United States Environmental Protection Agency

EPA/600/R-20/279 | February 2021

Performance Testing Protocols, Metrics, and Target Values for Ozone 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-20/279
February 2021

Performance Testing Protocols, Metrics, and Target Values for Ozone 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 in the report 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

Disclaimer ii

List of Figures vi

List of Tables vii

Acronyms and Abbreviations ix

Acknowledgments x

Executive Summary xi

1.0 Introduction 1

1.1 Background 1

1.2 Motivation 2

1.3 Objectives 3

2.0 Performance Testing Protocols for O_3 Air Sensors 5

2.1 Base Testing 6

2.1.1 Materials and Equipment 6

2.1.2 Selecting and Setting Up a Test Site 7

2.1.3 Setting Up the Air Sensors 9

2.1.4 Conduct Base Testing 10

2.2 Enhanced Testing 11

2.2.1 Materials and Equipment 12

2.2.2 Equipment Set Up in Exposure Chamber 14

2.2.3 Initial Testing Conditions 15

2.2.4 Effect of Interferents 15

2.2.5 Effect of Relative Humidity (RH) 16

2.2.6 Effect of Temperature (T) 16

2.2.7 Drift 17

2.2.7.1 Drift (Day 1) – Low and Mid Concentration Drift 17

2.2.7.2 Drift (Day 60) to Evaluate Sensor Aging 18

2.2.8 Accuracy at High Concentration 19

3.0 Performance Metrics and Supporting Calculations for Evaluating O_3 Air Sensors 20

3.1 Base Testing Calculations 22

3.1.1 Hourly Averages 22

3.1.2 Deployment Averages 22

3.1.3 Precision 23

3.1.4 Bias and Linearity 24

3.1.5 Error 24

3.1.6 Exploring Effect of Meteorology 25

3.1.6.1 Potential Scatter Plots 25

3.1.6.2 Normalized Concentration 26

3.1.6.3 Concentration Difference and Absolute Concentration Difference 27

3.1.6.4 Dew Point (DP) 28

3.2 Enhanced Testing Calculations 28

3.2.1 Data Averages 28

3.2.2 Test Averages 29

3.2.3 Precision 29

3.2.4 Bias and Linearity 30

3.2.5 Error 30

3.2.6 Effect of Interferents 31

3.2.7 Effect of Relative Humidity (RH) 32

3.2.8 Effect of Temperature (T) 32

3.2.9 Drift 33

3.2.10 Accuracy at High Concentration 33

4.0 Target Values for O_3 Air Sensors 34

4.1 Approach 34

4.2 List of Target Values 34

5.0 References 37

Appendix A: Definitions 40

Appendix B: Supporting Information for Testing Protocols 43

Appendix C: Supporting Information for Performance Metrics 46

Appendix D: Supporting Information for Target Values 48

Appendix E: Checklist for Base Testing 56

Appendix F: Example Reporting Template for Base Testing 57

Appendix G: Checklist for Enhanced Testing 60

Appendix H: Example Reporting Template for Enhanced Testing 61

List of Figures

Figure 2-1. Example Approach for Conducting Base and Enhanced Testing of a Single Set of Sensors 5

Figure 2-2. Overview of the Enhanced Testing Procedure 12

Figure 2-3. Drift Testing to Determine Changes After 60 days or More of Continuous Operation 17

List of Tables

Table ES-1. Recommended Testing Protocols for Understanding O3 Air Sensor Performance xi

Table ES-2. Base and Enhanced Testing – Recommended Performance Metrics and Target Values for O_3 Air Sensors xii

Table ES-3. Enhanced Testing – Additional Recommended Performance Metrics and Test Conditions for O_3 Air Sensors..... xiii

Table 1-1. NSIM Categories and Specific Examples (adapted from U.S. EPA, 2018)..... 3

Table 2-1. Test Site Selection Criteria 8

Table 2-2. Sampling Probes or Monitoring Path Siting Criteria..... 9

Table 2-3. Guidance on Air Sensor Setup at Testing Site 10

Table 2-4. Initial Testing Conditions 15

Table 2-5. Interferent Pollutant Test Concentrations 16

Table 2-6. Elevated RH Test Conditions..... 16

Table 2-7. Elevated T Test Conditions 17

Table 2-8. Low Concentration Drift Test Conditions 18

Table 2-9. Mid Concentration Drift Test Conditions 18

Table 2-10. High O_3 Concentration Test Conditions 19

Table 3-1. Summary of Recommended Performance Metrics for O_3 Air Sensors 21

Table 4-1. O_3 Sensor Performance Field Evaluation Results from Available Resources..... 34

Table 4-2. Base and Enhanced Testing – Recommended Performance Metrics and Target Values for O_3 Air Sensors Used in Ambient, Outdoor, Fixed Site NSIM Applications 35

Table 4-3. Enhanced Testing – Additional Recommended Performance Metrics and Test Conditions for O_3 Air Sensors Used in Ambient, Outdoor, Fixed Site NSIM Applications . 36

Table D-1. Performance Requirements for O_3 FRM/FEM Regulatory Monitors (adapted from U.S. EPA, 2018)..... 49

Table D-2. Summary of U.S. EPA O_3 Sensor Evaluation Results 50

Table D-3. Summary of AQ-SPEC O_3 Sensor Field Evaluation Results..... 50

Table D-4. Summary of Literature Reviews used to Inform Target Values..... 51

Table D-5. Summary of Available Resources Used to Inform Target Values 53

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
CO	Carbon monoxide
CEMM	Center for Environmental Measurement and Modeling
CEN	European Committee for Standardization
CFR	Code of Federal Regulations
CSN	Chemical Speciation Network
CV	coefficient of variation
°C	degrees Celsius
DL	detection limit
DP	dew point
Eq	equation
EU	European Union
FEM	Federal Equivalent Method
FRM	Federal Reference Method
m	slope
MEE	Ministry of Ecology and Environment (People's Republic of China)
NCore	National Core Network
NIST	National Institute for Standards and Technology
NO_2	nitrogen dioxide
NSIM	non-regulatory supplemental and informational monitoring
O_3	ozone
OAQPS	Office of Air Quality Planning and Standards
ORD	Office of Research and Development
ppbv	part(s) per billion by volume
ppmv	part(s) per million by volume
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
SO_2	sulfur dioxide
T	temperature
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, Ron Evans, Sue Kimbrough (retired), and Joann Rice. We also acknowledge the following external reviewers: John Kato, Walter Ham, and Ranjit Bhullar (California EPA/Air Resources Board); Keith Hoffman (Delaware Department of Natural Resources and Environmental Control); 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 ozone (O_3) 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 O3 Air Sensor Performance

Test Type	Setting	Description	Purpose
Base Testing	Field	Consists of two field deployments of at least three replicate O_3 air sensors with collocated FRM/FEM monitors for a minimum of 30 days each, at a single test site or two test sites. At least one deployment should occur during the typical summer O_3 season (May – August).	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 O_3 air sensors in controlled laboratory conditions to understand the effect of interferents, 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.

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 is 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 test site(s), at least one day of the testing period has a 1-hour average O_3 concentration of at least 60 parts per billion by volume (ppbv) 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 O_3 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 O_3 Air Sensors

Performance Metric		**Target Value**	
		Base Testing	**Enhanced Testing***
Precision	Standard Deviation (SD)	≤ 5 ppbv	No target values recommended; report results
	-OR- Coefficient of Variation (CV)	≤ 30%	
Bias	Slope	1.0 ± 0.2	
	Intercept (b)	-5 ≤ b ≤ 5 ppbv	
Linearity	Coefficient of Determination (R^2)	≥ 0.80	
Error	Root Mean Square Error (RMSE)	≤ 5 ppbv	

*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. See Appendix D for further discussion.

Table ES-3. Enhanced Testing – Additional Recommended Performance Metrics and Test Conditions for O_3 Air Sensors

Performance Metric	Test Conditions
Effect of Interferents	Carbon monoxide (CO): 35 ppmv ± 5%*
	Nitrogen dioxide (NO_2): 100 ppbv ± 5%
	Sulfur dioxide (SO_2): 75 ppbv ± 5%
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: 15 ppbv ± 10%
	Mid concentration: 70 ppbv ± 5%
Accuracy at High Concentration	High concentration: 125 ppbv ± 5%

*ppmv = parts per million by volume

The performance metrics and target values for base and enhanced testing are recommended based on the current knowledge of O_3 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.

It is recognized that O_3 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 O_3 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 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. 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., Sadighi et al., 2018; Masio1 et al., 2018; Feinberg et al., 2018; Masey et al., 2018; Jiao et al., 2016; Pang et al., 2017; Lin et al., 2015; Bart et al., 2014; Williams et al., 2013).

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 07/22/2020*).

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 O_3 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 varying pollutant mixtures can impact measurements;
- Understanding whether the device will measure the target pollutant in a mixture of other pollutants;
- 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 (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 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 in 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, 2014a). 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 ozone (O_3) 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 O_3 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 be best suited for 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 sensors technologies that would be best suited for their NSIM application and understand the performance of the 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 O_3 Air Sensors

The procedures outlined in this section provide standardized test protocols for evaluating the performance of O_3 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 the drift evaluation [i.e., Drift (Day 1) and Drift (Day 60)] in the enhanced testing can be conducted simultaneously.

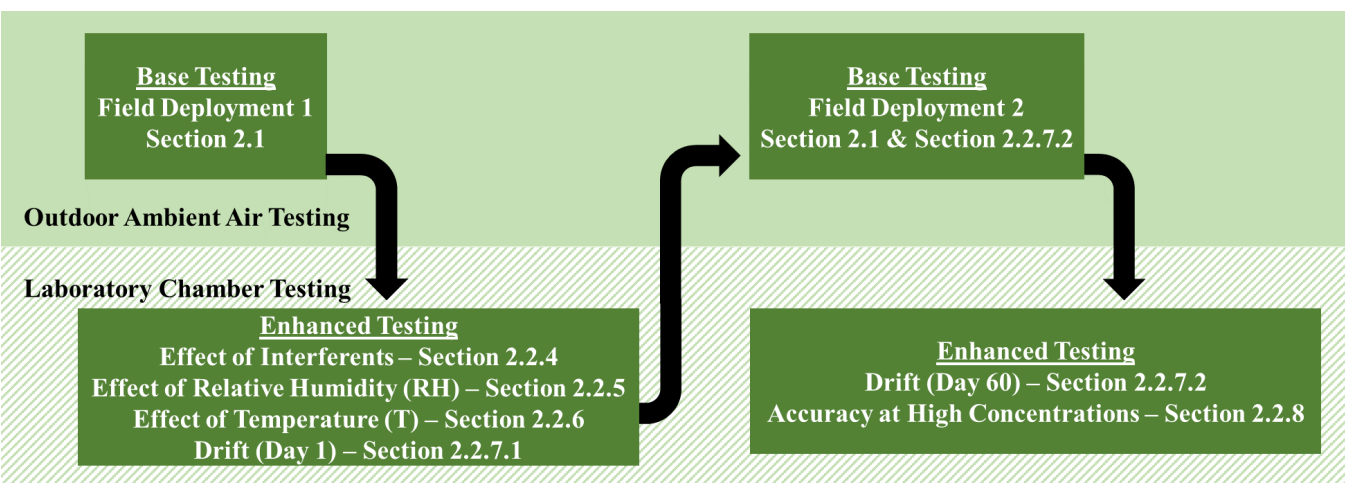


Figure 2-1. Example Approach for Conducting Base and Enhanced Testing of a Single Set of Sensors

2.1 Base Testing

Base testing consists of two (2) field deployments of O_3 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 Ambient Monitoring Technology Information Center (AMTIC) webpage (https://www.epa.gov/amtic, *last accessed 07/22/2020*). 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/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 07/22/2020*).

Base testing can occur at either a single test site during two (2) different seasons or at two (2) different test sites. The combination of field tests should demonstrate sensor performance over a range of T, RH, weather, and O_3 concentration conditions that will inform on sensor use across the U.S. For base testing, the sensor and FRM/FEM data should be compared at 1-hour averages.

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 O_3 air sensors having the same make, model, and firmware version
- Calibrated O_3 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 40 CFR Parts 50, 53, and Appendix A of Part 58. 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).

Testers may also wish to measure CO, NO_2, and SO_2 using calibrated FRM/FEM monitors at the base testing site(s). FRM/FEM monitors should be calibrated using transfer standards that are certified and NIST traceable. Each gas is a potential interferent that we recommend exploring during enhanced testing (see Section 2.2). Simultaneously measuring these pollutants during base testing may aid in interpreting the field testing results and in verifying the enhanced testing results.

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 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. FRM/FEM monitors typically operate every hour of every day except during periods of maintenance. This should also be considered when time aligning data.

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 site(s).

Table 2-1. Test Site Selection Criteria

Base Testing Plan	Location(s)	Season	Goal 1-Hour Average O_3 Concentration (for at least one day)
Single test site	Site 1	O_3 season (May to August*)	≥ 60 ppbv
	Site 1	Year-round†	None
Two test sites	Site 1	O_3 season (May to August*)	≥ 60 ppbv
	Site 2	Year-round†	None

*Typical O_3 season, but may be longer in some locations
†Not limited to the typical O_3 season

As shown in Table 2-1, it is recommended that base testing is conducted in two (2) locations, or alternatively, at one (1) site but during two different seasons. Testing under a range of O_3 concentrations is preferred, therefore one test should be conducted during the O_3 season (typically May through August) in a location likely to encounter 1-hour average O_3 concentrations greater than or equal to 60 ppbv for at least one day (as a goal) of the 30-day testing period. Since most elevated O_3 concentrations are likely to happen during a summer O_3 season, the range of T conditions encountered may be limited. Thus, a second test during a different season or at another site may offer performance information over a range of meteorological conditions and/or where other potentially interferent co-pollutants may be present. As an additional consideration, testers may wish to perform testing in locations which may experience wintertime O_3 episodes (Lyman and Tran, 2015; Ahmadov et al, 2015; Rappenglück et al, 2013; Schnell, et al. 2009).

Based on historical data, there are a number of sites across the U.S. that should offer the conditions shown in Table 2-1. 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)], historical air quality data can be found on the U.S. EPA AirData website (https://www.epa.gov/outdoor-air-quality-data, *last accessed 07/06/2020*). 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.

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) O_3 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 setup, 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 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 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., vehicle exhaust) • 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 outlined in 40 CFR Parts 50, 53, and Appendix A of Part 58.
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 quality control (QC) check on the FRM/FEM monitor (specified in 40 CFR Part 58, Appendix A, Section 3.1.1), and record the date of the QC 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 QC check).
9. Record and report the rationale for missing or invalidated data. For a full 30 consecutive days run, at least 75% uptime, with all instruments reporting is ideal. This corresponds to all

equipment reporting at least 540 valid 1-hour time-matched data points over the course of the 30-day deployment (720 hours total).

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. A preliminary report could present results with documentation as to why the sensors failed as these details may be useful to 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 540 valid 1-hour pairs of time-matched data points are collected.
c. If data from any piece of equipment is not available during each 1-hour sampling period, record and report the reason (e.g., outage, maintenance).
d. If any 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 effect of interferents, 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. An overview of enhanced testing procedures is shown in Figure 2-2. The procedure in this section outlines the materials and equipment needed, 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 enhanced testing procedure is documented, a checklist is provided in Appendix F. All information for this testing procedure should be recorded in the enhanced testing report (Appendix G). As mentioned previously, it is recommended that testers make the testing report(s) available on their respective websites to inform consumers.

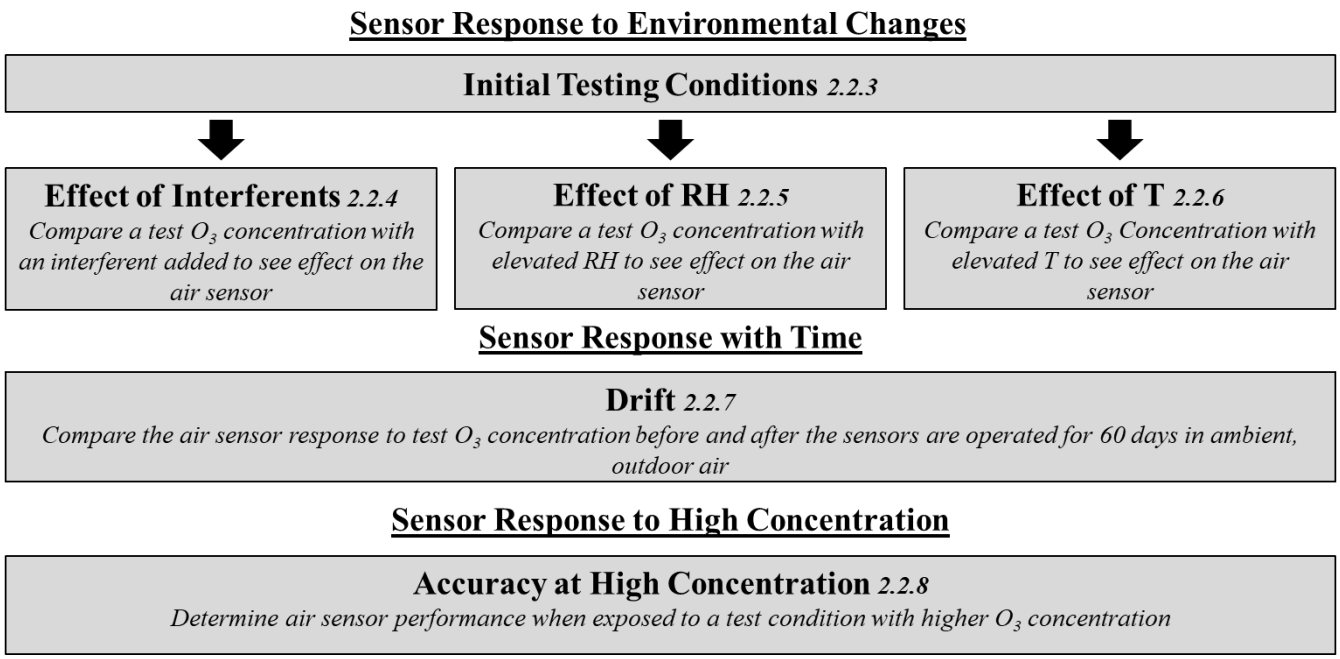


Figure 2-2. Overview of the Enhanced Testing Procedure

2.2.1 Materials and Equipment

The following materials are needed for enhanced testing:

- Three (3) or more O_3 air sensors having the same make, model, and firmware version*
- Calibrated O_3 FRM/FEM monitor†
- Exposure chamber that can control environmental conditions
- O_3 generator‡
- Carbon monoxide (CO) calibration cylinder‡
- Nitrogen dioxide (NO_2) calibration cylinder‡
- Sulfur dioxide (SO_2) calibration cylinder‡
- Zero air generator‡
- Dynamic calibration system
- Calibrated RH monitor§
- Calibrated T monitor§
- Calibrated CO FRM/FEM monitor†
- Calibrated NO_2 FRM/FEM monitor†

- Calibrated SO_2 FRM/FEM monitor[†]

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

[†]The FRM/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 40 CFR Parts 50, 53, and Appendix A of Part 58. 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).

[‡]O_3 generator (if being used to calibrate the FRM/FEM monitor), calibration cylinder, and 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).

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 (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), 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 pollutant concentrations. Approximate recommended ranges based on testing conditions outlined in this report: T – 19 to 41°C; RH – 35 to 90%; O_3 – 10 to 140 ppbv; CO – 30 to 40 ppmv; NO_2 – 90 to 115 ppbv; SO_2 – 65 to 85 ppbv.

- 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 nonreactive or passivated tubing connecting the chamber to the FRM/FEM monitors should be short so as not to affect what is sampled.

- Contain nonreactive or passivated chamber walls.

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

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 the FRM/FEM, T, and RH monitors are properly calibrated. Record the calibration date for each piece of equipment (as applicable). The FRM/FEM monitors should be calibrated following the procedures outlined in 40 CFR Parts 50, 53, and Appendix A of Part 58.
2. Conduct a one-point QC check on all the FRM/FEM monitors (as specified in 40 CFR Part 58, Appendix A, Section 3.1.1), and record the date of the QC check.
3. Verify that there are at least three (3) O_3 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 FRM/FEMs should not be used during this testing procedure.
5. In the enhanced testing report (Appendix H), record information about the equipment and setup, 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 (as 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 a .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. Throughout testing, follow the manufacturer's maintenance recommendations, as applicable, for all equipment (e.g., sensors, FRM/FEM). Record and report all maintenance or troubleshooting performed, including dates/times, on the instruments (e.g., power cycling, FRM/FEM one-point QC check).

2.2.3 Initial Testing Conditions

Take the following steps to begin the enhanced testing:

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 FRM/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. O_3 concentration from each sensor (ppbv)
    b. FRM/FEM O_3 concentration (ppbv)
    c. RH (%)
    d. T (°C)

Table 2-4. Initial Testing Conditions

Parameter	Reference Setpoint
O_3 Concentration	70 ppbv ± 5%
T	20°C ± 1°C
RH	40% ± 5%

2.2.4 Effect of Interferents

To evaluate the effect of the presence of CO, NO_2, and SO_2 on sensor performance, take the following steps:

1. Repeat the procedure outlined in Section 2.2.3.
2. Supply a single interferent pollutant to the exposure chamber at the concentration level shown in Table 2-5 after steady state is achieved at the initial conditions shown in Table 2-4. As the total gas flow rate to the chamber changes, adjust the exhaust to prevent pressure buildup.
3. Allow the measurements to stabilize and the interferent pollutant concentration to reach the appropriate level within the tolerances listed in Table 2-5.
4. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FRM/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. O_3 concentration from each sensor (ppbv)
    b. FRM/FEM O_3 concentration (ppbv)
    c. Interferent concentration (ppbv or ppmv depending on the pollutant)
    d. RH (%)
    e. T (°C)
5. Flush the exposure chamber with zero air until the interferent pollutant concentration reads zero ppbv or ppmv (depending on the pollutant).
6. Repeat steps 1-5 in this section (Section 2.2.4) for each interferent pollutant shown in Table 2-5.

Table 2-5. Interferent Pollutant Test Concentrations

Interferent Pollutant*	Reference Setpoint
CO	35 ppmv ± 5%
NO_2	100 ppbv ± 5%
SO_2	75 ppbv ± 5%

*A manufacturer(s) and/or scientific literature may identify additional interferents that are not listed in Table 2-5. A tester may wish to conduct additional tests with mixtures of interferents. Testers should report results from any additional interferent tests. Note that additional materials may be needed to conduct these tests.

2.2.5 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 2-6.
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 FRM/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. O_3 concentration from each sensor (ppbv)
    b. FRM/FEM O_3 concentration (ppbv)
    c. RH (%)
    d. T (°C)

Table 2-6. Elevated RH Test Conditions

Parameter	Reference Setpoint
O_3 Concentration	70 ppbv ± 5%
T	20°C ± 1°C
RH	85% ± 5%

2.2.6 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 2-7.
3. Allow the measurements to stabilize within the tolerances shown in Table 2-7.
4. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FRM/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. O_3 concentration from each sensor (ppbv)
b. FRM/FEM O_3 concentration (ppbv)
c. RH (%)
d. T (°C)

Table 2-7. Elevated T Test Conditions

Parameter	Reference Setpoint
O_3 Concentration	70 ppbv ± 5%
T	40°C ± 1°C
RH	40% ± 5%

2.2.7 Drift

A summary of the entire drift testing procedure is shown in Figure 2-3.

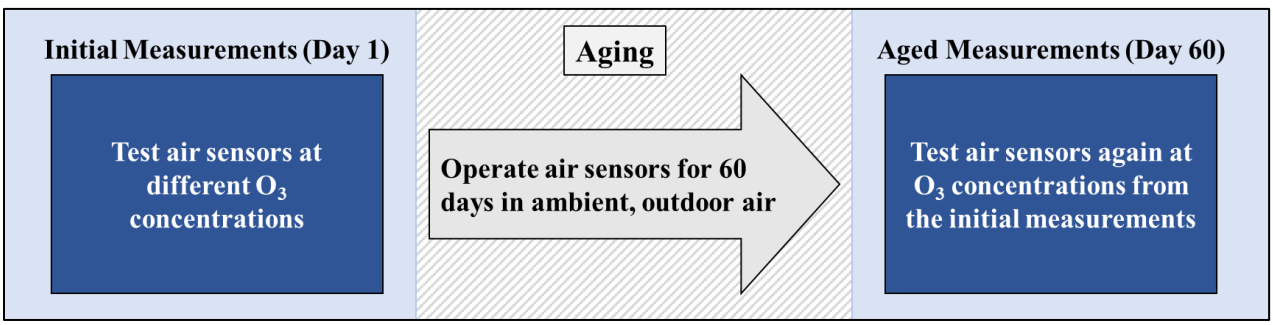


Figure 2-3. Drift Testing to Determine Changes After 60 days or More of Continuous Operation

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

To assess the drift, begin testing at a low and mid level O_3 concentration 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 2-8.
2. Allow the measurements to stabilize within the tolerances shown in Table 2-8.
3. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FRM/FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on the time resolution for data collection is provided in Appendix B.
    a. O_3 concentration from each sensor (ppbv)
    b. FRM/FEM O_3 concentration (ppbv)
    c. RH (%)
    d. T (°C)
4. Supply the exposure chamber with the conditions shown in Table 2-9.
5. Allow the measurements to stabilize within the tolerances shown in Table 2-9.

6. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FRM/FEM data points or three (3) consecutive hours for the parameters listed below. Additional information on the time resolution for data collection is provided in Appendix B.
    a. O_3 concentration from each sensor (ppbv)
    b. FRM/FEM O_3 concentration (ppbv)
    c. RH (%)
    d. T (°C)

Table 2-8. Low Concentration Drift Test Conditions

Parameter	Reference Setpoint
O_3 Concentration	15 ppbv ± 10%
T	20°C ± 1°C
RH	40% ± 5%

Table 2-9. Mid Concentration Drift Test Conditions

Parameter	Reference Setpoint
O_3 Concentration	70 ppbv ± 5%
T	20°C ± 1°C
RH	40% ± 5%

2.2.7.2 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 2.2.7.1 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 reduce burden on testers. 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.

2.2.8 Accuracy at High Concentration

To evaluate sensor accuracy at a high O_3 concentration, take the following steps:

1. Supply the exposure chamber with the conditions in Table 2-10.
2. Allow the measurements to stabilize within the tolerances shown in Table 2-10.
3. Once steady state is achieved, collect either a minimum of 20-30 pairs of time-matched sensor and FRM/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. O_3 concentration from each sensor (ppbv)
    b. FRM/FEM O_3 concentration (ppbv)
    c. RH (%)
    d. T (°C)

Table 2-10. High O_3 Concentration Test Conditions

Parameter	Reference Setpoint
O_3 Concentration	125 ppbv ± 5%
T	20°C ± 1°C
RH	40% ± 5%

3.0 Performance Metrics and Supporting Calculations for Evaluating O_3 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 O_3 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*);
- South Coast Air Quality Management District Air Quality Sensor Performance Evaluation Center (AQ-SPEC) sensor evaluations (http://www.aqmd.gov/aq-spec/evaluations/summary-gas, *last accessed 09/19/2020*; 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 including the European Union/European Committee for Standardization (EU/CEN; Gerboles 2018 and 2019) and the People's Republic of China Ministry of Environment and Ecology (MEE; Environmental Protection Department of Hebei Province, 2017).

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 is 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 O_3 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).
	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 Interferents	A measurement response due to any non-target pollutants that might skew or influence a sensor's response to the target pollutant.
	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 Concentration	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.

FRM/FEM monitors typically operate every hour of every day except during periods of maintenance. This should also be considered when time aligning data.

3.1.1 Hourly Averages

For base testing, performance metrics are calculated from hourly averaged data. Any FRM/FEM, sensor, RH and/or T data collected as sub-hourly time intervals will first need to be averaged up to hourly averages (Eq. 1). In calculating these averages, a 75% data completeness requirement for each 1-hour interval should be imposed. For example, an O_3 sensor recording concentration measurements every 15 minutes would require a minimum of three (3) valid measurements in order to calculate a valid 1-hour averaged concentration [i.e., (3/4) * 100% = 75%].

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

where:

x_{khj} = 1-hour averaged measurement *k* for hour *h* and instrument *j* (ppbv, °C, % RH)

n = number of instrument measurements per 1-hour period

c_{ij} = measurement from instrument *j* for time *i* of the 1-hour period (ppbv, °C, % RH)

As a reminder, x_{khj} is considered a valid 1-hour average if at least 75% of the expected data points over a 1-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 1-hour averaged data (Eq. 2) for each field test.

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

where:

$\overline{x_k}$= deployment averaged measurement *k* for a field test (ppbv, °C, % RH)

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

N = number of 1-hour periods during which all identical instruments are operating and returning valid averages over the duration of the field test

x_{hj} = valid 1-hour averaged measurement for hour *h* and instrument *j* (ppbv, °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 1-hour averaged measurements.

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

where:

SD = standard deviation of 1-hour averaged sensor O_3 concentration measurements (ppbv)

N = number of 1-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_{hj} = 1-hour averaged sensor O_3 concentration for hour *h* and sensor *j* (ppbv)

$\overline{x}_h$= 1-hour averaged sensor O_3 concentration for hour *h* from the three (3) sensors (ppbv)

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

where:

CV = coefficient of variation (%)

SD = standard deviation of 1-hour average sensor O_3 concentration measurements (ppbv)

$\overline{x}$ = deployment averaged sensor O_3 concentration for a field test (ppbv)

3.1.4 Bias and Linearity

A simple linear regression model can demonstrate the relationship between paired 1-hour averaged sensor and FRM/FEM O_3 measurements. Using a simple linear regression model (y = mx + b) with the sensor O_3 measurements as the dependent variable (y) and the FRM/FEM O_3 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 07/06/2020*), 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 O_3 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 1-hour averaged measurements, the sensor and FRM/FEM O_3 measurements 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.

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

where:

RMSE = root mean square error (ppbv)

N = number of 1-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_{hj} = valid 1-hour averaged sensor O_3 concentration for hour h and instrument j (ppbv)

R_h = valid 1-hour averaged FRM/FEM O_3 concentration for hour h (ppbv)

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.

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 O_3 sensor technologies (Williams et al., 2013; U.S. EPA 2014b and 2015; Lin et al., 2015; Pang et al., 2017; Borrego et al., 2018). 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 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.

- 1-hour averaged normalized sensor O_3 concentration vs. 1-hour averaged DP
- 1-hour averaged normalized sensor O_3 concentration vs. 1-hour averaged RH
- 1-hour averaged normalized sensor O_3 concentration vs. 1-hour averaged T
- 1-hour averaged concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged DP
- 1-hour averaged concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged RH

- 1-hour averaged concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged T
- 1-hour averaged absolute concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged DP
- 1-hour averaged absolute concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged RH
- 1-hour averaged absolute concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged T

Optional additional scatterplots if interferent data is also available:

- 1-hour averaged normalized sensor O_3 concentration vs. 1-hour averaged CO
- 1-hour averaged normalized sensor O_3 concentration vs. 1-hour averaged NO_2
- 1-hour averaged normalized sensor O_3 concentration vs. 1-hour averaged SO_2
- 1-hour averaged absolute concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged CO
- 1-hour averaged absolute concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged NO_2
- 1-hour averaged absolute concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged SO_2
- 1-hour averaged concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged CO
- 1-hour averaged concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged NO_2
- 1-hour averaged concentration difference between the sensor and FRM/FEM O_3 concentration vs. 1-hour averaged SO_2
- 1-hour averaged FRM/FEM O_3 concentration vs. 1-hour averaged NO_2
- 1-hour averaged FRM/FEM O_3 concentration vs. 1-hour averaged SO_2
- 1-hour averaged FRM/FEM O_3 concentration vs. 1-hour averaged CO

3.1.6.2 Normalized Concentration

Normalized 1-hour averaged sensor O_3 concentrations are derived by dividing the 1-hour averaged sensor O_3 concentration by the paired 1-hour averaged FRM/FEM O_3 concentration (Eq. 6). 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.

$$NormC_{hj} = \frac{x_{hj}}{R_h} \quad \text{Eq. 6}$$

where:

$NormC_{hj}$ = normalized 1-hour averaged sensor O_3 concentration for hour *h* and instrument *j* (unitless)

x_{hj} = valid 1-hour averaged sensor O_3 concentration for hour *h* and instrument *j* (ppbv)

R_h = valid 1-hour averaged FRM/FEM O_3 concentration for hour *h* (ppbv)

3.1.6.3 Concentration Difference and Absolute Concentration Difference

The 1-hour averaged concentration difference is derived by subtracting the 1-hour averaged FRM/FEM O_3 concentration from the 1-hour averaged sensor O_3 concentration (Eq. 7a).

$$\Delta C_{hj} = x_{hj} - R_h \quad \text{Eq. 7a}$$

where:

ΔC_{hj} = concentration difference between valid 1-hour averaged sensor and FRM/FEM O_3 concentration values for hour *h* and sensor *j* (ppbv)

x_{hj} = valid 1-hour averaged sensor O_3 concentration for hour *h* and instrument *j* (ppbv)

R_h = valid 1-hour averaged FRM/FEM O_3 concentration for hour *h* (ppbv)

The 1-hour averaged absolute concentration difference for sensor O_3 concentrations is derived by taking the absolute value of the difference between the 1-hour averaged sensor O_3 concentration and the 1-hour averaged FRM/FEM O_3 (Eq. 7b). Equations 7a and 7b assume only one FRM/FEM instrument will be running. If multiple FRM/FEM instruments are running, separate testing reports can be generated for each.

$$Abs\Delta C_{hj} = |x_{hj} - R_h| \quad \text{Eq. 7b}$$

where:

$Abs\Delta C_{hj}$ = absolute concentration difference between valid 1-hour averaged sensor and FRM/FEM O_3 concentration values for hour *h* and sensor *j* (ppbv)

x_{hj} = valid 1-hour averaged sensor O_3 concentration for hour *h* and instrument *j* (ppbv)

R_h = valid 1-hour averaged FRM/FEM O_3 concentration for hour *h* (ppbv)

3.1.6.4 Dew Point (DP)

The 1-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. 8). DP should not be calculated using on-board T and RH sensor measurements (if applicable), as these measurement may not accurately represent ambient T and RH conditions.

$$\mathrm{DP}_h = 243.04 \times \left[\frac{\ln\left(\frac{\mathrm{RH}_h}{100}\right) + \frac{(17.625 \times \mathrm{T}_h)}{(243.04 + \mathrm{T}_h)}}{17.625 - \ln\left(\frac{\mathrm{RH}_h}{100}\right) - \frac{(17.625 \times \mathrm{T}_h)}{(243.04) + \mathrm{T}_h)}} \right] \qquad \text{Eq. 8}$$

where:

DP_h = valid 1-hour averaged ambient DP for hour *h* (°C)

RH_h = valid 1-hour averaged ambient RH for hour *h* (%)

T_h = valid 1-hour averaged ambient T for hour *h* (°C)

3.2 Enhanced 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 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

For enhanced testing, the time interval to which all data must be averaged may be variable depending on the FRM/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 FRM/FEM is recorded at a 10-minute time resolution, all data should be averaged to the 10-minute time resolution. In Equation 9 (Eq. 9), 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, an O_3 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%).

$$x_{ktj} = \frac{1}{n}\sum_{i=1}^{n} c_{ij} \qquad \text{Eq. 9}$$

where:

x_{ktj} = averaged measurement *k* for time interval *t* and instrument *j* (ppbv, ppmv, °C, % RH)

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

c_{ij} = measurement from instrument *j* for time *i* of time interval *t* (ppbv, ppmv, °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. 10) collected during the steady state period for each test.

$$\overline{x_k} = \frac{1}{M}\sum_{j=1}^{M}\left[\frac{1}{N}\sum_{h=1}^{N} x_{ktj}\right] \qquad \text{Eq. 10}$$

where:

$\overline{x_k}$ = test averaged measurement *k* for the chamber test (ppbv, ppmv, °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* (ppbv, ppmv, °C, % RH)

3.2.3 Precision

Precision between identical sensors can be characterized by two metrics: standard deviation (SD) between measurements (Eq. 11) and coefficient of variation (CV; Eq. 12). 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 2.2.7).

$$\text{SD} = \sqrt{\frac{1}{(N \times M) - 1}\sum_{j=1}^{M}\left[\sum_{t=1}^{N}(x_{tj} - \overline{x}_t)^2\right]} \qquad \text{Eq. 11}$$

where:

SD = standard deviation of test averaged sensor O_3 concentration measurements (ppbv)

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 O_3 concentration for time interval *t* and sensor *j* (ppbv)

$\overline{x}_t$ = test averaged sensor O_3 concentration for time interval *t* (ppbv)

$$CV_{Enhanced} = \frac{\text{SD}}{\overline{x}} \times 100 \qquad \text{Eq. 12}$$

where:

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

SD = standard deviation of test averaged sensor O_3 concentration measurements (ppbv)

$\overline{x}$ = test averaged sensor O_3 concentration for the chamber test (ppbv)

3.2.4 Bias and Linearity

A simple linear regression model can demonstrate the relationship between paired averaged sensor and FRM/FEM O_3 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 2.2.7) and the high concentration test (Section 2.2.8) will reflect data collected under similar T and RH conditions. Using a simple linear regression model (y = mx + b) with the sensor O_3 measurements as the dependent variable (y) and the FRM/FEM O_3 measurements as the independent variable (x), calculate the slope (m), intercept (b), and the coefficient of determination (R^2).

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 07/06/2020*), 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.

3.2.5 Error

The root mean square error (RMSE) is one metric that can be used to help understand the error associated with sensor O_3 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 2.2.7). Using data during which all sensors are reporting valid time averaged measurements, the sensor and FRM/FEM O_3 measurement calculations are compared (Eq. 13). 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.

$$\text{RMSE} = \sqrt{\frac{1}{N \times M} \sum_{j=1}^{M} \left[\sum_{t=1}^{N} (x_{tj} - R_t)^2 \right]} \qquad \text{Eq. 13}$$

where:

RMSE = root mean square error (ppbv)

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 O_3 concentration for time interval *t* and instrument *j* (ppbv)

R_t = averaged FRM/FEM O_3 concentration for time *t* (ppbv)

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 FRM/FEM instrument measurements or true values (see Eq. 13). 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.

3.2.6 Effect of Interferents

As described in Section 2.2.4, the interferent tests involve two steps: 1) collecting data during steady state at a prescribe O_3 concentration, and 2) collecting data during steady state when the prescribed concentrations of O_3 and the prescribed interferent is present. The effect of each interferent is the difference between these two measurements (Eq. 14).

$$\overline{x_{int}} = \overline{x_{(O_3 + int)}} - \overline{x_{O_3}} \qquad \text{Eq. 14}$$

where:

$\overline{x_{int}}$ = test averaged influence of the interferent on sensor measurements (ppmv or ppbv, dependent upon interferent)

$\overline{x_{(O_3 + int)}}$ = test averaged sensor O_3 concentration for the portion of the chamber test when both O_3 and the interferent are present (ppmv or ppbv, dependent upon interferent)

$\overline{x_{O_3}}$ = test averaged sensor O_3 concentration for the portion of the chamber test when only O_3 is present (ppmv or ppbv, dependent upon interferent)

3.2.7 Effect of Relative Humidity (RH)

As described in Section 2.2.5, the RH tests on sensor measurements involve two steps: 1) collecting data during steady state at a prescribed O_3 concentration at 40% RH, and 2) collecting data during steady state at the same prescribed O_3 concentration at 85% RH. The effect of RH is the difference between these two measurements (Eq. 15).

$$\overline{x_{RH}} = \overline{x_{(RH=85\%)}} - \overline{x_{(RH=40\%)}} \qquad \text{Eq. 15}$$

where:

$\overline{x_{RH}}$ = test averaged influence of RH on sensor measurements (ppbv)

$\overline{x_{(RH=85\%)}}$ = test averaged sensor O_3 concentration for the portion of the chamber test when the RH is 85% (ppbv)

$\overline{x_{(RH=40\%)}}$ = test averaged sensor O_3 concentration for the portion of the chamber test when the RH is 40% (ppbv)

3.2.8 Effect of Temperature (T)

As described in Section 2.2.6, the T tests on sensor measurements involve two steps: 1) collecting data during steady state at a prescribed O_3 concentration at 20°C, and 2) collecting data during steady state at the same prescribed O_3 concentration at 40°C. The effect of T is the difference between these two measurements (Eq. 16).

$$\overline{x_T} = \overline{x_{(T=40)}} - \overline{x_{(T=20)}} \qquad \text{Eq. 16}$$

where:

$\overline{x_T}$ = test averaged influence of T on sensor measurements (ppbv)

$\overline{x_{(T=40)}}$ = test averaged sensor O_3 concentration for the portion of the chamber test when the T is 40°C (ppbv)

$\overline{x_{(T=20)}}$ = test averaged sensor O_3 concentration for the portion of the chamber test when the T is 20°C (ppbv)

3.2.9 Drift

As described in Section 2.2.7, the drift tests involve measuring the drift at two O_3 concentrations: 1) at a low concentration of 15 ppbv, and 2) at a mid concentration of 70 ppbv which is relevant for health messaging. For each O_3 concentration, the drift measurement includes two separate chamber tests. The first will be conducted to determine the steady state concentration for the prescribed O_3 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 O_3 concentrations by the difference in the measurement over the 60-day period (Eq. 17).

$$\overline{x_{C_{drift}}} = \overline{x_{C(day=60)}} - \overline{x_{C(day=1)}} \qquad \text{Eq. 17}$$

where:

$\overline{x_{C_{drift}}}$ = test averaged sensor drift at O_3 concentration *C* over the course of 60 days (ppbv)

$\overline{x_{C(day=60)}}$ = test averaged sensor O_3 concentration at O_3 concentration *C* after 60 days of operation following the start of the drift test (ppbv)

$\overline{x_{C(day=1)}}$ = test averaged sensor O_3 concentration at O_3 concentration *C* at the beginning of the drift test (ppbv)

3.2.10 Accuracy at High Concentration

As described in Section 2.2.8, the high concentration accuracy test involves testing the sensor response at a high O_3 concentration which is relevant for health messaging. The accuracy of the sensor measurement will be determined by the difference between the sensor and FRM/FEM measurements (Eq. 18).

$$\overline{x_{\Delta}} = \overline{x_{sensor}} - \overline{x_{ref}} \qquad \text{Eq. 18}$$

where:

$\overline{x_{\Delta}}$ = test averaged difference between the sensor and FRM/FEM O_3 concentrations (ppbv)

$\overline{x_{sensor}}$ = test averaged sensor O_3 concentration (ppbv)

$\overline{x_{ref}}$ = test averaged FRM/FEM O_3 concentration (ppbv)

4.0 Target Values for O_3 Air Sensors

4.1 Approach

To inform the development of the target values for the performance metrics for O_3 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-the-box sensor performance before additional corrections were made by the user.

Table 4-1. O_3 Sensor Performance Field Evaluation Results from Available Resources

Performance Metric		Range	Average	Median
Precision	SD (ppbv)	no data	no data	no data
Bias	Slope*	0.66 to 1.23	0.95	0.98
	Intercept* (ppbv)	-7.14 to 14.53	0.79	0.00
Linearity	$R^{2,\dagger}$	0.74 to 0.98	0.90	0.89
Error	RMSE (ppbv)	3.65 to 7.22	5.71	6.25

Note: Resources include AQ-SPEC sensor evaluations, the U.S. EPA sensor evaluations, and peer-reviewed literature. Table only includes 1-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, as available, related to air sensor standards/certification programs in development from the EU/CEN (Gerboles 2018 and 2019) and 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 key performance metrics and target values recommended for the base and enhanced testing protocols for O_3 air sensors used in ambient, outdoor, fixed site NSIM applications. The recommended target values reflect the current state-of-the-science as the range of observed performance (Table 4-1) demonstrates they should be possible to achieve. Encouraging development of sensors which meet these target values should help ensure that sensor data can be well characterized and

understood. Additional performance metrics and test conditions for enhanced testing protocols 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 (see Appendix D for more detailed discussion).

Table 4-2. Base and Enhanced Testing – Recommended Performance Metrics and Target Values for O_3 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 ppbv	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 ± 0.2		3.1.4 and 3.2.4
	Intercept (b)	-5 ≤ b ≤ 5 ppbv		3.1.4 and 3.2.4
Linearity	Coefficient of Determination (R^2)	≥ 0.80		3.1.4 and 3.2.4
Error	Root Mean Square Error (RMSE)	≤ 5 ppbv		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 results due to the limited amount of data that will be collected. See Appendix D for further discussion.

Table 4-3. Enhanced Testing – Additional Recommended Performance Metrics and Test Conditions for O_3 Air Sensors Used in Ambient, Outdoor, Fixed Site NSIM Applications

Performance Metric	Test Conditions	Associated Section Describing Calculation
Effect of Interferents	Carbon monoxide (CO): 35 ppmv ± 5%	3.2.6
	Nitrogen dioxide (NO_2): 100 ppbv ± 5%	3.2.6
	Sulfur dioxide (SO_2): 75 ppbv ± 5%	3.2.6
Effect of Relative Humidity (RH)	Moderate RH: 40% ± 5%	3.2.7
	Elevated RH: 85% ± 5%	3.2.7
Effect of Temperature (T)	Moderate T: 20°C ± 1°C	3.2.8
	Elevated T: 40°C ± 1°C	3.2.8
Drift	Low concentration: 15 ppbv ± 10%	3.2.9
	Mid concentration: 70 ppbv ± 5%	3.2.9
Accuracy at High Concentration	High concentration: 125 ppbv ± 5%	3.2.10

It is recognized that the information in this report is based on the current knowledge of O_3 air sensors at the time this report was released and that O_3 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 O_3 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. Ahmadov, R., McKeen, S., Trainer, M., Banta, R., Brewer, A., Brown, S., Edwards, P.M., et al. Understanding high wintertime ozone pollution events in an oil- and natural gas-producing region of the western U.S. Atmospheric Chemistry and Physics, 15, 411-429, 2015.
2. Bart, M., Williams, D.E., Ainslie, B., McKendry, I., Salmond, J., Grange, S.K., Alavi-Shoshtari, M., Steyn, D., Henshaw, G.S. High density ozone monitoring using gas sensitive semi-conductor sensors in the lower Fraser Valley, British Columbia. Environmental Science & Technology, 48, 3970-3977, 2014.
3. Borrego, C., Ginja, J., Coutinho, M., Ribeiro, C., Karatzas, K., Sioumis, T., Natsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Salvato, M., et al. Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise – Part II. Atmospheric Environment, 193, 127-142, https://doi.org/10.1016/j.atmosenv.2018.08.028, 2018.
4. Borrego, C., Costa, A.M., Ginga, J., Amorim, M., Coutinho, M., Karatzas, K., Sioumis, Th., et al. Assessment of air quality microsensors versus reference methods: the EuNetAir joint exercise. Atmospheric Environment, 193, 246-263, https://doi.org/10.1016/j.atmosenv.2016.09.050, 2016.
5. Collier-Oxandale, A., Coffey, E., Thorson, J., Johnston, J., Hannigan, M. Comparing building and neighborhood-scale variability of CO_2 and O_3 to inform deployment considerations for low-cost sensor system use. Sensors, 18, 1349, doi:10.3390/s18051349, 2018.
6. Duvall, R.M., Long, R.W., Beaver, M.R., Kronmiller, K.G., Wheeler, M.W., Szykman, J.J. Performance evaluation and community application of low-cost sensors for ozone and nitrogen dioxide. Sensors, 16, 1698, doi:10.3390/s16101698, 2016.
7. Environmental Protection Department of Hebei Province. Technical Regulation for Selecting the Location of Air Pollution Control Gridded Monitoring System, DB13/T 2545-2017, 2017.
8. Feinberg, S., Williams, R., Hagler, G.S.W., Rickard, J., Brown, R., Garver, D., Harshfield, G., Stauffer, P., Mattson, E., Judge, R., Garvey, S. Long-term evaluation of air sensor technology under ambient conditions in Denver, Colorado. Atmospheric Measurement Techniques, 11, 4605-4615, https://doi.org/10.5194/amt-11-4605-2018, 2018.
9. Gerboles, M. European Union perspective on sensor standards – What factors were used to develop EN standards? [PowerPoint slides], 2019. Retrieved from: https://www.epa.gov/sites/production/files/2019-08/documents/gerboles_european_union_perspective_on_sensor_standards.pdf.
10. Gerboles, M. The road to developing performance standards for low cost sensors in Europe [PowerPoint slides], 2018. Retrieved from: https://www.epa.gov/sites/production/files/2020-02/documents/session_01_b_gerboles.pdf.
11. Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R. Caudill, M., et al. Community Air Sensor Network (CAIRSENSE) project: Evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. Atmospheric Measurement Techniques, 9, 5281-5292, https://doi.org/10.5194/amt-9-5281-2016, 2016.
12. Lin, C., Gillespie, J., Schuder, M.D., Duberstein, W., Beverland, I.J., Heal, M.R. Evaluation and calibration of Aeroqual series 500 portable gas sensors for accurate measurements of ambient ozone and nitrogen dioxide. Atmospheric Environment, 100, 111-116, http://dx.doi.org/10.1016/j.atmosenv.2014.11.002, 2015.
13. Lyman, S., Tran, T. Inversion structure and winter ozone distribution in the Uintah Basin, Utah, U.S.A. Atmospheric Environment, 123, 156-165, 2015.

14. Masiol, M., Squizzato, S., Chalupa, D., Rich, D.Q, Hopke, P.K. Evaluation and field calibration of a low-cost ozone monitor at a regulatory urban monitoring station. Aerosol and Air Quality Research, 18, 2029-2037, https://doi.org/10.4209/aaqr.2018.02.0056, 2018.
15. Masey, N., Gillespie, J., Ezani, E., Lin, C., Wu, H., Ferguson, N.S., Hamilton, S., Heal, M.R., Beverland, I.J. Temporal changes in field calibration relationships for Aeroqual S500 O_3 and NO_2 sensor-based monitors. Sensors and Actuators B: Chemical, 273, 1800-1806, https://doi.org/10.1016/j.snb.2018.07.087, 2018.
16. Pang, X., Shaw, M.D., Lewis, A.C., Carpenter, L.J., Batchellier, T. Electrochemical ozone sensors: A miniaturised alternative for ozone measurements in laboratory experiments and air-quality monitoring. Sensors and Actuators B: Chemical, 240, 829-837, https://doi.org/10.1016/j.snb.2016.09.020, 2017.
17. 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.
18. Rappenglück, B., Ackermann, L., Alvarez, S., Golovko, J., Buhr, M., Field, R., et al. Strong wintertime ozone events in Upper Green River Basin, Wyoming. Atmospheric Chemistry and Physics Discussions, 3, 17953-18005, 2013.
19. Sadighi, K., Coffey, E., Polidori, A., Feenstra, B., Lv, Q., Henze, D.K., Hannigan, M. Intra-urban spatial variability of surface ozone in Riverside, CA: viability and validation of low-cost sensors. Atmospheric Measurement Techniques, 11, 1777-1792, https://doi.org/10.5194/amt-11-1777-2018, 2018.
20. Schneider, P., Castell, N., Vogt, M., Dauge, F.R., Lahoz, W.A., Bartonova, A. Mapping urban air quality in near real-time using observations from low-cost sensors and model information. Environment International, 106, 234-247, https://doi.org/10.1016/j.envint.2017.05.005, 2017.
21. Schneider, P., Bartonova, A., Castell, N., Dauge, F.R., Gerboles, M., Hagler, G.S.W., Hüglin, C., Jones, R.L., Khan, S., Lewis, A.C., Mijling, B., Müller, M., Penza, M., Spinelle, L., Stacey, B., Vogt, M., Wesseling, J., Williams, R.W. 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.
22. Schnell, R.C., Oltmans, S.J., Neely, R.R., Endres, M.S., Molenar, J.V., White, A.B. Rapid photochemical production of ozone at high concentrations in a rural site during winter. Nature Geoscience, 2, 120-122, 2009.
23. Sonoma Technology, Inc. (STI), Ozone Concentrations In and Around the City of Arvin, California, Final Report prepared for San Joaquin Valley Unified Air Pollution Control District, Fresno, CA, STI-913040-5865-FR2, 2014. Available at: http://www.valleyair.org/air_quality_plans/docs/2013attainment/ozonesaturationstudy.pdf.
24. South Coast Air Quality Management District (SCAQMD), Air Quality Sensor Performance Evaluation Center (AQ-SPEC). Gas-phase sensor evaluations, available at: http://www.aqmd.gov/aq-spec/evaluations/summary-gas.
25. 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.
26. U.S. EPA. Air Sensor Guidebook. U.S. Environmental Protection Agency, Research Triangle Park, NC, EPA 600/R-14/159, 2014a.
27. U.S. EPA. Sensor Evaluation Report. U.S. Environmental Protection, Agency, Research Triangle Park, NC, EPA 600/R-14/143, 2014b.

28. 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
29. 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.
30. 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.
31. 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.
32. U.S. EPA, 40 CFR Part 50 – National Primary and Secondary Ambient Air Quality Standards.
33. U.S. EPA, 40 CFR Part 53 – Ambient Air Monitoring Reference and Equivalent Methods.
34. U.S. EPA 40 CFR Part 58 – Ambient Air Quality Surveillance.
35. University of British Columbia (UBC). Review of Next Generation Air Monitors for Air Pollution. Report prepared for Environment Canada, Gatineau, QC, 2014. Available at: https://open.library.ubc.ca/cIRcle/collections/facultyresearchandpublications/52383/items/1.0132725.
36. Williams, D.E., Henshaw, G.S., Bart, M., Laing, G., Wagner, J., Naisbitt, S., Salmond, J.A. Validation of low-cost ozone measurements instruments suitable for use in an air-quality monitoring network. Measurement Science and Technology, 24, doi:10.1088/0957-0233/24/6/065803, 2013.
37. 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.

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 sources (e.g., AQ-SPEC, EU/CEN, 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 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 T and not by changes in the concentration of the target pollutant.

Effect of Interferents: A measurement response due to any non-target pollutants that might skew or influence a sensor's response to 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 07/06/2020.*

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.*

Interferent: Any non-target pollutants that might skew or influence a sensor's response to the target pollutant.

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): A 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 of the random disagreement between the measurements reported by the sensor and the reference instrument. 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.

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 (696 hours) of a 30-day base test field deployment (720 hours expected), the uptime for the deployment can be expressed as 96.7% (i.e., 696 hours/720 hours * 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. However, testers 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 O_3 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, 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. Consistent with most current evaluation efforts, one FRM/FEM monitor is the minimum recommended for comparison. Testing recommends comparison of 1-hour averaged data to reflect the shorter response times of O_3 air sensors and FRM/FEM monitors. Consistent 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. 40 CFR Part 53.35 also requires field test deployments at multiple sites or during multiple seasons. Discussion with experts and review of current practices determined that two (2) field deployments is likely sufficient to show sensor performance over a range of conditions including temperature (T), relative humidity (RH), weather, O_3 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.

Enhanced testing protocols were modeled after laboratory evaluations conducted by a variety of organizations seeking to quantify the effect of RH, T, and interferents; drift; and accuracy at high concentrations. Other tests, specifically the detection limit, were considered but ultimately not included at this time due to limited feasibility and inconsistency in results. Testing protocols outlined in this report specify initial conditions of 20°C and 40% RH to maintain consistency with other laboratory sensor performance evaluations. Mid and high O_3 concentration setpoints were determined based on breakpoints in the Air Quality Index (AQI) where sustained high concentration measurements would be important for health messaging.

Testing duration and data time averaging during the enhanced tests can vary dependent on the equipment used for testing. The enhanced testing protocols describe a test duration as the time needed to collect either a minimum of 20-30 pairs of time-matched sensor and FRM/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 marker, handle the time resolution available on current FRM/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 FRM/FEM monitors that may be used for this work often report at 1-minute, 10-minute, or 1-hour averages. A pair of high time resolution instruments (sensor and FRM/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 FRM/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 Interferents for current electrochemical O_3 sensors include NO_2, SO_2, and CO. Interferences from these pollutants were identified from peer-reviewed literature and manufacturer specification sheets. These testing protocols were simplified from previous U.S. EPA sensor performance evaluations.

Effect of Relative Humidity (RH) testing protocols are based on the AQ-SPEC and previous U.S. EPA laboratory evaluations. Two RH set points (40% and 85%) are recommended to simplify testing as lower and higher setpoints may be difficult to achieve in some laboratory exposure chambers. An RH of 40% commonly represents average RH conditions. For the protocols in this report, an elevated RH of 85% was selected as it is consistent with previous U.S. EPA evaluations and is important to better characterize sensor performance in areas like the Southeast U.S. that can experience high RH levels.

Effect of Temperature (T) testing protocols are based on the AQ-SPEC and previous U.S. EPA laboratory evaluations. Two T set points (20°C and 40°C) are compared. A T of 20°C commonly represents average T conditions. The elevated T condition of 40°C is important to better characterize sensor performance during summer when O_3 concentrations are typically higher and in areas like the Southwest U.S. that can experience 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. In addition, currently some sensors do not provide true measurements for zero air making those measurements more difficult to interpret in context of DL. Sensor evaluation programs typically determine DL from noise levels within the data at low concentrations or during exposure to zero air. Most manufacturers typically list the range of values a device can measure as being zero to some positive value (e.g., 0 – 50 ppbv). 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 and O_3 sensor is challenged with zero air followed by small step changes in O_3 concentrations to determine a 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 O_3 concentrations more quickly than a test chamber can obtain equilibrium and/or an FRM/FEM instrument can confirm it. It would be advantageous for this test, as well as for rise/lag testing, if O_3 sensors were designed with a remote activitation 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 loosely based on the U.S. EPA FRM/FEM testing procedures outlined in 40 CFR Section 53.23(e) for 24-hour drift. During the low concentration drift test in this testing protocol, a test gas with low O_3 concentration (15 ppbv), rather than zero air, was used because some sensors do not provide true measurements for zero air. The mid concentration drift test setpoint was determined based on the green/yellow (good/moderate) breakpoint in the AQI where sustained concentration measurements would be important for health messaging. A 60-day testing, or aging, period was recommended as opposed to a short-term 24-hour test period as it was 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 Concentration testing protocols are loosely based on the FRM/FEM testing procedures outlined in 40 CFR Sections 53.21(b) and 53.23(a), which prescribe multi-point calibration curves. The test in this report adds a third calibration point (in addition to a low and mid concentration) and prescribes that the high concentration testing be performed last, as some literature indicates that exposures to high concentrations can accelerate sensor aging and reduce sensor response, both of which can damage the sensor. The high O_3 concentration setpoint was determined based on the yellow/orange (moderate/unhealthy for sensitive groups) breakpoint in the AQI where sustained high concentration measurements would be important for health messaging.

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 protocols. Some of the metrics recommended in this report are not included in the FRM/FEM 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 sensors devices. Many studies use precision, or a metric called repeatability (used by EU/CEN and the People's Republic of China MEE), to describe this agreement. 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.

Bias within O_3 air sensors was not explicitly discussed by other organizations. However, performing a linear regression to determine slope and intercept was a standard procedure in sensor evaluations and literature. This metric 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 literature. Further, simple linear regression is simpler and easier to communicate 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) be calculated. 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.

Exploring Meteorological Effects with respect to temperature (T) and relative humidity (RH) are common exploration 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 data from reference instruments. Understanding this dependence can be important for some NSIM applications or sensor environments.

Additional Enhanced Testing Performance Metrics

Interferents can potentially skew air sensor readings. It is important to understand which pollutants may influence O_3 air sensors and the degree to which they skew the data. For example, if an O_3 air sensor is extremely sensitive to NO_2 (a common interferent for O_3 sensors), then the O_3 air sensor readings might respond to fluctuations in NO_2 concentrations rather than the O_3 concentrations.

Effect of Relative Humidity (RH) is a thoughtfully designed experiment to determine the degree to which RH introduces a positive or negative bias to sensor measurements. For electrochemical O_3 sensors, the sensing element can be affected by the presence of water vapor which can coat the sensor surface or change the concentration of an electrolyte solution. These reactions may create deviations in the sensor response to the target pollutant, especially under high RH conditions. 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.

Effect of Temperature (T) is a thoughtfully designed experiment to determine the degree to which 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.

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 Concentration is important to in order to evaluate the suitability of a sensor for NSIM applications where high O_3 concentrations are expected. Additionally, this test is used to determine the linearity of sensor response relative to reference measurements across a wide range of concentrations.

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
- Comparison to other organizations developing sensor standards/certification programs (Table D-5).

Information from these resources listed are summarized as indicated in Tables D-1 to D-5.

Table D-1. Performance Requirements for O_3 FRM/FEM Regulatory Monitors (adapted from U.S. EPA, 2018)

Performance Attribute	Specification for Regulatory Monitoring	Notes (Based on 40 CFR Part 53 Subpart B)
Accuracy/Uncertainty	± 4 ppbv	Denoted by zero drift, 12 hour and 24 hour.
Measurement Range	0 – 500 ppbv	Referred to as 'range'.
Detection Limit	Detection Limit: 5 ppbv	Referred to as 'lower detection limit'.
	Noise: 2.5 ppbv	Defined as spontaneous, short duration deviations in measurements or measurement signal output, about the mean output, that are not caused by input concentration changes.
Selectivity	± 5 ppbv	Referred to as "Interference Equivalent".
Response Time	Lag time: 2 minutes	Defined as the time interval between a step change in input concentration and the first observable corresponding change in measurement response.
	Rise Time: 2 minutes	Defined as the time interval between initial measurement response and 95% percent of the final response after a step increase in input concentration.
	Fall Time: 2 minutes	Defined as the time interval between initial measurement response and 95% percent of the final response after a step decrease in input concentration.
Precision	20% of Upper Range Limit: 2%	Denoted as the standard deviation expressed as percent of the upper range limit.
	80% of Upper Range Limit: 2%	

Table D-2. Summary of U.S. EPA O_3 Sensor Evaluation Results

Source	Field or Lab	Sensor	Averaging Time	Concentration Range (ppbv)	Precision (SD)	Slope	Intercept (ppbv)	R^2
U.S. EPA, 2015	Field	PerkinElmer Elm	5-minute	0-60	N/A	0.875	6.2	0.73
U.S. EPA, 2014b	Lab	AGT Environmental Sensor	1-minute	0-400	2.6-10.3	N/A	N/A	0.97-0.99
	Lab	CairClip O3/NO2	1-minute	0-200	2.8-9.5	N/A	N/A	0.99
	Lab	Dynamo	1-minute	0-450	3.3-7.0	N/A	N/A	0.94-0.99
	Lab	U-POD	5-second	not specified	6.5-46.2	N/A	N/A	0.87-0.95

Table D-3. Summary of AQ-SPEC O_3 Sensor Field Evaluation Results

Sensor Manufacturer/Model	Averaging Time	Concentration Range (ppbv)	Slope*	Intercept* (ppbv)	R^2
Aeroqual/AQY v0.5	1-hour	0-60	0.66 to 0.81	0.89 to 14.53	0.96-0.98
Aeroqual/AQY v1.0	1-hour	0-60	1.02 to 1.23	-0.11 to 0.90	0.98
Aeroqual/model OZU S500	1-hour	0-75	0.98 to 1.00	0.00	0.86-0.88
AQMesh/v4.0	15-minute	0-120	1.16 to 1.51	-17.10 to -12.32	0.46-0.83
AQMesh/v5.1 Gas Unit	1-hour	5-200	1.80 to 2.23	-38.90 to -5.75	0.89-0.91
APIS	1-hour	0-100	2.16 to 2.75	-62.80 to -25.29	0.74-0.84
Kunak/Air A10	1-hour	0-100	0.94 to 1.10	-7.14 to -0.85	0.87-0.89
Magnasci SRL/uRADMonitor INDUSTRIAL HW103	1-hour	0-60	-19.61 to 4.84	-83.66 to 406.53	0.00-0.08
Perkin Elmer/ELM	5-minute	0-140	0.60 to 0.91	0.42 to 16.95	0.89-0.96
Spec Sensor	5-minute	0-40	5.12 to 25.25	-363.06 to -29.62	0.01-0.24
uHoo Sensor/old firmware	5-minute	0-150	0.20 to 0.37	-0.30 to 0.82	0.54-0.65
uHoo Sensor/new firmware	5-minute	0-100	1.56 to 4.85	65.50 to -9.38	0.43-0.72
Vaisala/AQT410 v.1.15	5-minute	0-80	1.01 to 1.33	-13.15 to -2.60	0.67-0.81
Vaisala/AQT410 v.1.11	5-minute	0-100	-5.39 to -3.62	0.27 to 0.45	0.40-0.58
Wicked Devices/Air Quality Egg v2	5-minute	0-50	-12.72 to 7.98	372.95 to 398.77	0.14-0.17

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 used to Inform Target Values

Source	Location	Sensor Manufacturer/ Model	Averaging Time	Concentration Range (ppbv)	Slope	Intercept (ppbv)	R^2	RMSE (ppbv)
STI, 2014	Arvin, CA	Aeroqual/Series 500	1-hour	0-100	1.001 to 1.051	-3.28 to -0.015	0.92-0.98	N/A
Feinberg et al., 2018	Denver, CO	Aeroqual/SM-50	1-hour	N/A	0.56 to 0.77	-0.0004	0.85-0.92*	N/A
		Cairpol/CairClip	1-hour	N/A	-0.04 to 1.03	-23.6 to -39.0	0.00-0.21*	N/A
Masiol et al., 2018	Rochester, NY		1-hour	0-90	0.81	8.64	0.87	N/A
Jiao et al., 2016	Atlanta, GA	Aeroqual/SM-50	1-hour	0-90	0.001*	-0.003*	0.82-0.94	N/A
		Cairpol/CairClip	1-hour	0-90	1.17 to 1.46*	-14.6 to -17.8*	0.68-0.88	N/A
		AQMesh/Gen. 3	1-hour	0-70	N/A	N/A	≤ 0.25	N/A
Masey et al., 2018	Glasgow, United Kingdom	Aeroqual/Series 500	1-hour	0-100	0.56 to 1.02	6.23 to 12.7	0.91-0.93	6.25-7.22
Collier-Oxandale et al., 2018	Los Angeles, CA	University of Colorado Boulder/Y-Pods	1-hour	0-80	N/A	N/A	0.97	3.65
Sadighi et al., 2018	Riverside, CA	University of Colorado Boulder/U-Pod	1-minute	0-100	0.5	0.05	0.97-0.99	4.4-5.9
Borrego et al., 2016	Aveiro, Portugal	ENEA	1-hour	0-90	N/A	N/A	0.13	N/A
		NanoEnvi	1-hour	0-90	N/A	N/A	0.77	N/A
		CAM_11	1-hour	0-90	N/A	N/A	0.14	N/A
		AQMesh	1-hour	0-90	N/A	N/A	0.7	N/A
		ISAG	1-hour	0-90	N/A	N/A	0.12	N/A
Borrego et al., 2018	Aveiro, Portugal	AQMesh	15-minute	5-45	N/A	N/A	0.68	N/A
		AUTh	5-minute	5-45	N/A	N/A	0.11	N/A
		CAM_11	20-second	5-45	N/A	N/A	0.05	N/A
		ENEA	15-minute	5-45	N/A	N/A	0.25	N/A
		NanoEnvi	5-minute	5-45	N/A	N/A	0.05	N/A
Schneider et al., 2017	Oslo, Norway	AQMesh	1-hour	N/A	N/A	N/A	0.29	22.2
UBC, 2014	Leamington Spa, United Kingdom	AQMesh	15-minute	0-100	N/A	N/A	> 0.80	N/A

Source	Location	Sensor Manufacturer/ Model	Averaging Time	Concentration Range (ppbv)	Slope	Intercept (ppbv)	R^2	RMSE (ppbv)
	London, United Kingdom	AirBase/CanarIT	15-minute	0-30	N/A	N/A	0.81	N/A
Pang et al., 2017	University of York, United Kingdom	Alphasense/ OXB421	5-minute	0-65	1.09	N/A	0.84	N/A
Williams et al., 2013	Houston, TX	Aeroqual/Early S500	1-hour	0-140	0.99*	-5.35*	N/A	N/A
	Auckland, New Zealand	Aeroqual/Early S500	1-hour	N/A	N/A	N/A	N/A	N/A
	Rocky Mountain National Park, CO	Aeroqual/Early S500	1-hour	N/A	N/A	N/A	N/A	N/A
	Sequoia & King Canyon National Parks, CA	Aeroqual/Early S500	1-hour	0-120	1.08*	-4.19*	N/A	N/A
	Research Triangle Park, NC	Aeroqual/Early S500	1-hour	0-60	1.33*	0.8*	N/A	N/A
Lin et al., 2015	Edinburgh, United Kingdom	Aeroqual/Series 500	1-hour	0-90	N/A	-6.82	0.91	N/A
Bart et al., 2014	Lower Fraser Valley of British Columbia	Early Aeroqual S500	1-minute	0-50	0.9 ± 0.2	6 ± 6	N/A	N/A
	10 sites in British Columbia	Early Aeroqual S500	1-hour	0-08	1	-1	N/A	N/A
Duvall et al., 2016	Houston, TX	Cairpol CairClip	1-hour	0-125	0.93	4.25	0.80	N/A
	Golden, CO	Cairpol CairClip	1-hour	0-90	1.19	0.37	0.77	N/A

*Indicates that the value was calculated based on data within the cited source. For slope/intercept, results were mathematically manipulated to present slopes and intercepts in a similar form as is requested in this report. It should be noted that these results are 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 (ppbv)	Slope	Intercept (ppbv)	R^2	RMSE (ppbv)
Literature and Evaluation Data Summary*	Range†	No data	0.66 to 1.23	7.14 to 14.53	0.74 to 0.98	3.65 to 7.22
	Average†	No data	0.95	0.79	0.90	5.71
	Median†	No data	0.98	0.00	0.89	6.25
People's Republic of China MEE		5% (relative SD)	Not discussed/ not listed	Not discussed/not listed	$\geq$ 0.64	Not discussed/ not listed
This Report	Target (Base Testing Only)	$\leq$ 5 (SD) OR $\leq$ 30% (CV)	1.0 ± 0.2	$-5 \leq b \leq 5$	$\geq$ 0.80	$\leq$ 5

*Data only includes 1-hour averaged data from field evaluations.
†Several values were excluded when summarizing the range, average, and median of data collected from peer-reviewed literature and sensor evaluation programs. Values were often excluded because they reflect sensor malfunction or poor performance which would prevent the use of the sensor for all potential applications.

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

Precision: $\leq$ 5 ppbv (SD) OR $\leq$ 30% (CV)

Precision metrics in peer-reviewed literature are presented in a variety of ways leading to difficulty in comparing results. Common to all sources, strong precision between sensors is defined by similar measurements from identical sensors when subjected to various 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 ppbv 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 O_3 concentration in the environment, they have perfect precision even though the O_3 sensors are non-functional. Results from current sensor evaluation efforts observe slight differences between measurements from even the best performing sensors which is allowed for within a range of precision target values.

Slope (Bias): 1.0 ± 0.2

The target value for the slope component of bias is 1.0 ± 0.2. A slope around 1 indicates that O_3 sensors respond similarly to reference instruments at various O_3 concentrations. This is extremely important for NSIM applications where relative difference, or the amount of change, is important. This target value proposed in this report prescribes a confidence interval to assist testers in evaluating performance.

Intercept (Bias): $-5 \leq b \leq 5$ ppbv

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

R^2 (Linearity): $\geq$ 0.80

Higher R^2 values are associated with closer agreement and better linearity between two data sets being compared. 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 (Error): $\leq$ 5 ppbv

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. RMSE is an important metric for NSIM applications where sensor and reference instrument measurements need to be compared. RMSE is sensitive to data points with large differences between sensor and reference instrument measurements. 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).

Effect of Interferents: No Target Value Established

The role of chemical interferents on a sensor's O_3 concentration measurements is of importance to many NSIM applications, whether the presence of these interferences is known or unknown. The U.S. EPA FRM/FEM Program allows for ± 5 ppbv in the O_3 concentration with respect to each interferent in chemiluminescent and gas-phase spectrophotometric O_3 analyzers, though this standard is likely too stringent for air sensor performance. At this time, the performance data for the effect of interferents on O_3 sensor measurements are quite limited, therefore a target level has not been established for this performance metric. Rather, the protocols outlined in this report recommend that testers quantify the influence of several interferents (CO, NO_2, and SO_2) known to influence electrochemical and metal oxide-based O_3 sensors. Additional interferents may be identified through continued research and may

change as a result of technology advancements. This work may inform the future establishment of a target value.

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

The effect of RH on O_3 sensor measurements is an important performance metric for all NSIM application areas especially because RH varies across the U.S. and can change rapidly throughout the day. Many studies have shown that currently available O_3 sensors are affected by RH (Williams et al., 2013; U.S. EPA 2014b and 2015; Lin et al., 2015; Pang et al., 2017; Borrego et al., 2018). Currently, O_3 sensors are not typically heated to reduce moisture content in the sample and can be affected by condensation. Literature sources attempting to quantify the effect 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. This influence of T is particularly hard to understand since O_3 and T are also typically correlated. Considering the very limited performance data available, this report does not establish a target value for T. 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. At this time, 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 recommend 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 Concentration: No Target Value Established

Many sensor manufacturers/developer claim the ability to accurately measure O_3 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 O_3 concentrations is important for areas that experience such conditions, for NSIM applications focused on exceptional events, 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**	**O_3 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 QC 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 for O_3 - Base Testing
Manufacturer & Air Sensor Name

Deployment Number
Testing Organization
Contact Email / Phone Number
Date

Image of device during deployment

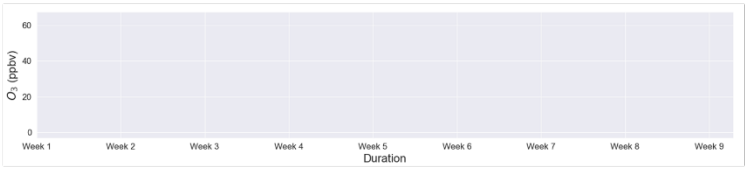
Deployment Details

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	

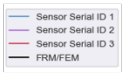
Sensor Information			
Manufacturer, model			
Device firmware version			
Sampling time interval			
Sensor serial numbers	#1	#2	#3
Issues faced during deployment?	☐	Brief summary of issues	

FRM/FEM Monitor Information	
Manufacturer, model	
Sampling time interval	
Date of calibration	
Date of one-point QC check	
Description, date(s) of maintenance activities	

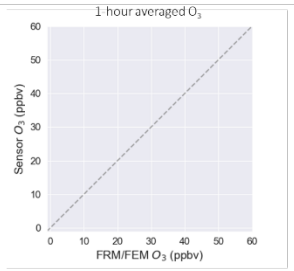
Time Series Plot: 1-hour Average O_3



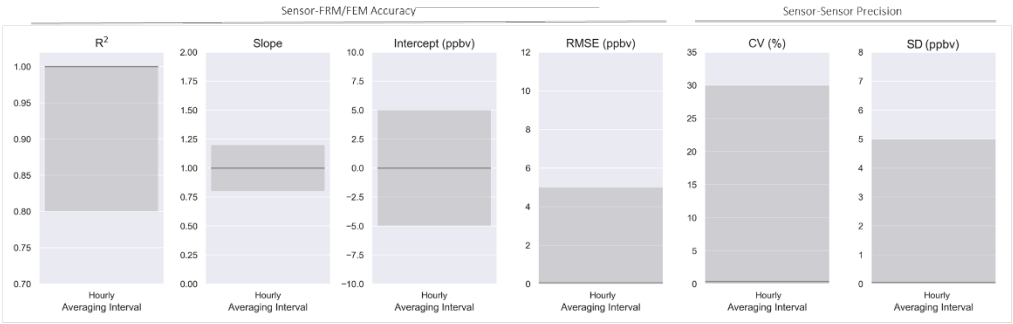
Range of FRM/FEM monitor concentrations over duration of test (ppbv)	
Number of 1-hour periods in FRM/FEM monitor measurements with a goal concentration of ≥ 60 ppbv (*as applicable*)	



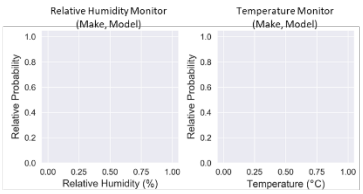
Scatter Plot: Comparison to FRM/FEM Monitor



Performance Metrics*

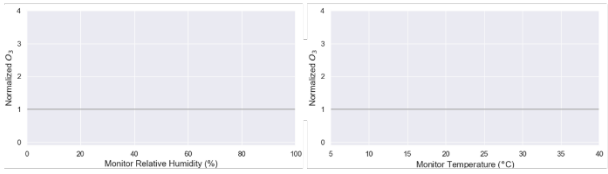


Hourly Meteorological Conditions During Deployment



Number of 1-hour periods outside manufacturer-listed temperature target criteria	
Number of 1-hour periods outside manufacturer-listed relative humidity target criteria	

Hourly 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, CV, and standard deviation.

1

Testing Report for O_3 - 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) (ppbv)	Uptime (%)	Number of paired sensor and FRM/FEM concentration values
	1-Hour ○○○	1-Hour ○○○	1-Hour ○○○	1-Hour ○○○	1-Hour
Metric Target Range	≥ 0.80	1.0 ± 0.2	-5 ≤ b ≤ 5	75%	-
Sensor Serial #1					
Sensor Serial #2					
Sensor Serial #3					
Mean					

	Error
	RMSE (ppbv)
	1-Hour ○
Metric Target Range	≤ 5
Deployment Value	

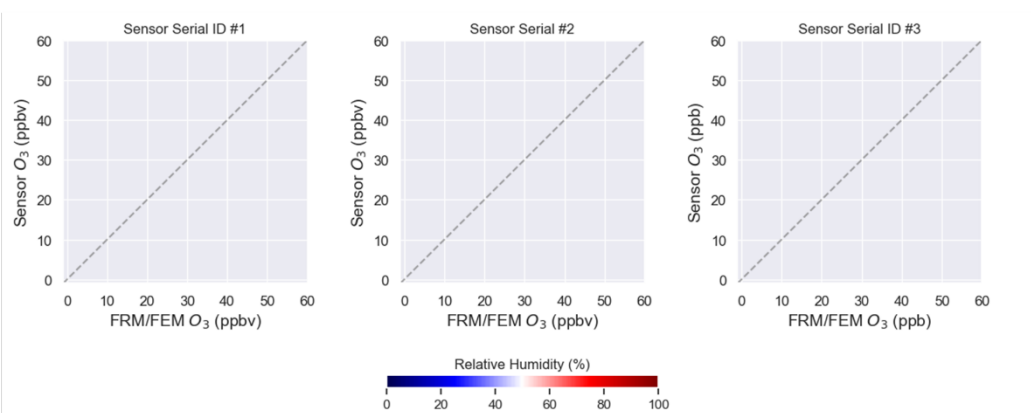
Device-specific metrics (computed for each sensor in evaluation)
○○○ Indicates that the metric value for none of the devices tested falls within the target range
●○○ Indicates that the metric value for one of the devices tested falls within the target range
●●○ Indicates that the metric value for two of the devices tested falls within the target range
●●● Indicates that the metric value for three of the 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 (ppbv)	Uptime (%)	Number of concurrently reported sensor concentration values
	1-Hour ○	1-Hour ○	1-Hour ○	1-Hour
Metric Target Range	≤ 30	≤ 5	75%	-
Deployment Value				

Individual Sensor-FRM/FEM Scatter Plots for 1-hour Averaged O_3



Testing Report for O_3 - 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 specifications 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 FRM/FEM monitor information, including:

Item (as applicable)	RH Monitor	T Monitor	O_3 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				
Date of Calibration at Test Location				
Date of One-point QC 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 2.2.8)

- ☐ 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 – O_3 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)	

O_3 FRM/FEM Monitor Information	
Manufacturer, model	
Sampling time interval	
Date of calibration	
Date of one-point QC 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	
FRM/FEM Monitor Documentation	
Description, date(s) of maintenance activities	☐
Sensor Documentation	
Additional interference testing information	☐
Product Specification Sheet	☐
Product Manual	☐
Description of parameters measured and units, and data flow	☐
Data storage and transmission method	☐
Data correction method	☐
Data analysis/correction script and version	☐
Testing Chamber Documentation	
Description of chamber and O_3 test gas generator system	☐

Effect of Interferents

			Interferent test concentration	Average RH (%)	Average T (°C)	Average FRM/FEM monitor O_3 concentration with interferent (ppbv)	Average sensor O_3 concentration with interferent (ppbv)	Average sensor O_3 concentration without interferent (ppbv)	Average influence of interferent on sensor measurements (ppbv)
Effect of Interferents	Interferent Pollutant: CO	Setpoint	35 ppmv ± 5%	40 ± 5	20 ± 1	70 ± 5%			
		Measured Value							
	Interferent Pollutant: NO_2	Setpoint	100 ppbv ± 5%	40 ± 5	20 ± 1	70 ± 5%			
		Measured Value							
	Interferent Pollutant: SO_2	Setpoint	75 ppbv ± 5%	40 ± 5	20 ± 1	70 ± 5%			
		Measured Value							

1

Testing Report – O_3 Enhanced Testing
Manufacturer & Air Sensor Name

Testing Organization
Contact Email / Phone Number
Date

Image of device during chamber evaluation

Effect of Relative Humidity (RH)

RH Monitor	Manufacturer	
	Model	

			Average RH (%)	Average T (°C)	Average FRM/FEM monitor O_3 concentration of test gas (ppbv)	Average sensor O_3 concentration (ppbv)	Averaged influence of RH on sensor measurements (ppbv)
Effect of RH	Initial Testing Conditions	Setpoint	40 ± 5	20 ± 1	70 ± 5%		
		Measured Value					
	High RH Conditions	Setpoint	85 ± 5	20 ± 1	70 ± 5%		
		Measured Value					

Effect of Temperature (T)

T Monitor	Manufacturer	
	Model	

			Average RH (%)	Average T (°C)	Average FRM/FEM monitor O_3 concentration of test gas (ppbv)	Average sensor O_3 concentration (ppbv)	Averaged influence of T on sensor measurements (ppbv)
Effect of T	Initial Testing Conditions	Setpoint	40 ± 5	20 ± 1	70 ± 5%		
		Measured Value					
	High T Conditions	Setpoint	40 ± 5	40 ± 1	70 ± 5%		
		Measured Value					

2

Testing Report – O_3 Enhanced Testing
Manufacturer & Air Sensor Name

Testing Organization
Contact Email / Phone Number
Date

Image of device during chamber evaluation

60-Day Low Concentration Drift

			Average RH (%)	Average T (°C)	Average FRM/FEM monitor O_3 concentration of test gas (ppbv)	Average sensor O_3 concentration (ppbv)	Sensor drift after 60 days (ppbv)
60-Day Low Concentration Drift	Day 1	Setpoint	40 ± 5	20 ± 1	15 ± 10%		
		Measured Value					
	Day 30	Setpoint	40 ± 5	20 ± 1	15 ± 10%		
		Measured Value					

60-Day Mid Concentration Drift

			Average RH (%)	Average T (°C)	Average FRM/FEM monitor O_3 concentration of test gas (ppbv)	Average sensor O_3 concentration (ppbv)	Sensor drift after 60 days (ppbv)
60-Day Mid Concentration Drift	Day 1	Setpoint	40 ± 5	20 ± 1	70 ± 5%		
		Measured Value					
	Day 30	Setpoint	40 ± 5	20 ± 1	70 ± 5%		
		Measured Value					

Accuracy at High Concentration

		Average RH (%)	Average T (°C)	Average FRM/FEM monitor O_3 concentration of test gas (ppbv)	Average sensor O_3 concentration (ppbv)	Test averaged difference between sensor and FRM/FEM O_3 concentrations (ppbv)
Accuracy at High Concentration	Setpoint	40 ± 5	20 ± 1	125 ± 5%		
	Measured Value					

3

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

United States
Environmental Protection
Agency

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

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

Official Business
Penalty for Private Use
$300