averaged sensor $PM_{2.5}$ concentration measurements ($\mu g/m^3$)

$\overline{x}$ = test averaged sensor $PM_{2.5}$ concentration for the chamber test ($\mu g/m^3$)

3.2.4 Bias and Linearity

A simple linear regression model can demonstrate the relationship between paired averaged sensor and FEM $PM_{2.5}$ measurements. During enhanced testing, pooling the data collected during the steady state period of the low and mid concentration test conditions during the post-aging (Day 60) drift test (Section 2.2.6) and the high and higher concentration tests (Section 2.2.7) will reflect data collected under similar T and RH conditions. Using a simple linear regression model (y = mx + b) with the sensor $PM_{2.5}$ measurements as the dependent variable (y) and the FEM $PM_{2.5}$ measurements as the independent variable (x), calculate the slope (m), intercept (b), and the coefficient of determination (R^2) for each test.

A function for determining a simple linear regression model is well established in many software packages (e.g., Excel, R) and readily available using the U.S. EPA Excel-based Macro Analysis Tool (https://www.epa.gov/air-sensor-toolbox/air-sensor-collocation-macro-analysis-tool, *last accessed 07/25/2020*), thus the equations are not presented here. Caution should be taken to appropriately select the FEM measurements as the independent (x) variable and sensor measurements as the dependent (y) variable when using these tools.

3.2.5 Error

The root mean square error (RMSE) is one metric that can be used to help understand the error associated with sensor $PM_{2.5}$ concentration measurements. The interpretation of this value is slightly more straightforward because it is calculated in concentration units. This metric should be calculated from valid averaged data collected during the mid concentration test condition during the post-aging (Day 60) drift test (Section 2.2.6). Using data during which all sensors are reporting valid time averaged measurements, the sensor and FEM $PM_{2.5}$ measurement calculations are compared (Eq. 14). This equation assumes only one FEM instrument will be running. If multiple FEM instruments are running, separate testing reports can be generated for each.

$$\text{RMSE} = \sqrt{\frac{1}{N \times M}\sum_{j=1}^{M}\left[\sum_{t=1}^{N}(x_{tj} - R_t)^2\right]} \qquad \text{Eq. 14}$$

where:

RMSE = root mean square error (µg/m^3)

N = n umber of valid time intervals during which all identical instruments are operating and returning valid averages over the duration of the chamber test

M = number of identical sensors operated simultaneously during the chamber test

x_{tj} = averaged sensor $PM_{2.5}$ concentration for time interval t and instrument j (µg/m^3)

R_t = averaged FEM $PM_{2.5}$ concentration for time t (µg/m^3)

As a caution, RMSE is not defined in a consistent way throughout available resources. It has commonly been defined in two ways: 1) describing the difference between a measurement and the true value, and 2) describing the difference between a measurement and a linear regression best fit line of a measurement and a corresponding true value. In this report, RMSE is defined as the error between the sensor measurements and the reference instrument measurements or true values (see Eq. 14). This approach is presumed to provide the best indication of out-of-the-box sensor performance and the error that can be expected prior to any data corrections. Further, this approach is how RMSE is commonly calculated in air sensor literature to date.

The normalized root mean square error (NRMSE) is also included as a metric to account for testing conditions where the ambient concentrations may be much higher than typical U.S. ambient levels (e.g., wildfires). This metric should be calculated from valid averaged data collected during the high and higher concentration tests (Section 2.2.7). The RMSE calculated value (Eq. 14) is normalized using the average of the valid time averaged FEM $PM_{2.5}$ concentrations over the testing period (Eq. 15).

$$\text{NRMSE} = \frac{\text{RMSE}}{\overline{R_t}} \times 100 \qquad \text{Eq. 15}$$

where:

NRMSE = normalized root mean square error (%)

RMSE = root mean square error as calculated in Eq. 14 (μg/m^3)

$\overline{R_t}$ = valid test averaged FEM $PM_{2.5}$ concentration over the test period (μg/m^3)

3.2.6 Effect of Relative Humidity (RH)

As described in Section 2.2.4, the RH tests on sensor measurements involve two steps: 1) collecting data during steady state at a prescribed $PM_{2.5}$ concentration at 40% RH, and 2) collecting data during steady state at the same prescribed $PM_{2.5}$ concentration at 85% RH. The effect of RH is the difference between these two measurements (Eq. 16).

$$\overline{x_{RH}} = \overline{x_{(RH=85\%)}} - \overline{x_{(RH=40\%)}} \qquad \text{Eq. 16}$$

where:

$\overline{x_{RH}}$ = test averaged influence of RH on sensor measurements (μg/m^3)

$\overline{x_{(RH=85\%)}}$ = test averaged sensor $PM_{2.5}$ concentration for the portion of the chamber test when the RH is 85% (μg/m^3)

$\overline{x_{(RH=40\%)}}$ = test averaged sensor $PM_{2.5}$ concentration for the portion of the chamber test when the RH is 40% (μg/m^3)

3.2.7 Effect of Temperature (T)

As described in Section 2.2.5, the T tests on sensor measurements involve two steps: 1) collecting data during steady state at a prescribed $PM_{2.5}$ concentration at 20°C, and 2) collecting data during steady state at the same prescribed $PM_{2.5}$ concentration at 40°C. The effect of T is the difference between these two measurements (Eq. 17).

$$\overline{x_T} = \overline{x_{(T=40)}} - \overline{x_{(T=20)}} \qquad \text{Eq. 17}$$

where:

$\overline{x_T}$ = test averaged influence of T on sensor measurements (μg/m^3)

$\overline{x_{(T=40)}}$ = test averaged sensor $PM_{2.5}$ concentration for the portion of the chamber test when the T is 40°C (μg/m^3)

$\overline{x_{(T=20)}}$ = test averaged sensor $PM_{2.5}$ concentration for the portion of the chamber test when the T is 20°C (µg/m^3)

3.2.8 Drift

As described in Section 2.2.6, the drift tests involve measuring the drift at two $PM_{2.5}$ concentrations: 1) at a low concentration of 10 µg/m^3, and 2) at a mid concentration of 35 µg/m^3 which is relevant for health messaging. For each $PM_{2.5}$ concentration, the drift measurement includes two separate chamber tests. The first will be conducted to determine the steady state concentration for the prescribed $PM_{2.5}$ concentration. The sensors will then be operated continuously and tested again at least 60 days later to see if the measurement has drifted. The amount of drift will be quantified for both $PM_{2.5}$ concentrations by the difference in the measurement over the 60-day period (Eq. 18).

$$\overline{x_{C_{drift}}} = \overline{x_{C(day=60)}} - \overline{x_{C(day=1)}} \qquad \text{Eq. 18}$$

where:

$\overline{x_{C_{drift}}}$ = test averaged sensor drift at $PM_{2.5}$ concentration *C* over the course of 60 days (µg/m^3)

$\overline{x_{C(day=60)}}$ = test averaged sensor $PM_{2.5}$ concentration at $PM_{2.5}$ concentration *C* after 60 days of operation following the start of the drift test (µg/m^3)

$\overline{x_{C(day=1)}}$ = test averaged sensor $PM_{2.5}$ concentration at $PM_{2.5}$ concentration *C* at the beginning of the drift test (µg/m^3)

3.2.9 Accuracy at High Concentrations

As described in Section 2.2.7, the accuracy at high concentrations test involves testing the sensor response at a high $PM_{2.5}$ concentration which is relevant for health messaging and a higher $PM_{2.5}$ concentration which is relevant for $PM_{2.5}$ events such as wildfires. The accuracy of the sensor measurement will be determined by the difference between the sensor and FEM measurements (Eq. 19).

$$\overline{x_{\Delta}} = \overline{x_{sensor}} - \overline{x_{ref}} \qquad \text{Eq. 19}$$

where:

$\overline{x_{\Delta}}$ = test averaged difference between the sensor and FEM $PM_{2.5}$ concentrations (µg/m^3)

$\overline{x_{sensor}}$ = test averaged sensor $PM_{2.5}$ concentration (µg/m^3)

$\overline{x_{ref}}$ = test averaged FEM $PM_{2.5}$ concentration (µg/m^3)

4.0 Target Values for $PM_{2.5}$ Air Sensors

4.1 Approach

To inform the development of the target values for the performance metrics for $PM_{2.5}$ air sensors (outlined in Section 3.0), the U.S. EPA considered the same resources used to inform the selection of performance metrics (i.e., workshop discussions, FRM/FEM performance specifications, U.S. EPA sensor evaluation results, AQ-SPEC sensor field evaluations, peer-review literature findings, and target levels proposed by organizations developing sensor standards/certification programs).

The sensor performance evaluation results gathered from the available resources are summarized in Table 4-1 (more detail available in Appendix D). In summarizing the performance results, the U.S. EPA did not consider results deemed to be outliers or unrepresentative of normal sensor operation to avoid significantly biasing the recommended target values. These results reflect out-of-box sensor performance before additional corrections were made by the user.

Table 4-1. $PM_{2.5}$ Sensor Performance Field Evaluation Results from Available Resources

Performance Metric		Range	Average	Median
Precision	CV (%)	0.89 to 31.03	12.78	11.62
Bias	Slope*	0.50 to 1.49	1.09	1.12
	Intercept* (μg/m^3)	-19.08 to 0.91	-3.75	-3.19
Linearity	$R^{2,\dagger}$	0.52 to 0.97	0.80	0.83
Error	RMSE (μg/m^3)	2.41 to 7.64	5.28	5.52

Note: Resources include AQ-SPEC sensor evaluations, the U.S. EPA sensor evaluations, and peer-reviewed literature. Table only includes 24-hour averaged data.
*Slopes outside of 0.5 to 1.5 were not considered; the intercept was not considered if the slope was discarded.
†R^2 values greater than or equal to 0.5 were considered; R^2 values less than 0.5 were not considered.

Performance metrics and target values related to air sensor standards/certification programs from the MEE (Environmental Protection Department of Hebei Province, 2017) are summarized in Appendix C and D. Additionally, the performance specifications for FRM/FEM monitors used for regulatory compliance are discussed in Appendix C and D.

4.2 List of Target Values

Table 4-2 summarizes the performance metrics and target values recommended for the base and enhanced testing protocols for $PM_{2.5}$ air sensors used in ambient, outdoor, fixed site NSIM applications. The recommended performance metrics and target values for base and enhanced testing reflect the current state-of-the-science as the range of observed performance (Table 4-1) demonstrates that $PM_{2.5}$ sensors should be possible to achieve. Encouraging development of sensors which meet these target values should help ensure that sensor performance can be well characterized and understood. Additional performance metrics and test conditions for the enhanced testing protocol are shown in Table 4-3. Target values for enhanced testing are not included at this time due to limited feasibility, lack of consistency regarding testing protocols, and inconsistency in sensor evaluation results that can result due to the limited amount of data that will be collected and the variation in the tester's choice of PM surrogate (see Appendix D for more detailed discussion).

Table 4-2. Base and Enhanced Testing – Recommended Performance Metrics and Target Values for $PM_{2.5}$ Air Sensors Used in Ambient, Outdoor, Fixed Site NSIM Applications

Performance Metric		**Target Value**		**Associated Section Describing Calculation**
		Base Testing	**Enhanced Testing***	
Precision	Standard Deviation (SD)	≤ 5 μg/m³	No target values recommended; report results	3.1.3 and 3.2.3
	-OR- Coefficient of Variation (CV)	≤ 30%		3.1.3 and 3.2.3
Bias	Slope	1.0 ± 0.35		3.1.4 and 3.2.4
	Intercept (b)	-5 ≤ b ≤ 5 μg/m³		3.1.4 and 3.2.4
Linearity	Coefficient of Determination (R^2)	≥ 0.70		3.1.4 and 3.2.4
Error	Root Mean Square Error (RMSE) or Normalized Root Mean Square Error (NRMSE)	RMSE ≤ 7 μg/m³ or NRMSE ≤ 30%†		3.1.5 and 3.2.5

*No specific target values are recommended due to limited feasibility, lack of consensus regarding testing protocols, and inconsistency in sensor evaluation results that can result due to the limited amount of data that will be collected and variation in the tester's choice of PM surrogate. See Appendix D for further discussion.

†A sensor will meet this target if either the RMSE or NRMSE meet this criterion. See Appendix D for further discussion.

Table 4-3. Enhanced Testing – Additional Recommended Performance Metrics and Test Conditions for $PM_{2.5}$ Air Sensors Used in Ambient, Outdoor, Fixed Site NSIM Applications

Performance Metric	Test Conditions	Associated Section Describing Calculation
Effect of Relative Humidity (RH)	Moderate RH: 40% ± 5%	3.2.6
	Elevated RH: 85% ± 5%	3.2.6
Effect of Temperature (T)	Moderate T: 20°C ± 1°C	3.2.7
	Elevated T: 40°C ± 1°C	3.2.7
Drift	Low concentration: 10 $\mu g/m^3$ ± 10%	3.2.8
	Mid concentration: 35 $\mu g/m^3$ ± 5%	3.2.8
Accuracy at High Concentrations	High concentration: 150 $\mu g/m^3$ ± 5%	3.2.9
	Higher concentration: 250 $\mu g/m^3$ ± 5%	3.2.9

It is recognized that the information in this report is based on the current knowledge of $PM_{2.5}$ air sensors at the time this report was released and that $PM_{2.5}$ sensor technologies will likely continue to develop and improve over time. The U.S. EPA anticipates updating Tables 4-2 and 4-3 as well as other information in this report, as feasible, to reflect advances in $PM_{2.5}$ sensor technologies and knowledge gained from sensor evaluation results. Updates will likely be shared as an addendum to this report.

Testing results do not constitute certification or endorsement by the U.S. EPA. It is recommended that testers make the testing reports available on their respective websites to inform consumers on the testing results.

5.0 References

1. Badura, M., Batog, P., Drzeniecka-Osiadacz, A., Modzel, P. Evaluation of low-cost sensors for ambient $PM_{2.5}$ monitoring. Journal of Sensors, 1-16, https://doi.org/10.1155/2018/5096540, 2018.
2. Bulot, F.M.J., Johnston, S.J., Bashford, P.J., Easton, N.H.C., Apetroaie-Cristea, M., Foster, G.L., Morris, A.K.R., Cox, S.J., Loxham, M. Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. Scientific Reports, 9, 7497, 2019.
3. Crilley, L.R., Shaw, M., Pound, R., Kramer, L.J., Price, R., Young, S., Lewis, A.C., Pope, F.D. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. Atmospheric Measurement Techniques, 11, 709-720, https://doi.org/10.5194/amt-11-709-2018, 2018.
4. Environmental Protection Department of Hebei Province. Technical Regulation for Selecting the Location of Air Pollution Control Gridded Monitoring System, DB13/T 2545-2017, 2017.
5. Feenstra, B., Papapostolou, V., Hasheminassab, S., Zhang, H., Der Boghossian, B., Cocker, D., Polidori, A. Performance evaluation of twelve low-cost $PM_{2.5}$ sensors at an ambient air monitoring site. Atmospheric Environment, 216, 116946, https://doi.org/10.1016/j.atmosenv.2019.116946, 2019.
6. Frederick, S., Johnson, K., Johnson, C., Yaga, R., Clements, A. Performance Evaluations of $PM_{2.5}$ Sensors in Research Triangle Park, NC: PurpleAir PA-II-SD, Aeroqual AQY, Applied Particle Technology Maxima, Vaisala AQT420, Sens-it RAMP, and Clarity Node-S. Presented at EPA Air Sensor Brownbag (Research Triangle Park, NC), 2020a. Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=CEMM&dirEntryId=348487
7. Frederick, S., Johnson, K., Johnson, C., Yaga, R., Clements, A. Performance Evaluations of $PM_{2.5}$ Sensors in Research Triangle Park, NC. Presented at the Air Sensors International Conference (Pasadena, CA; May 12-15, 2020), 2020b. Available at: https://cfpub.epa.gov/si/si_public_record_Report.cfm?dirEntryId=349512&Lab=CEMM.
8. Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., Morawska, L. The influence of humidity on the performance of low-cost air particle mass sensor and the effect of atmospheric fog. Atmospheric Measurement Techniques, 11, 4883-4890, https://doi.org/10.5194/amt-11-4883-2018, 2018.
9. Kelly, K.E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., Butterfield, A. Ambient and laboratory evaluation of a low-cost particulate matter sensor. Environmental Pollution, 221, 491-500, http://dx.doi.org/10.1016/j.envpol.2016.12.039, 2017.
10. Li, J., Mattewal, S.K., Patel, S., Biswas, P. Evaluation of nine low-cost-sensor based particulate matter monitors. Aerosol and Air Quality Research, 20, 254-270, https://doi.org/10.4209/aaqr.2018.12.0485, 2020.
11. Mukherjee, A., Stanton, L.G., Graham, A.R., Roberts, P.T. Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama Valley of California. Sensors, 17, 1805, http://dx.doi.org/10.3390/s17081805, 2017.
12. Nakayama, T., Matsumi, Y., Kawahito, K., Watabe, Y. Development and evaluation of a palm-sized optical $PM_{2.5}$ sensor. Aerosol Science and Technology, 52, 2-12, https://doi.org/10.1080/02786826.2017.1375078, 2018.
13. Papapostolou, V., Zhang, H., Feenstra, B.J., Polidori, A. Development of an environmental chamber for evaluating the performance of low-cost air quality sensors under controlled conditions. Atmospheric Environment, 171, 82-90, https://doi.org/10.1016/j.atmosenv.2017.10.003, 2017.
14. Schneider, P., Bartonova, A., Castell, N., Dauge, F.R., Gerboles, M., Hagler, G.S.W., Hügline, C., Jones, R.L., Khan, S., Lewis, A.C., et al. Toward a unified terminology of processing levels

for low-cost air-quality sensors. Environmental Science & Technology, 53, 8485-8487, http://dx.doi.org/10.1021/acs.est.9b03950, 2019.
15. South Coast Air Quality Management District (SCAQMD), Air Quality Sensor Performance Evaluation Center (AQ-SPEC). PM sensor evaluations, available at: http://www.aqmd.gov/aq-spec/evaluations/summary-pm
16. South Coast Air Quality Management District (SCAQMD), Air Quality Sensor Performance Evaluation Center (AQ-SPEC), Laboratory Evaluation of Low-Cost Air Quality Sensors: Laboratory Setup and Testing Protocol, 2016. Available at: http://www.aqmd.gov/docs/default-source/aq-spec/protocols/sensors-lab-testing-protocol6087afefc2b66f27bf6fff00004a91a9.pdf
17. U.S. EPA. Air Sensor Guidebook. U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA 600/R-14/159, 2014.
18. U.S. EPA. Evaluation of Elm and Speck Sensors. U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-15/314, 2015. Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&dirEntryId=310285
19. U.S. EPA. Quality Assurance Guidance Document 2.12: Monitoring $PM_{2.5}$ in Ambient Air Using Designated Reference or Class I Equivalent Methods, Research Triangle Park, NC, EPA-454/B-16-001, 2016. Available at: https://www3.epa.gov/ttn/amtic/files/ambient/pm25/qa/m212.pdf
20. U.S. EPA. Quality Assurance Handbook for Air Pollution Measurement Systems, Volume IV: Meteorological Measurements Version 2.0 (Final). U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA-454/B-08-002, 2008.
21. U.S. EPA. Peer Review and Supporting Literature Review of Air Sensor Technology Performance Targets. U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA/600/R-18/324, 2018.
22. U.S. EPA. Peer Review and Supporting Literature Review of Air Sensor Technology Performance Targets: 2019 Supplemental. U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA/600/R-20/120, 2020.
23. U.S. EPA, 40 CFR Part 50 – National Primary and Secondary Ambient Air Quality Standards.
24. U.S. EPA, 40 CFR Part 53 – Ambient Air Monitoring Reference and Equivalent Methods.
25. U.S. EPA 40 CFR Part 58 – Ambient Air Quality Surveillance.
26. Williams, R., Duvall, R., Kilaru, V., Hagler, G., Hassinger, L., Benedict, K., Rice, J., Kaufman, A., Judge, R., Pierce, G., et al. Deliberating performance targets workshop: Potential paths for emerging $PM_{2.5}$ and O_3 air sensor progress. Atmospheric Environment: X, 2, 100031, doi:10.1016/j.aeaoa.2019.100031, 2019.
27. Zamora, M.L., Rice, J., Koehler, K. One year evaluation of three low-cost $PM_{2.5}$ monitors. Atmospheric Environment, 235, 117615, 2020.
28. Zheng, T., Bergin, M., Johnson, K., Tripathi, S., Shirodkar, S., Landis, M., Sutaria, R., and Carlson, D. Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. Atmospheric Measurement Techniques, 11, 4823–4846, https://doi.org/10.5194/amt-11-4823-2018, 2018.
29. Zikova, N., Hopke, P.K., Ferro, A.R. Evaluation of new low-cost particle monitors for $PM_{2.5}$ concentrations measurements. Journal of Aerosol Science, 105, 24-34, https://doi.org/10.1016/j.jaerosci.2016.11.010, 2017.

Appendix A: Definitions

This Appendix summarizes the definitions for the commonly used terms included throughout this report. In developing these definitions, we consulted a variety of resources (e.g., AQ-SPEC, People's Republic of China MEE, 40 CFR Part 53, peer-reviewed literature) to try to provide consistency in the use of these terms among documents and an appropriate level of detail to support testers and consumers.

Accuracy: A measure of the agreement between the pollutant concentrations reported by the sensor and the reference instrument. This includes a combination of random error (precision) and systematic error (bias) components which are due to sampling and analytical operations. One way to measure this agreement is by calculating the root mean square error (RMSE; calculation described in Section 3.1.5)

Air Sensor: A class of non-regulatory technology that are lower in cost, portable, and generally easier to operate than regulatory monitors. Air sensors often provide relatively quick or instant air pollutant concentrations (both gas-based and particulate matter) and allow air quality to be measured in more locations. The term 'air sensor' often describes an integrated set of hardware and software that uses one or more sensing elements (also sometimes called sensors) to detect or measure pollutant concentrations.

Bias: The systematic (non-random) or persistent disagreement between the concentrations reported by the sensor and reference instruments. It is often determined using the linear regression slope and intercept of a simple linear regression, fitting sensor measurements (y-axis) to reference measurements (x-axis).

Coefficient of Variation (CV): The ratio of the standard deviation (SD) to the mean among a group of collocated sensors of the same type, used to show the precision between sensors.

Collocation: The process by which a sensor and a reference instrument are operated at the same time and place under real world conditions. The siting criteria (e.g., proximity and height of the sensor and the reference monitor) should follow procedures outlined in 40 CFR Part 58 as closely as possible. For example, sensors should be placed within 20 meters horizontal of the reference instrument, positioned such that the sample air inlets for the sensors are within a height of ± 1 meter vertically of the sample air inlets of the reference instrument, and placed as far as possible from any obstructions (e.g., trees, walls) to minimize spatial and wind turbulence effects on sample collection.

Comparability: The level of overall agreement between two separate data sets. This term is often used to describe how well sensor data compares with reference instrument data. Comparability is a combination of accuracy, precision, linearity, and other performance metrics.

Completeness: In determining averages, completeness describes the amount of valid data obtained relative to the averaging period. In this report, a completeness threshold is prescribed to make sure that the average is representative of the concentrations observed within the averaging period. For example, if a sensor collects measurements every 5 minutes, it can return 12 measurements every hour. To obtain 75% data completeness for a calculated hourly average, at least 9 valid measurements are needed (i.e., 9/12 * 100% = 75%).

Concurrent: Operating a series of instruments at the same time and place. Concurrent measurements cover the same period of time and are time aligned so that they can be compared.

Drift: A change in the response or concentration reported by a sensor when challenged by the same pollutant concentration over a period during which the sensor is operated continuously and without adjustment.

Dew Point (DP): The temperature (T) to which air must be cooled to become saturated with water vapor.

Error: A measure of the disagreement between the pollutant concentrations reported by the sensor and the reference instrument. One way to measure error is by calculating the root mean square error (RMSE; calculation described in Sections 3.1.5 and 3.2.5).

Effect of Dew Point (DP), Relative Humidity (RH), or Ambient Temperature (T): A positive or negative measurement response caused by variations in DP, RH, or ambient T and not by changes in the concentration of the target pollutant.

Federal Equivalent Method (FEM): A method for measuring the concentration of an air pollutant in the ambient air that has been designated as an equivalent method in accordance with 40 CFR Part 53. An FEM does not include a method for which an equivalent method designation has been canceled in accordance with 40 CFR Parts 53.11 or 53.16. A list of designated FEMs can be found here: https://www.epa.gov/amtic/air-monitoring-methods-criteria-pollutants, *last accessed 07/26/2020*.

Federal Reference Method (FRM): A method of sampling and analyzing the ambient air for an air pollutant that is specified as a reference method in 40 CFR Part 50, or a method that has been designated as a reference method in accordance with 40 CFR Part 53. An FRM does not include a method for which the U.S. EPA has cancelled a reference method designation in accordance with 40 CFR Parts 53.11 or 53.16. A list of designated FRMs can be found here: https://www.epa.gov/amtic/air-monitoring-methods-criteria-pollutants, *last accessed 07/06/2020*.

Linearity: A measure of the extent to which the measurements reported by a sensor can explain the concentrations reported by the reference instrument. It is often quantified by the coefficient of determination (R^2) obtained from the simple linear regression fitting sensor measurements (y-axis) to reference instrument measurements (x-axis) with values closer to 1 generally indicating better linearity. In some cases, sensor measurements can be linear with a near perfect R^2 but may differ significantly from the reference instrument measurements. For example, a linear regression can result in an R^2 of 0.99 and slope of 5. This indicates that the reported sensor measurement is always 5 times higher than the reference instrument measurements.

Performance Metric: A parameter used to describe the data quality of a measurement device.

Precision: Variation around the mean of a set of measurements obtained concurrently by two (2) or more sensors of the same type collocated under the same sampling conditions. The consistency in measurements from identical sensors is often quantified by standard deviation (SD) or the coefficient of variation (CV; calculation described in Sections 3.1.3 and 3.2.3) with lower values indicating a more precise measurement.

Relative Humidity (RH): The measure of the amount of moisture or water vapor in the air as a function of temperature (T).

Representativeness: A description of how closely a sample reflects the characteristics of the whole. Although challenging to verify, effort should be made to ensure that a sample is representative using techniques such as thorough mixing to obtain homogeneity, duplicate analyses, etc. For example, the data completeness threshold suggested in this report is meant to ensure that measurements averaged to longer time intervals are as representative as possible by covering at least 75% of the time period.

Root Mean Square Error (RMSE): A measure of the random disagreement between the measurements reported by the sensor and the reference measurements. RMSE is one of several ways to measure error. It penalizes large deviations from the reference measurements and is therefore, sensitive to outliers. It should be noted that in this report, RMSE is not quantified by the linear regression best fit line of the sensor measurements and corresponding reference instrument measurements. See Section 3.1.5 which describes the RMSE definition and corresponding calculation for base testing and Section 3.2.5 which describes the calculation for enhanced testing.

Normalized Root Mean Square Error (NRMSE): A measure of the overall difference (error) between the measurements made by the sensor and the reference instrument measurement but accounts for average concentration resulting in a percentage that can be more helpful if tests occur over a range of average concentrations.

Standard Deviation (SD): A measure of the amount of variation in measurements from sensors of the same type reported in the same units as the concentration measurement.

Uptime: A measure of the amount of valid data obtained by all tested sensors relative to the amount of data that was expected to be obtained under correct, normal operation for the entire length of a test. For example, if valid data is collected by all three sensors for 29 days of a 30-day base test field deployment the uptime for the deployment can be expressed as 96.7% (i.e., 29 days/30 days * 100%). Operation may be interrupted by sensor failure, connectivity issues, equipment maintenance, extreme weather events, etc. No matter the reason for missing data, all downtime should be included in the uptime calculation. However, tests may report more information such as specifying the percent of downtime attributed to various types of interruptions.

Appendix B: Supporting Information for Testing Protocols

Testing protocols for $PM_{2.5}$ air sensors were drafted based on best known practices in the literature to date with the goal of collecting an array of comparable data on air sensors without overstraining resources. The methodology considered the air sensor testing protocols performed by AQ-SPEC (Papapostolou et al., 2017; SCAQMD, 2016), the U.S. EPA, and the People's Republic of China MEE, as well as protocols used for FRM/FEM regulatory monitors (40 CFR Part 53) that test the capabilities and constraints of measurement devices.

Base testing protocols were modeled after field evaluations conducted by a variety of organizations with slight variations. Sensor studies most commonly compare sensor measurements at 24-hour or 1-hour averages. This protocol emphasizes 24-hour averaged data to compare air sensor and FRM/FEM monitor information as there is little variation between different types of FRM/FEM monitors at this scale per federal requirements. Thus, with 24-hour averaged data there is more confidence that the type of FRM/FEM monitor will not influence results. Testers are encouraged to include 1-hour average information if it can be reasonably assumed that this averaging interval will be used in practice by consumers. As previously noted, there may be unquantified variation between FEM instruments at the 1-hour time averaging interval versus the 24-hour averaging interval. The base testing protocol uses a single FRM/FEM monitor as the recommended minimum for comparison, to reduce equipment needed for testing and to simplify calculations. In concurrence with current sensor evaluation efforts and 40 CFR Part 53.35, these testing protocols recommend testing three (3) or more sensors simultaneously for at least 30 days with a 75% data completeness threshold. Testing three (3) or more identical air sensors can help consumers understand the variation in performance that may exist among identical sensors. Discussion with experts and review of current practices determined that two (2) field deployments are likely sufficient to show sensor performance over a range of conditions including T, RH, weather, $PM_{2.5}$ concentrations, and other factors that provide information about the sensor's potential performance in a variety of other areas of the U.S. The deployment site criteria (Table 2-1) are meant to be achievable at a variety of locations across the U.S. For NSIM applications where high $PM_{2.5}$ concentrations are expected (e.g., wildfire smoke applications), it is recommended that testers conduct base testing in more than two (2) locations and include sites impacted by wildfire smoke and higher concentrations.

Identification of field sites for base testing involved evaluating the past three years (2017-2019) of $PM_{2.5}$ data, obtained from AQS, across the U.S. To determine a reasonable threshold concentration for base testing, a concentration was selected that could likely be achieved at a single site in half of the U.S. states. $PM_{2.5}$ data from 2017-2019 was separated by month and sites with three points above the target threshold concentration in the past three years were identified (e.g., three points total above 25 $\mu g/m^3$ in July 2017, 2018, and 2019). Target concentrations ranging from 25-55 $\mu g/m^3$ were initially considered. Analysis of the data showed that a target of 35 $\mu g/m^3$ or lower was achievable in a single site in half of the U.S. states. A goal concentration of 25 $\mu g/m^3$ for one day during the 30-day test period was chosen as it could most likely be achieved and would provide the most flexibility in site selection. Note that some historically high $PM_{2.5}$ concentrations may be caused by wildfires and other extreme events not present every year while others are caused by specific seasonal events, typical meteorology, or specific holiday events (e.g., 4th of July, New Year's Eve fireworks). Table B-1 summarizes locations and months that are likely to meet the one day, 24-hour goal $PM_{2.5}$ concentration as specified in the test site selection criteria (Table 2-1).

Table B-1. General locations likely to meet 25 µg/m³ criteria by month based on analysis of 2017-2019 $PM_{2.5}$ data. Locations typically referred to in terms of climate zones shown in Figure 2-2.

Month	Locations Likely to Meet Test Site Criteria
January and February	Many sites in near the border of the Southwest and Northwest regions, many sites across West Coast states, along with some sites in the Northeast, Ohio Valley and Upper Midwest, and a few sites in the Southeast and South.
March	Some sites in most climate regions.
April, May, and June	Concentrations are typically lowest across the U.S. during these months so there are few sites that will likely exceed 25 µg/m³ although some near the southern border may.
July	Many sites in the center of the country (Upper Midwest, Ohio Vally, and South) along with sites in west coast states.
August	Many sites in the western half of the country. Almost 20% of sites in the U.S. will likely achieve concentrations above 25 µg/m³.
Sept	Sites in the Northwest, Northern Rockies and Plains, and West.
October	Some sites in West Coast states.
November	Sites in West Coast States and some scattered across every other region of the U.S. Almost 15% of sites across the country are likely to achieve concentrations above 25 µg/m³.
December	Roughly 25% of the sites across the U.S. are likely to achieve concentrations above 25 µg/m³. Many sites in the West, Northwest, Southwest, Upper Midwest, Ohio Valley, and Northeast.

Enhanced testing protocols were modeled after laboratory evaluations conducted by a variety of organizations seeking to quantify the effect of RH and T, drift, and accuracy at high concentrations. Other tests, specifically the detection limit (DL), were considered but ultimately not included at this time due to limited feasibility and inconsistency in results. Testing protocols outlined in this document specify initial conditions of 20°C and 40% RH to maintain consistency with other laboratory sensor performance evaluations. The $PM_{2.5}$ concentration levels used in testing align with Air Quality Index (AQI) breakpoints for 24-hour averaged $PM_{2.5}$ concentrations.

Testing duration and data time averaging during the tests can vary dependent on the equipment being used for testing. The enhanced testing protocol describes a test duration as the time needed to collect either a minimum of 20-30 pairs of time-matched sensor and FEM data points or three (3) consecutive hours of steady state data. This language reflects the need to maintain a level of flexibility to collect a sufficient amount of data to produce statistically significant results, handle a wide variety of sensors presently on the market, accommodate the time resolution available on current FEM instruments, and prudently minimize the cost and effort involved in maintaining steady state conditions within a test chamber for extended periods of time. Many sensors on the market today provide measurements at high time resolutions (between 1-minute and 5-minute averages). Current FEMs that may be used for this work often report at 1-minute, 10-minute rolling, or 1-hour averages. A pair of high time resolution instruments (sensor and FEM both reporting 1-minute averages) could collect 20 or more pairs of time-matched data quickly thereby minimizing the cost and duration of the test. A chamber using an FEM that only reports hourly averaged data would require a day to collect 20 time-matched data pairs but maintaining steady state conditions for that long would be extremely difficult, if not impossible. However, 3 time-matched data pairs (3 hours of testing) would provide a minimum number of data points for a statistical analysis. Testers should collect as many time-matched data pairs as possible, within the constraints of the testing setup, with a suggestion that 20-30 time-matched data pairs would be an ideal dataset.

Effect of Relative Humidity (RH) testing protocols use two RH condition set points (40% and 85%) to simplify testing as higher and lower setpoints may be difficult to achieve in some laboratory exposure chambers. Further, an elevated RH condition (85% RH) is important to better characterize performance in areas like the Southeast U.S that can experience high RH levels.

Effect of Temperature (T) testing protocols compare 20°C condition to the elevated T condition of 40°C. The elevated T condition of 40°C is important to better characterize sensor performance in areas like the Southwest U.S that can experiences high T levels.

Detection Limit (DL) testing protocols were not included in this report at this time. Several methodologies were considered but none seemed to provide consistent results across a variety of sensor devices. Observations from recent evaluation efforts suggest that some PM sensors do not report negative concentrations and some report zero concentration for a range of low pollutant concentrations. Additionally, a sensor's response to low concentrations may have a different slope and/or variable uncertainty across the low concentration regime (0-8 $\mu g/m^3$). This makes determination and interpretation of a DL difficult. Understanding the lowest concentration a device can measure is useful in knowing when the NSIM measurement needs cannot be met by a given device. Testers are encouraged to provide the manufacturer reported DL (typically found in sensor specification sheets) in the testing report. A future enhanced testing protocol may be designed in which a PM sensor is challenged with zero air followed by small step changes in PM concentrations to determine the point at which the sensor starts to respond reliably and systematically and to look for any observable change in the slope of response. However, air sensors often respond to changes in PM concentration more quickly than a test chamber can obtain equilibrium and/or an FEM instrument can confirm it. It would be advantageous for this test, as well for rise/lag testing, if PM sensors were designed with a remote activation feature so that they can be switched on remotely after the test chamber has been equilibrated. Most sensors on the market today do not offer this feature.

Drift testing protocols were informed by other sensor evaluation tests. The low concentration drift test in this testing protocol uses a low $PM_{2.5}$ concentration test rather than zero air because some sensors do not provide true measurements for zero air. This concentration was chosen to be above the DL of most air sensors (as specified by sensor manufacturers) and set around the AQI green/yellow (good/moderate) breakpoint. The mid concentration drift test setpoint was determined based on the yellow/orange (moderate/unhealthy for sensitive groups) breakpoint in the AQI where sustained concentration measurements would be important for health messaging. Additionally, the mid concentration $PM_{2.5}$ setpoint of 35 $\mu g/m^3$ is the current primary (health-based), 24-hour National Ambient Air Quality Standard (NAAQS). A 60-day test, or aging, period was a compromise between a short-term 24-hour test period and a longer term (e.g., months to years) test period. 60-days is presumed important to measure potential changes in sensor performance over a longer timeframe but not too long as to be an undue burden and hinder the completion of sensor performance testing. These protocols require that sensors be aged by 60 days of continual operation in outdoor, ambient air to be most representative of routine operation with good variation in T and RH conditions.

Accuracy at High Concentrations testing protocols were informed by other sensor evaluation tests. The testing protocol in this report adds more calibration points at the higher AQI breakpoints (in addition to a low and mid concentration) but prescribes that this test be conducted last as some literature indicates that exposure to high concentrations can accelerate sensor aging and reduce sensor response, both of which can damage the sensor. The high $PM_{2.5}$ concentration setpoints were determined based on the red/purple (unhealthy/very unhealthy) and the purple/maroon (very unhealthy/hazardous) breakpoints in the AQI where sustained high concentration measurements would be important for health messaging. Additionally, the high concentration setpoint (150 $\mu g/m^3$) has often been measured under ambient conditions in some U.S. locations and the higher concentration setpoint (250 $\mu g/m^3$) has often been measured in areas impacted by wildfire smoke.

Appendix C: Supporting Information for Performance Metrics

As mentioned in Section 3.0, the performance metrics selected were based on workshop discussions, literature reviews, performance specifications for FRM/FEM monitors, and metrics being employed or considered by other organizations that are implementing sensor standards/certification programs. These metrics are deemed important for providing information on a range of properties that describe the performance of air sensors, while also recognizing that it may not be practical to include every possible metric due to cost and time considerations. Some of the metrics recommended are best assessed under controlled, laboratory conditions. It should be noted that air sensors currently do not have testing requirements nor conform to the U.S. EPA FRM/FEM Program quality assurance (QA) protocols. Some of the metrics recommended are not included in the FRM/FEM certification Program. The metrics presented in this report are recommended in order to better understand, and account for, the unknown data quality from air sensor devices. More details are provided below on the recommended performance metrics for base testing and enhanced testing.

Base and Enhanced Testing Performance Metrics

Precision is a measure of how a set of identical air sensors perform relative to each other and how closely the sensor concentrations agree. The better the precision, the less variability will be seen between any randomly chosen set of identical sensor devices. Two possible statistical expressions of precision are standard deviation (SD), reported in the units of measurement, or coefficient of variation (CV), reported as a percentage when divided by the mean and then multiplied by 100. These expressions of precision were chosen as they are used in 40 CFR Section 53.58.

Bias is not commonly discussed explicitly in sensor evaluation studies; however, it is common practice to perform a linear regression to determine slope and intercept. For FRM/FEM monitors, bias is included as a performance metric and is represented as slope and intercept [40 CFR Section 53.35(g)]. Bias quantifies systemic under- or over-reporting of air sensor measurements from true values determined by reference instruments. Poor calibration can be one source of such a systematic error.

Linearity was calculated with linear regression (rather than orthogonal regression) to determine the correlation of the collocated sensors and reference instrument measurements. This is a common metric used in the sensor evaluation programs and in literature. Further, simple linear regression is simpler and more familiar than orthogonal regression. Additionally, the coefficient of determination (R^2) is calculated instead of the Pearson correlation coefficient (r) because R^2 indicates the proportion of variability in the dependent variable that is predicted from the independent variable; r only describes the degree of linear correlation. One major limitation of the use of R^2 is that an instrument can score well on this measure (near to 1, which indicates perfect agreement) but still be very inaccurate. To help compensate for this limitation, other metrics like error and bias are also used.

Error can be described by several metrics including standard error, absolute error, mean absolute error, root mean square error, and normalized root mean square error. Each metric has its merits but, this report requests that the root mean square error (RMSE) or the normalized root mean square error (NRMSE) be calculated (the greater of the two values should be reported). RMSE penalizes large deviations of the sensor measurements from the reference instrument measurements and is therefore, sensitive to outliers. As a caution, RMSE is not defined consistently based on available resources. It has commonly been defined in two ways: 1) describing the difference between a measurement and the true value, and 2) describing the difference between a measurement and a linear regression best fit line of a measurement and a corresponding true value. In this report, RMSE is defined as the error between the sensor measurement and the reference instrument measurement (true value). This approach is presumed to provide the best indication of out-of-the-box sensor performance and the error that can be expected prior to any data corrections. Further, this approach is how RMSE is commonly calculated in air sensor literature to date. NRMSE was included to account for testing conditions where the ambient concentrations may be much higher than typical U.S. ambient levels (e.g., wildfires).

Exploring Meteorological Effects allows for a greater understanding of sensor performance in different conditions. The greater the variety in conditions, the better the understanding of how a sensor might perform in different environments. Analyzing sensor response with respect to temperature (T) and relative humidity (RH) are common exploratory analyses conducted to better understand air sensor performance using field data. Some air sensors show a dependence on T or RH when comparing sensor measurements with reference instrument measurements. Aerosols may swell due to uptake of water under higher humidity conditions. Fog conditions can cause the sensor to respond more to atmospheric water than particles. Passive heat generated by a sensors can cause condensation in the sample stream going through the sensor on high T and high dew point (DP) days. Current evidence suggests that the magnitude of these impacts is variable by sensor and may be difficult to observe. Understanding meteorological impacts on sensor performance can be important for some NSIM applications or sensor environments.

Additional Enhanced Testing Performance Metrics

Effect of Relative Humidity (RH) is important to understand sensor performance because RH can introduce a positive or negative bias to sensor measurements. Because $PM_{2.5}$ sensors do not typically heat the inlet sample stream to drive off water vapor, high moisture content in the ambient, outdoor air can change the refractive indices of particulate matter (PM) in the sensor sample stream and can also lead to hygroscopic growth of particles (e.g., at high RH levels, mist or fog could be detected as PM). Understanding this response helps determine the environmental conditions that a sensor may be expected to reasonably perform and can allow for the development of corrections to address the influence of RH on sensor measurements. The AQ-SPEC program also evaluates the effect of RH.

Effect of Temperature (T) is important to understand sensor performance because T can introduce positive or negative bias in sensor response and thus cause deviation from a linear response. This can happen at both very high and low T. Given that outdoor, ambient field conditions can vary due to daily T extremes or seasonal variations, an understanding of the T response helps determine the conditions that a sensor may be expected to reasonably perform and can allow for the development of corrections to address the influence of T changes on sensor measurements. The AQ-SPEC program also evaluates the effect of T.

Drift measurement is important for understanding the magnitude by which a sensor measurement may vary over time leading to erroneous, biased, and inaccurate readings. Understanding drift allows for development of a calibration check and/or recalibration plan and may be used to compensate for changes in the sensor's response over time.

Accuracy at High Concentrations is an important metric in order to evaluate the suitability of a sensor for NSIM applications where high $PM_{2.5}$ concentrations are expected (e.g., wildfires, dust storms). This performance metric helps determine the degree to which sensor measurements can be trusted in high $PM_{2.5}$ environments.

Appendix D: Supporting Information for Target Values

As mentioned in Section 4.0, the target values were informed by the following:

- Workshop discussions;
- The FRM/FEM certification program (Table D-1);
- The U.S. EPA's findings on air sensor evaluations (Table D-2);
- AQ-SPEC air sensor field evaluations (Table D-3);
- Peer-reviewed literature reporting data quality levels (Table D-4); and
- Sensor standards/certification programs in development by other organizations (Table D-5).

Table D-1. Performance Requirements for $PM_{2.5}$ FRM/FEM Regulatory Monitors (adapted from U.S. EPA, 2018)

Performance Attribute	Specification for Regulatory Monitoring	Notes (Based on 40 CFR Part 53 Subpart C and 40 CFR Part 50)
Accuracy/Uncertainty	R^2: 0.7225-0.9025 Slope: 1 ± 0.10 Intercept: 0 ± 2 μg/m^3	
Measurement Range	3-200 μg/m^3	Referred to as 'acceptable concentration range (R_j)'.
Detection Limit	2 μg/m^3	Referred to as 'lower detection limit'.
Precision	CV_{conc}: ≤ 5% SD: ≤ 2 μg/m^3 Root mean square: 15%	CV_{conc} represents the concentration coefficient of variation.

Table D-2. Summary of U.S. EPA $PM_{2.5}$ Sensor Evaluation Field Results (24-hour Average)

Sensor Manufacturer/ Model	Sensor Concentration Range (μg/m³)	Reference Concentration Range (μg/m³)	Precision (CV, %)	Slope*	Intercept* (μg/m³)	$R^{2,*}$	RMSE (μg/m³)
Aeroqual/AQY	1-12	4-15	15.32†	0.61 (0.42 to 0.87)	-0.99 (-2.0 to -0.19)	0.77 (0.70 to 0.89)	4.52
AirVisual/ AirVisual Pro	1-16	3-21	16.90	0.95 (0.88 to 1.07)	-1.51 (-1.81 to -1.19)	0.82 (0.65 to 0.91)	2.41
Airviz/Speck v2	0-25			-6.54	119.11	0.03	
APT/Maxima	1-24	3-16	9.91	1.83 (1.69 to 1.96)	-7.72 (-.823 to -6.42)	0.89 (0.87 to 0.95)	3.50
Cairpol/CairClip PM Prototype	0-25			-0.01	0.13	0.01	
Clarity/Node	2-21	5-15	13.32	1.84 (1.42 to 2.07)	-5.94 (-6.84 to -4.69)	0.84 (0.76 to 0.88)	3.59
Clarity/Node-S	3-27	5-15	4.62	2.28 (2.18 to 2.38)	-5.38 (-5.62 to -5.12)	0.77 (0.75 to 0.79)	7.64
Dylos/DC 1100	0-30					0.42 (0.40 to 0.46)	
HabitatMap/AirBeam	0-30					0.47 (0.45 to 0.48)	
MetOne/831	0-30			0.30 (0.27 to 0.31)	0.15 (0.01 to 0.18)	0.17 (0.15 to 0.19)	
PurpleAir/PA-II-SD	2-36	3-17	0.89	2.27 (2.26 to 2.29)	-6.72 (-6.61 to -6.88)	0.81 (0.81 to 0.81)	6.52
Sensit/RAMP	0-4	2-16	8.67‡	0.29 (0.24 to 0.37)	-1.08 (-1.44 to -0.90)	0.91 (0.91 to 0.93)	7.07
Shinyei/PM Evaluation Kit	0-30			0.52 (0.47 to 0.56)	1.31 (-0.08 to 2.70)	0.36 (0.31 to 0.41)	
Vaisala/AQT420	1-5	4-20	31.03	0.01 (-0.02 to 0.07)	1.71 (1.41 to 1.89)	0.01 (0.00 to 0.04)	6.98
Wicked Device/Air Quality Egg	0-30			0.04 (-1.80 to 1.80)	129.07 (2.00 to 297)	0.06 (0.01 to 0.16)	

Note: Data from Frederick et al. (2020a and 2020b). The field evaluations were conducted in Research Triangle Park, NC.

*Values represent mean with range in parenthesis.

†Not all AQY sensors were deployed at the same time. CV of 8.37% was calculated for sensors 317A, 318A, and 319A. A subsequent deployment with devices 522 through 527, resulted in a CV of 15.32%. The higher value is reported here as an upper estimate for CV.

‡RAMP devices SN 1016, 1017, 1018, 1019, 1020, 0181 were deployed concurrently and resulted in a CV of 8.67%. RAMP devices SN 1015, 0182, 0183 were deployed concurrently and resulted in a CV of 5.65%. The higher value is reported here as an upper estimate for CV.

Table D-3. Summary of AQ-SPEC $PM_{2.5}$ Sensor Field Evaluation Results (24-hour Average)

Sensor Manufacturer/ Model/Version	Concentration Range ($\mu g/m^3$)	Slope*	Intercept* ($\mu g/m^3$)	$R^{2,*}$
Aeroqual/AQY/v0.5	0-70	1.09 to 1.10	-4.29 to -4.18	0.92-0.93
Aeroqual/AQY/v1.0	0-25	0.46 to 0.61	-1.13 to -0.54	0.87-0.88
Aeroqual S500-PM	5-23	1.80 to 2.25	-19.68 to -14.63	0.65- 0.77
AirThinx/IAQ	0-30	1.84 to 1.94	-12.82 to -11.57	0.52-0.57
Alphasense/OPC-N2	5-30	1.05 to 1.26	-8.16 to -2.88	0.76-0.90
Alphasense/OPC-N3	0-35	0.70 to 0.84	-2.86 to -2.13	0.75-0.81
AQMesh/v3.0	5-22	0.84 to 1.18	-9.51 to -1.63	0.64-0.85
AS-LUNG/Air Quality Station	0-45	1.59 to 1.91	-8.31 to -6.46	0.86-0.89
AS-LUNG/Portable	5-45	1.50 to 1.58	-5.93 to -5.43	0.91
Atmotube/Pro	0-35	1.05 to 1.31	-3.88 to -2.60	0.91-0.92
Clarity/Node	0-30	1.26 to 1.42	-4.11 to -2.50	0.85-0.87
Dylos/DC1100 Pro	0-45	N/A	N/A	0.81
Dylos/DC1700-PM	0-40	4.12 to 4.63	22.44 to -18.73	0.75-0.78
Ecowitt/WH415B	0-20	2.28 to 3.24	-8.04 to -1.21	0.50-0.60
Edimax/Airbox	5-35	1.87 to 2.04	-12.33 to -10.58	0.59-0.65
Edimax/Edigreen Home	5-35	1.64 to 1.82	-10.68 to -8.81	0.63-0.68
Elitech/Temtop LKC-1000S+	0-35	1.35 to 1.53	-4.77 to -4.31	0.95-0.96
Elitech/Temtop M2000	0-25	0.94 to 1.17	-2.63 to -2.34	0.86-0.89
Fablab/Smart Citizen Kit/v2.1	0-30	2.47 to 2.60	-16.97 to 15.62	0.90-0.91
Foobot	5-25	2.16 to 2.78	-16.87 to -14.24	0.57-0.66
HabitatMap/AirBeam	0-40	3.18 to 6.72	-67.82 to -34.59	0.77-0.78
HabitatMap/AirBeam2	5-35	1.31 to 1.37	-7.68 to -6.73	0.84
Hanvon/Hanvon N1	0-60	2.61 to 3.06	-15.65 to -12.57	0.63-0.67
IQAir/AirVisual Pro FW1.1683	2-32	0.89 to 1.21	-5.69 to -2.35	0.86-0.93
IQAir/AirVisual Pro	5-30	1.92 to 2.10	-15.86 to -15.68	0.80-0.81
Kaiterra/Laser Egg 2+	0-20	2.20 to 2.35	-10.56 to -7.73	0.69-0.72
Kunak/Air A10	0-60	0.58 to 0.73	-0.60 to -0.24	0.67-0.81
Lunar Outpost/Canary-S	5-60	1.33 to 1.45	-2.86 to -1.97	0.68-0.69
Magnasci SRL/uRADMonitor A3/HW105	0-34	0.81 to 1.03	-4.35 to -2.20	0.79-0.83
Magnasci SRL/uRADMonitor INDUSTRIAL/HW103	0-45	0.89 to 1.20	-4.32 to -2.56	0.77-0.83
MagnasciSRL/uRADMonitor SMOGGIE-PM/v1.101	5-23	1.26 to 1.68	-7.98 to -3.60	0.55-0.66
MetOne/Neighborhood Monitor	8-27	1.65 to 2.14	-21.09 to -18.18	0.65-0.66
Moji China/Airnut	3-37	1.03 to 1.19	-5.50 to -3.96	0.65-0.80
Origins/Laser Egg	5-22	1.74 to 2.09	-13.23 to -10.66	0.66-0.77
Plume Labs/Flow 2	3-23	0.25 to 1.62	-12.17 to -0.51	0.02 to 0.72
Purple Air/PA-II	0-40	1.64 to 2.02	-6.31 to -2.83	0.93-0.97
PurpleAir/PA-I	3-32	1.35 to 1.44	-5.21 to -3.86	0.91-0.93

Sensor Manufacturer/ Model/Version	Concentration Range (μg/m^3)	Slope*	Intercept* (μg/m^3)	$R^{2,*}$
PurpleAir/PA-I Indoor	2-30	1.71 to 1.80	-7.57 to -5.86	0.85-0.86
RTI/MicroPEM	0-50	1.53 to 1.69	-14.40 to -7.94	0.77-0.91
SainSmart/Pure Morning P3	0-30	1.77 to 2.12	9.07 to -5.07	0.78-0.83
Samyoung S&C/SY-DS-DK3	0-20	2.26 to 3.01	-7.65 to -5.3	0.65-0.66
Sensirion/Nubo	0-40	1.30 to 1.36	-5.18 to -4.80	0.90-0.91
Sensirion/SPS30	0-20	1.27 to 1.30	-4.56 to -4.37	0.68-0.69
Shinyei/PM Evaluation Kit	0-45	1.28 to 1.44	-5.17 to -4.64	0.92-0.93
TSI/AirAssure	0-40	1.34 to 1.49	-3.18 to -0.28	0.87-0.89
TSI/BlueSky	0-25	0.72 to 0.91	-3.75 to -1.85	0.77-0.81
Wicked Device/Air Quality Egg/v2	0-35	0.80 to 0.92	-5.78 to -2.83	0.40-0.93

Note: These field evaluations were conducted at the Rubidoux air monitoring station in Riverside, CA. Evaluations are current as of 08/26/2020.

*AQ-SPEC presents graphical results with reference instrument measurements on the y-axis and sensor measurements on the x-axis, which is the reverse of the recommended method in this report. The results shown in this Table mathematically manipulate the equations reported by AQ-SPEC to present slopes and intercepts in a similar form to that recommended in this report. It should be noted that these results are approximate as performing a least squares regression on the data with the x-axis and y-axis variables switched will produce different results.

Table D-4. Summary of Literature Reviews of $PM_{2.5}$ Sensor Evaluations (24-hour average) used to Inform Target Values

Source	Sensor Manufacturer/ Model	Concentration Range (μg/m^3)	Slope*, †	Intercept† (μg/m^3)	$R^{2,*,\dagger}$
Nakayama et al., 2018	Panasonic	0-40	1.11 (0.97 to 1.23)		
Badura et al., 2018	Alphasense/OPC-N2	0-120	4.76	-37.52	0.60 (0.53 to 0.69)
	Plantower/PMS7003	0-120	3.57	-11.82	0.91 (0.88 to 0.93)
	Winsen/ZH03A	0-120	2.86	-7.26	0.84 (0.78 to 0.89)
	NovaFitness/SDS011	0-120	2.50	-11.65	0.88 (0.87 to 0.90)

*Values represent mean with range in parenthesis (where applicable).

†Indicates that the value was calculated based on data within the cited source. For slope and intercept, results were mathematically manipulated to present slopes and intercepts in a similar form to that recommended in this report (i.e., sensor measurements on the x-axis and reference measurements on y-axis). It should be noted that these results will be approximations. For R^2, results were calculated from a reported r value.

Using data from Tables D-2 through D-4, a summary of current air sensor capabilities from peer-reviewed literature and evaluation programs is presented in Table D-5 in conjunction with the target values recommended in this report.

Table D-5. Summary of Available Resources Used to Inform Target Values

		Precision (CV, %)	Slope	Intercept (μg/m³)	R^2	RMSE (μg/m³)
Literature and Evaluation Data*	Range†	0.89 to 31.03	0.50 to 1.49	-19.08 to 0.91	0.52 to 0.97	2.41 to 7.64
	Average†	12.78	1.09	-3.75	0.80	5.28
	Median†	11.62	1.12	-3.19	0.83	5.52
People's Republic of China MEE		5% (relative SD)	Not discussed/ not listed	Not discussed/not listed	≥ 0.64	Not discussed/ not listed
This Report	Target (Base Testing Only)	≤ 5 μg/m³ (SD) OR ≤ 30% (CV)	1.0 ± 0.35	-5 ≤ b ≤ 5	≥ 0.70	RMSE ≤ 7 μg/m³ or NRMSE ≤ 30%‡

*Data only includes 24-hour averaged data from field evaluations.
†Several values were excluded when summarizing these data. Slopes outside of 0.5 to 1.5 were not considered; the intercept was not considered if the slope was discarded. R^2 values greater than or equal to 0.5 were considered; R^2 values less than 0.5 were not considered.
‡A sensor will meet this target if either the RMSE or NRMSE meet this criterion. See Appendix D for further discussion.

Rationale is provided below for the recommended target values for each of the recommended performance metrics.

Precision: ≤ 5 μg/m³ (SD) OR ≤ 30% (CV)

Precision metrics in peer-reviewed literature are presented in a variety of manners leading to difficulty in comparing results. Common to all sources, strong precision between sensors is defined by sensors behaving in a similar manner to environmental conditions thus minimizing variations between individual sensor devices and measurement noise. Using the metrics identified in this report, the strongest precision possible is reflected by 0 μg/m^3 SD or 0% CV which means that all the sensors respond identically to environmental conditions. Strong precision is needed in a wide range of NSIM applications, especially those where concentrations from one sensor must be compared to that of another. While it is best to have precision values close to 0, precision is just one metric and needs to be viewed in conjunction with other metrics to better understand sensor performance. For example, if all sensors give measurements of zero regardless of the $PM_{2.5}$ concentration in the environment, they have perfect precision even though the $PM_{2.5}$ sensors are non-functional. Sensor evaluation programs and $PM_{2.5}$ monitor evaluations both note there may be slight differences between measurements from different devices, which is allowed for within a range of precision target values.

Slope (Bias): 1.0 ± 0.35

The target value for the slope component of bias is 1.0 ± 0.35. This goal indicates that $PM_{2.5}$ sensors show roughly the same increase or decrease in $PM_{2.5}$ concentration as the reference device, making changes in concentration comparable between the two different devices. This is extremely important for NSIM applications where relative difference, or the amount of change, is important. The target value proposed in this report prescribes a confidence interval to assist testers in evaluating performance.

Intercept (Bias): -5 ≤ b ≤ 5 μg/m^3

The target value for the intercept component of bias is near 0 ± 5 μg/m^3. This goal ensures that low concentrations measurements are still meaningful, and that systemic error is minimized. The target value proposed in this report prescribes a confidence interval to assist testers in evaluating performance.

R^2 (Linearity): > 0.70

Higher R^2 values are associated with closer agreement and better linearity between two data sets being compared. The target value for R^2 here is ≥ 0.70. R^2 should be considered in conjunction with other metrics because high linearity does not necessarily indicate perfect agreement between datasets (e.g., two datasets can have an R^2 close to 1 with a linear regression slope of 2, as a result of different absolute concentration values between the data sets). Care should be taken in interpreting R^2, as poor linearity can result from various reasons such as a non-linear relationship, outliers, or lack of precision in the sensor or reference instrument. Linearity can also be strongly influenced by just a few high concentration measurements.

RMSE or NRMSE (Error): ≤ 7 μg/m^3 (RMSE) or ≤ 30% (NRMSE)

RMSE quantifies the random disagreement between the pollutant concentrations reported by the sensor and the reference instrument, thus values closer to zero indicate better agreement and less uncertainty in the measurement. This is an important metric for NSIM applications where sensor and reference instrument concentrations need to be compared. RMSE is sensitive to data points with large differences between sensor and reference instrument concentrations. The target value for RMSE is ≤ 7 μg/m^3. Care should be taken to use the definition and recommended calculation for RMSE that is provided in this report (Sections 3.1.5 and 3.2.5). A target value for NRMSE of ≤ 30% is also included to account for testing conditions where the ambient $PM_{2.5}$ concentrations may be much higher than typical U.S. ambient levels (e.g., wildfires). Under wildfire smoke conditions, the RMSE metric may exceed the target because concentrations are high but the NRMSE may meet this target instead. Both RMSE and NRMSE may be reported, but the tester may report the more favorable result (the result meeting or closest to this target metric). The sensor will meet this target if either the RMSE or NRMSE meet this criterion.

Effect of Relative Humidity (RH): No Target Value Established

The effect of RH on sensor measurements is an important performance metric for all NSIM applications especially because RH varies across the U.S. and can change rapidly throughout the day. Many studies have shown that currently available $PM_{2.5}$ sensors are affected by RH (U.S. EPA, 2015; Jayaratne et al., 2018; Zheng et al., 2018; Feenstra et al., 2019; AQ-SPEC PM sensor evaluations). Because $PM_{2.5}$ sensors do not typically heat the inlet sample stream to drive off water vapor, high moisture content in the ambient, outdoor air can change the refractive indices of particulate matter (PM) in the sample stream and can also lead to hygroscopic growth of particles (e.g., at high RH levels, mist or fog could be detected as PM). Literature sources attempting to quantify the effects of RH did not report results in a consistent manner thus, a target level has not been established. The protocols outlined in this report request that testers quantify the influence of RH in a systematic way. This work may inform the future establishment of a target value.

Effect of Temperature (T): No Target Value Established

The effect of ambient T on sensor measurements is an important performance metric for all NSIM application areas because T varies significantly throughout the day, across different seasons, and across the U.S. Considering the very limited performance data available, this report does not establish a target value for this metric. The protocols outlined in this report request that testers quantify the influence of T in a systematic way which may inform the future establishment of a target value.

Drift: No Target Value Established

While little to no drift is ideal, the available information and literature suggests that measurements from many air sensors may drift over time. The literature suggests that drift may occur abruptly or steadily as the sensor ages, on the order of days, months, or years. The rate of drift is currently understood to be highly variable, may depend on the concentrations experienced, and may still occur even if the sensor is not being used. The rate and degree of drift has not been systemically quantified in the literature. Currently, there has been little testing on drift in air sensors on the 60-day scale therefore, this report does not establish a target value for drift. The protocols outlined in this report request that testers quantify the influence of drift in a systematic way after 60-days of operation in an outdoor, ambient environment. A 60-day evaluation period is recommended to reduce the burden on testers. These results will help establish whether drift can be observed within a 60-day period and may inform the future establishment of a target value.

Accuracy at High Concentrations: No Target Value Established

Many sensor manufacturers/developers claim the ability to accurately measure $PM_{2.5}$ concentrations at high concentrations. Discussions with groups that evaluate air sensors suggests that sensor measurements are more likely to differ from reference instrument measurements at high concentrations. Few field measurements are made at high concentrations because they occur less frequently. Understanding how accurately a sensor performs during higher $PM_{2.5}$ concentrations is important for areas that experience such conditions, for NSIM applications focused on exceptional events (e.g., wildfires, dust storms), and for verifying whether potential corrections still apply at higher concentrations. A target value has not been established at this time.

Appendix E: Checklist for Base Testing

Data Collected Before Base Testing (Sections 2.1.1 through 2.1.3)

☐ Testing Organization(s) Name and Contact Information [email, phone number, and/or website]
☐ Testing location [City, State; Latitude, Longitude; AQS site ID (*if applicable*)]
☐ Information about air sensor spacing relative to the FRM/FEM monitor and other air sensors
☐ Information about any weather-protective shelter/enclosure used (*if applicable*)
☐ Relative humidity (RH), temperature (T), and FRM/FEM monitor information, including:

Item (*as applicable*)	**RH Monitor**	**T Monitor**	**$PM_{2.5}$ FRM/FEM Monitor**	**Other FRM/FEM Monitor(s)**
Manufacturer/Model				
Firmware Version				
Parameter(s) Measured and Units				
Sampling Time Interval				
Manufacturer Specification Sheet				
Copy of Calibration Certificate				

☐ Air sensor equipment information, including:

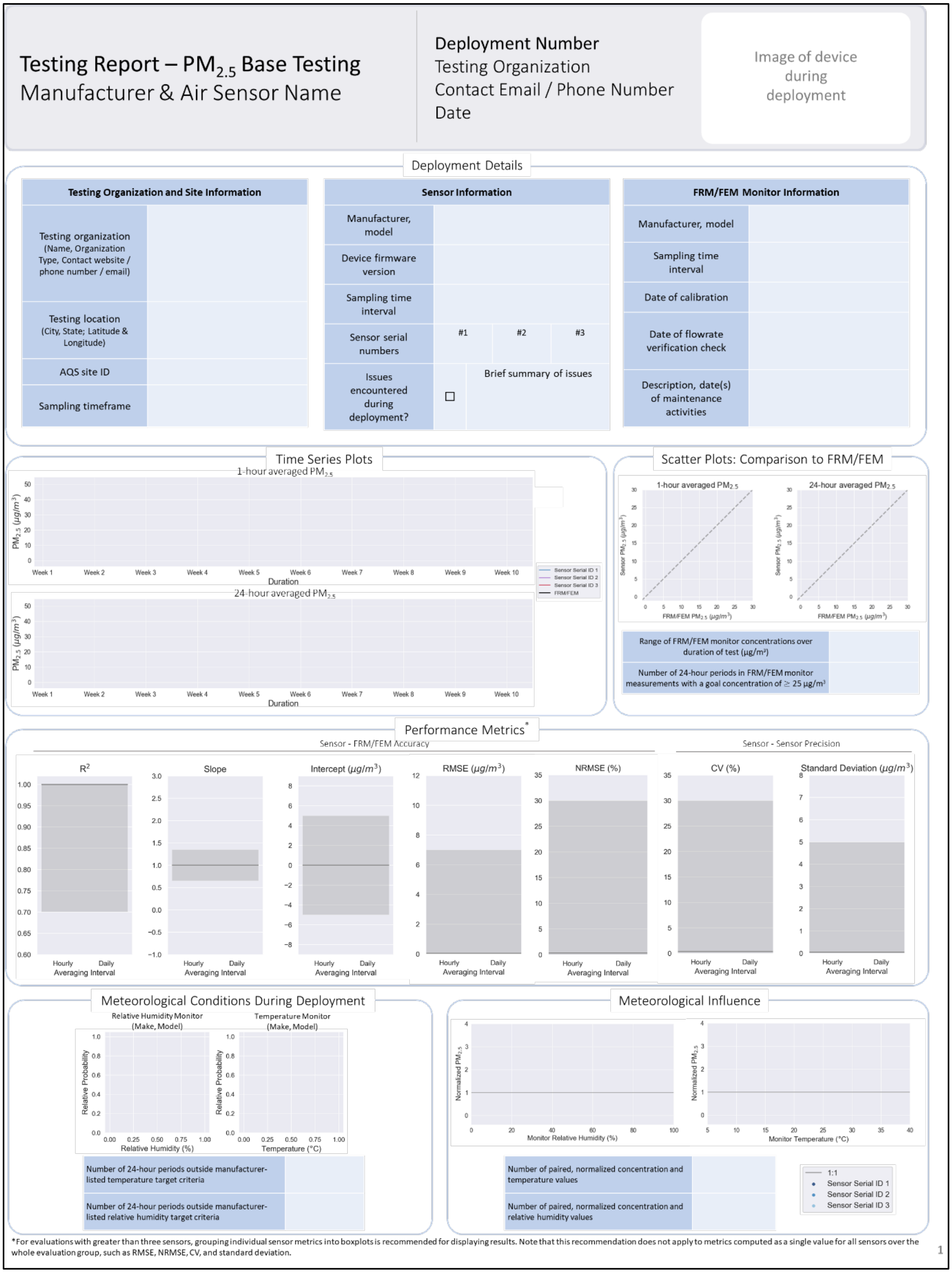
Item (as applicable)		**Sensor 1**	**Sensor 2**	**Sensor 3**
General Information	Manufacturer/Model			
	Firmware Version			
	Serial/Identification Number			
	Parameter(s) Measured and Units			
	Sampling Time Interval			
	Manufacturer Specification Sheet			
Data Storage Information	Where the data are stored			
	Where the data are transmitted			
	Form of data stored			
Data Correction Approach	Procedure used to correct data			
	If data correction does not change or is static, record approach			
	If data correction does change or is dynamic, record approach			
Data Analysis/Correction Script	Script used and version			
Final Data Reported	Location of final data			
	Format of final data			

☐ Photo(s) of entire equipment set up at test site

Data Collected During Base Testing (Section 2.1.4)

☐ Deployment number and sampling timeframe
☐ Dates for calibration and one-point flow rate verification check on the FRM/FEM monitor
☐ At least 30 consecutive days of measurements
☐ Description of QC criteria (*as applicable*)
☐ Time, dates, description, and rationale for any of the following (*as applicable*): 1) maintenance, 2) missing or invalidated data, and 3) any other issue(s) impacting data collection

Appendix F: Example Reporting Template for Base Testing



Testing Report – $PM_{2.5}$ Base Testing
Manufacturer & Air Sensor Name

Deployment Number
Testing Organization
Contact Email / Phone Number
Date

Image of device during deployment

Deployment Details

Testing Organization and Site Information	
Testing organization (Name, Organization Type, Contact website / phone number / email)	
Testing location (City, State; Latitude & Longitude)	
AQS site ID	
Sampling timeframe	

Sensor Information			
Manufacturer, model			
Device firmware version			
Sampling time interval			
Sensor serial numbers	#1	#2	#3
Issues encountered during deployment?	☐	Brief summary of issues	

FRM/FEM Monitor Information	
Manufacturer, model	
Sampling time interval	
Date of calibration	
Date of flowrate verification check	
Description, date(s) of maintenance activities	

Time Series Plots

1-hour averaged $PM_{2.5}$

24-hour averaged $PM_{2.5}$

Scatter Plots: Comparison to FRM/FEM

Range of FRM/FEM monitor concentrations over duration of test (µg/m³)	
Number of 24-hour periods in FRM/FEM monitor measurements with a goal concentration of ≥ 25 µg/m³	

Performance Metrics*

Meteorological Conditions During Deployment

Number of 24-hour periods outside manufacturer-listed temperature target criteria	
Number of 24-hour periods outside manufacturer-listed relative humidity target criteria	

Meteorological Influence

Number of paired, normalized concentration and temperature values	
Number of paired, normalized concentration and relative humidity values	

*For evaluations with greater than three sensors, grouping individual sensor metrics into boxplots is recommended for displaying results. Note that this recommendation does not apply to metrics computed as a single value for all sensors over the whole evaluation group, such as RMSE, NRMSE, CV, and standard deviation.

1

Testing Report – PM$_{2.5}$ Base Testing
Manufacturer & Air Sensor Name

Deployment Number
Testing Organization
Contact Email / Phone Number
Date

Image of device during deployment

Tabular Statistics

- Sensor – FRM/FEM Correlation

	Bias and Linearity						Data Quality			
	R^2		Slope		Intercept (b) (µg/m³)		Uptime (%)		Number of paired sensor and FRM/FEM concentration values	
	1-Hour ○○○	24-Hour ○○○	1-Hour ○○○	24-Hour ○○○	1-Hour ○○○	24-Hour ○○○	1-Hour ○○○	24-Hour ○○○	1-Hour	24-Hour
Metric Target Range	≥ 0.70	≥ 0.70	1.0 ± 0.35	1.0 ± 0.35	-5 ≤ b ≤ 5	-5 ≤ b ≤ 5	90%*	90%*		
Sensor Serial #1										
Sensor Serial #2										
Sensor Serial #3										
Mean										

	Error			
	RMSE (µg/m³)		NRMSE (%)	
	1-Hour ○	24-Hour ○	1-Hour ○	24-Hour ○
Metric Target Range	≤ 7	≤ 7	≤ 30	≤ 30
Deployment Value				

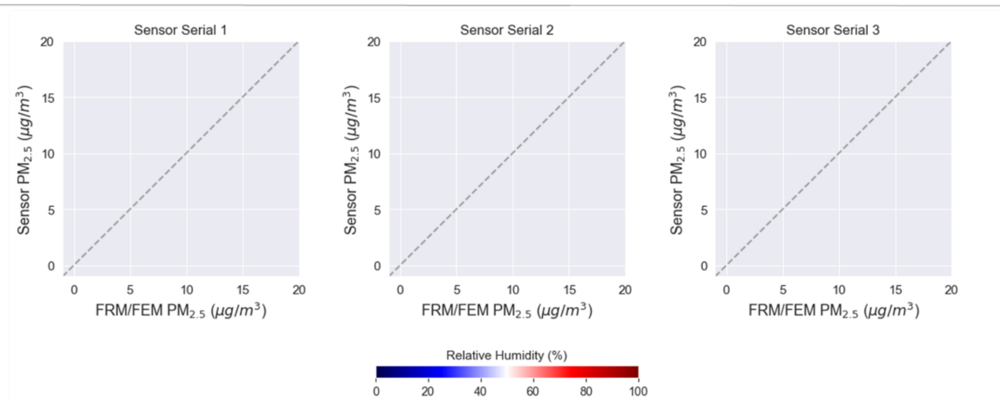
Device-specific metrics (computed for each sensor in evaluation)
○○○ Metric value for none of devices tested falls within the target range
●○○ Metric value for one of devices tested falls within the target range
●●○ Metric value for two of devices tested falls within the target range
●●● Metric value for three of devices tested falls within the target range

Single-valued metrics (computed via entire evaluation dataset)
○ Indicates that the metric value is not within the target range
● Indicates that the metric value is within the target range

- Sensor – Sensor Precision

	Precision (between collocated sensors)				Data Quality			
	CV (%)		SD (µg/m³)		Uptime (%)		Number of concurrently reported sensor concentration values	
	1-Hour ○	24-Hour ○	1-Hour ○	24-Hour ○	1-Hour ○	24-Hour ○	1-Hour	24-Hour
Metric Target Range	≤ 30	≤ 30	≤ 5	≤ 5	90%*	90%*		
Deployment Value								

Individual Sensor – FRM/FEM Scatter Plots



*This value is only a recommendation for ensuring data quality and is not included in the list of target values discussed in Section 4 of the *Performance Testing Protocols, Metrics, and Target Values for Fine Particulate Matter Air Sensors* document.

2

Testing Report – $PM_{2.5}$ Base Testing
Manufacturer & Air Sensor Name

Deployment Number
Testing Organization
Contact Email / Phone Number
Date

Image of device during deployment

Supplemental Information

Additional documentation may be attached or linked to digital versions alongside this report. Such documentation may include field reports and observations during the testing period, maintenance logs for sensors and FRM/FEM monitors, standard operating procedures, and other documentation relevant to this testing report (see below for examples).

Supplemental Documentation	Attached	Description & URL or file path to documentation
Field observations	☐	
Maintenance logs	☐	
Standard operating procedure(s)	☐	
Photos of equipment setup and testing	☐	
Product specification sheet(s)	☐	
Product manual(s)	☐	
Deployment issues	☐	
Data storage and transmission method	☐	
Data correction approach	☐	
Data analysis/correction scripts and version	☐	
Air Monitoring Station QAPP	☐	
Summary of FRM/FEM monitor QC checks	☐	
Other documents	☐	

3

Note: A fillable reporting template for base testing is also available with this report. See accompanying PowerPoint file.

Appendix G: Checklist for Enhanced Testing

Data Collected Before Enhanced Testing (Sections 2.2.1 and 2.2.2)

☐ Testing organization(s) name and contact information [email, phone number, and/or website]
☐ Testing address/location [City, State]
☐ Description of all chamber specifications and characterization
☐ Relative humidity (RH), temperature (T), and FEM monitor information, including:

Item (*as applicable*)	**RH Monitor**	**T Monitor**	**$PM_{2.5}$ FEM Monitor**	**Other FEM Monitor(s)**
Manufacturer/Model				
Firmware Version				
Parameter(s) Measured and Units				
Sampling Time Interval				
Manufacturer Specification Sheet				
Copy of Calibration Certificate				
Date of Calibration at Test Location				
Date of one-point flow rate verification check				

☐ Air sensor equipment information, including:

Item (as applicable)		**Sensor 1**	**Sensor 2**	**Sensor 3**
General Information	Manufacturer/Model			
	Firmware Version			
	Serial/Identification Number			
	Parameter(s) Measured and Units			
	Sampling Time Interval			
	Manufacturer Specification Sheet			
Data Storage Information	Where the data are stored			
	Where the data are transmitted			
	Form of data stored			
Data Correction Approach	Procedure used to correct data			
	If data correction does not change or is static, record approach			
	If data correction does change or is dynamic, record approach			
Data Analysis/Correction Scripts	Script used and version			
Final Data Reported	Location of final data			
	Format of final data			

☐ Photo(s) of entire equipment set up in exposure chamber (optional)

Data Collected During Enhanced Testing (Sections 2.2.3 through 2.2.7)

☐ All time-matched data points for each testing condition
☐ Description of QC criteria (*as applicable*)
☐ Time, dates, description, and rationale for any of the following (*as applicable*): 1) maintenance, 2) missing or invalidated data, and 3) any other issue(s) impacting data collection

Appendix H: Example Reporting Template for Enhanced Testing

Testing Report – $PM_{2.5}$ Enhanced Testing
Manufacturer & Air Sensor Name

Testing Organization
Contact Email / Phone Number
Date

Image of device during chamber evaluation

Testing Details

Testing Organization and Contact Information	
Testing organization (Name, Organization Type)	
Contact Information (Website, Phone Number, Email)	

$PM_{2.5}$ FEM Monitor Information	
Manufacturer, model	
Sampling time interval	
Date of calibration	
Date of flowrate verification check	

Sensor Information			
Manufacturer, model			
Device firmware version			
Sampling time interval			
Sensor serial numbers	#1	#2	#3
Manufacturer listed detection limit			
Manufacturer listed longevity, lifespan			
Manufacturer listed drift			

Attached Documentation	
FEM monitor documentation	
Description, date(s) of maintenance activities	☐
Sensor documentation	
Product specification sheet	☐
Product manual	☐
Description of parameters measured and units, and data flow	☐
Data storage and transmission method	☐
Data analysis/correction script and version	☐
Testing chamber documentation	
Description of chamber and $PM_{2.5}$ test aerosol generator system	☐

Effect of Relative Humidity (RH)

RH Monitor		
	Manufacturer	
	Model	

			Average RH (%)	Average T (°C)	Average FEM monitor $PM_{2.5}$ concentration of test aerosol (μg m^{-3})	Average sensor $PM_{2.5}$ concentration (μg m^{-3})	Averaged influence of RH on sensor measurements (μg m^{-3})
Effect of RH	Initial Testing Conditions	Setpoint	40 ± 5	20 ± 1	35 ± 5%		
		Measured Value					
	High RH Conditions	Setpoint	85 ± 5	20 ± 1	35 ± 5%		
		Measured Value					

Testing Report – $PM_{2.5}$ Enhanced Testing
Manufacturer & Air Sensor Name

Testing Organization
Contact Email / Phone Number
Date

Image of device during chamber evaluation

Effect of Temperature (T)

T Monitor	Manufacturer	
	Model	

			Average RH (%)	Average T (°C)	Average FEM monitor $PM_{2.5}$ concentration of test aerosol (μg m^{-3})	Average sensor $PM_{2.5}$ concentration (μg m^{-3})	Averaged influence of T on sensor measurements (μg m^{-3})
Effect of T	Initial Testing Conditions	Setpoint	40 ± 5	20 ± 1	35 ± 5%		
		Measured Value					
	High T Conditions	Setpoint	40 ± 5	40 ± 1	35 ± 5%		
		Measured Value					

60-Day Low Concentration Drift

			Average RH (%)	Average T (°C)	Average FEM monitor $PM_{2.5}$ concentration of test aerosol (μg m^{-3})	Average sensor $PM_{2.5}$ concentration (μg m^{-3})	Sensor drift after 60 days (μg m^{-3})
60-Day Low Concentration Drift	Day 1 Date	Setpoint	40 ± 5	20 ± 1	10 ± 10%		
		Measured Value					
	Day 60 Date	Setpoint	40 ± 5	20 ± 1	10 ± 10%		
		Measured Value					

60-Day Mid Concentration Drift

			Average RH (%)	Average T (°C)	Average FEM monitor $PM_{2.5}$ concentration of test aerosol (μg m^{-3})	Average sensor $PM_{2.5}$ concentration (μg m^{-3})	Sensor drift after 60 days (μg m^{-3})
60-Day Mid Concentration Drift	Day 1 Date	Setpoint	40 ± 5	20 ± 1	35 ± 5%		
		Measured Value					
	Day 60 Date	Setpoint	40 ± 5	20 ± 1	35 ± 5%		
		Measured Value					

Testing Report – $PM_{2.5}$ Enhanced Testing
Manufacturer & Air Sensor Name

Testing Organization
Contact Email / Phone Number
Date

Image of device during chamber evaluation

Accuracy at High Concentrations

		Average RH (%)	Average T (°C)	Average FEM monitor $PM_{2.5}$ concentration of test aerosol (µg m^{-3})	Average sensor $PM_{2.5}$ concentration (µg m^{-3})	Test averaged difference between sensor and FEM $PM_{2.5}$ concentrations (µg m^{-3})
Accuracy at High Concentrations	Setpoint	40 ± 5	20 ± 1	150 ± 5%		
	Measured Value					
	Setpoint	40 ± 5	20 ± 1	250 ± 5%		
	Measured Value					

3

Note: A fillable reporting template for enhanced testing is also available with this report. See accompanying PowerPoint file.

EPA
United States
Environmental Protection
Agency

Office of Research and
Development (8101R)
Washington, DC 20460

PRESORTED
STANDARD POSTAGE
& FEES PAID EPA
PERMIT NO. G-35