DRAFT

Revision 1

January 23, 2002

CONTINUOUS MONITORING IMPLEMENTATION PLAN

Table of Contents

List of Abbreviations and Terms

AQI	Air Quality Index - An numerical and color coded index for reporting timely air quality to the public for five major pollutants: ground-level ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide.
Bias (total)	The systematic or persistent distortion of a measurement process which causes errors in one direction.
BAM	Beta Attenuation Monitor - A monitor that uses a source and detector of emitted beta particles to determine the collection of particulate matter.
CAC	Correlated Acceptable Continuous *(as currently applied)* - A continuous PM2.5 monitor collocated with a FRM having sufficient comparability to allow for a reduction in sample frequency of the FRM from daily to one in three days.
CAMM	Continuous Ambient Mass Monitor - A monitor that measures changes in pressure drop across a filter tape with particulate matter collected on it to determine the concentration of fine particulate.
CASAC	Clean Air Science Advisory Committee - A group charged with statutorily mandated responsibility to review and offer scientific and technical to the Administrator on the air quality criteria and regulatory documents which form the basis for the National Ambient Air Quality Standards.
DQO	Data Quality Objectives - Are qualitative and quantitative statements that clarify study objectives, define the appropriate type of data, and specify tolerable levels of potential errors, that will be used as the basis for establishing the quality and quantity of data needed to support decisions.
FEM	Federal Equivalent Method - 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.
FRM	Federal Reference Method
Measurement Precision (total)	A measure of the mutual agreement among individual measurements of the same property, usually under prescribed similar conditions, expressed generally in terms of the standard deviation.

Primary Monitor	Identifies one instrument as the sanctioned monitor for comparison to the NAAQS when there are multiple instruments measuring the same pollutant at the same site.
REM	Regional Equivalent Monitor - A potential new type of equivalent monitor being proposed in this document that would be limited geographically in its approval to where its performance has been successfully demonstrated.
SAMWG	State Air Monitoring Working Group
SES	Sample Equilibration System - A technology utilizing a Naphion® dryer that allows sample flow streams to be conditioned to low humidity and temperature.
STAPPA/ALAPCO	State and Territorial Air Pollution Program Administrators / Association of Local Air Pollution Control Officers
TEOM	Tapered Element Oscillating Microbalance - A particulate matter continuous monitor that utilizes an inertial balance which directly measures the mass collected on a filter by measuring the frequency changes on a tapered element.

Executive Summary

An enlarged continuous PM monitoring network will improve public data reporting and mapping, support air pollution studies more fully by providing continuous (i.e., hourly) particulate measurements, and decrease the resource requirements of operating a large network of nearly 1200 filter-based particulate samplers. This document provides recommended directional guidance to move forward in deploying a valued continuous PM monitoring program operated by State and local agencies and tribal governments. A range of topics are addressed, including relationships between continuous and filter-based measurements, performance analyses of collocated continuous and filter based samplers, recommended performance criteria, regulatory modifications, and identification of outstanding technical issues and actions to be taken in the near future.

This plan proposes a hybrid network of filter based and continuous mass samplers. The hybrid network would include a reduced number of existing FRM samplers for direct comparison to the NAAQS and continuous samplers that meet specified performance criteria related to their ability to produce sound comparisons to FRM data. Two approaches for integrating continuous mass monitors are proposed to maximize flexibility for agencies; an expanded use of Correlated Acceptable Continuous Monitors (CAC), and a new Regional Equivalent Monitor (REM) program. The CAC approach would enable agencies to address any monitoring objective, other than ***direct*** comparisons to NAAQS for attainment and non-attainment designations, while the REM approach would serve any monitoring objective.

In either approach, if data produced by a continuous monitor differ from that produced by the filter-based method, then monitoring agencies should seek to optimize the continuous method to reduce those differences. If all established means to optimize the continuous method have been exhausted, and the differences in data from the filter-based method and continuous monitor are still not acceptable, then the continuous data should be adjusted to be more comparable to that of the filter-based method. Only simple adjustments will be allowed for the REM whereas any type of adjustment will be allowed for the CAC. At sites operating a continuous instrument that is not collocated with a filter-based instrument, assumptions will have to be made about the adjustment that is appropriate to produce data that is comparable to a filter-based sampler. The general approach proposed in this document is to determine geographical regions within which one adjustment is appropriate for all of the continuous measurements. There is flexibility in the adjustments and regions associated with a CAC, whereas the adjustments and regions associated with the REM will be restricted and subject to an independent review through EPA's Office of Research and Development or a similar entity.

Two performance criteria are proposed to determine whether the adjusted continuous measurements are sufficiently comparable to be integrated into the $PM_{2.5}$ network. These criteria are bias (relative to a filter-based method) between -10% and +10% and precision less than 20% coefficient of variance (CV). These criteria are the result of a data quality objective (DQO) analysis that is based on data from the existing $PM_{2.5}$ network and based on the assumption that the annual

$PM_{2.5}$ air quality standard is the principal decision driver. Also, the DQO result is conservative in that the goals guarantee decision error rates for the "worst case" scenarios. In cases that are not "worst," the DQO approach allows for additional flexibility beyond the stated bias and precision goals.   These performance criteria preferably would be demonstrated by monitoring agency staff under actual operational conditions, a departure from the very tightly controlled approach used for national equivalency demonstration, and validated periodically in recognition of changing aerosol composition and instrument performance.

Recommended changes in the PM monitoring regulations are proposed that will reduce the number of required FRM samplers nationally, a divestment needed to generate operational resources to stimulate deployment of continuous mass samplers.

Many of the proposals in this document require additional work before formal guidance can be completed. Some of the additional work includes increasing the number of sites with collocated FRM and continuous samplers in order to characterize relationships between sampling methods and the spatial extent of those relationships. Such characterization is necessary for initial and ongoing integration of the continuous monitors into the entire $PM_{2.5}$ mass network. The final section of this document lists the areas requiring additional work.

Section 1. Introduction

This document presents a proposal for enhancing the continuous particulate matter monitoring in the air monitoring networks operated by State and local agencies and tribal governments. The document addresses a range of topics including recommended performance requirements, regulatory modifications, and identification of outstanding technical issues and actions to be taken in the near future.

EPA is working with the Clean Air Science Advisory Committee (CASAC) technical subcommittee on particle monitoring; State and local agencies and tribal governments; and consortiums of State and local agencies on a strategy to enhance deployment and utility of continuous fine particulate mass monitors. This document is an important step in this cooperative effort as it provides a basis for comment on our intended approaches. Clearly, a substantial subsequent guidance development effort will be required to implement the directions in this proposal. Comments are welcome from all interested stakeholders on this document as well as the national air monitoring strategy it is intended to support.

The reader should be aware that the concepts and elements incorporated in this plan are singularly and collectively complex therefore creating a communications challenge. Other approaches were considered, but the potential drawbacks of a simplistic approach were not acceptable. That is, it would have been easy to develop a rigorous non-flexible program easily communicable but conveying little motivation for deployment. Similarly, a program without constraints would likely compromise data quality and interpretability. Thus, a decision was made to accommodate both flexibility and data comparability at the expense of developing and communicating a complex program.

The development of “acceptable” relationships between a Federal Reference Method (FRM) measurements and continuous monitors is stressed throughout this document. The reason for this is that so many objectives relate to the FRM measurement (e.g., NAAQS comparisons, air quality index reporting, air quality model application). In many instances, there is no technical reason to expect comparability between disparate measurement approaches. Such comparability is desired given the utility of relating continuous measurements to a wealth of existing FRM data and to incorporate a reference marker. The downside of this approach is that the value of an FRM measurement is assumed or inferred to be greater than that of a candidate method, when in some cases the candidate method may better reflect “true” characteristics of an aerosol. This topic is addressed in more detail in Section 7.

Background

EPA is motivated to develop the continuous monitoring program by the need to improve public data reporting and mapping, support air pollution studies more fully by providing continuous (i.e., hourly) particulate measurements, and to decrease the resource requirements of operating a large network of filter-based particulate samplers. This document also addresses an important gap in technical guidance for the continuous particulate matter program, created in part by a strong emphasis to date on compliance (Federal Reference Method) and chemical speciation sampling.

Approximately $170 million has been directed toward the deployment and operation of the $PM_{2.5}$ network since July 1997, and the $PM_{2.5}$ network continues to operate at a cost of $42 million annually. The majority of the annual expenses are for the operation and maintenance of the federal reference method samplers, $26.5 million. The introduction of continuous particulate matter monitors capable of addressing multiple objectives with reduced operator burden could produce desired network efficiencies. For example, the cost of operating a federal reference method sampler on a one-in-three day schedule for a year is approximately $19,000 (including operations, maintenance, data management, filters, and quality assurance audits). The cost of operating one of the available continuous (hourly) particulate matter samplers is approximately $8,000. EPA does not expect that all federal reference method samplers will be replaced; however, significant resources can be impacted by the use of more continuous samplers in lieu of some federal reference methods.

Assessments of existing criteria pollutant networks are being conducted as part of a separate but parallel National Air Monitoring Strategy effort. These assessments are providing direction for reducing the current number of $PM_{2.5}$ federal reference methods based on observed spatial redundancy (due to relatively broad homogeneous fine aerosol behavior throughout the eastern United States) and related factors. Such divestment in filter based methods is needed to support integration of a more comprehensive continuous mass network, as well as preparing for future coarse particulate monitoring requirements. This comprehensive air monitoring strategy also has defined progress in continuously operating PM monitors as a priority for implementation.

Over the last four years many monitoring agencies have expressed a strong desire for the development and acceptance of continuous methods for use as compliance samplers (i.e., federal equivalent methods). This sentiment has been expressed in a number of venues including the Air and Waste Management Association PM2000 conference; through the STAPPA/ALAPCO Monitoring Committee and the Standing Air Monitoring Work Group (SAMWG); and the CASAC Technical Subcommittee on Particle Monitoring. The CASAC Technical Subcommittee on Particle Monitoring met on January 22, 2001 in a workshop session dedicated to continuous particulate matter monitoring. This document provides further details on EPA's proposal to enhance continuous PM monitoring as a follow-up to that CASAC workshop. The approach utilizes the data quality objective process to develop continuous monitor performance specifications. State and local agencies and tribal governments would have a set of parallel options through a new Regional Equivalent Method program and an modification of the existing Correlated Acceptable Continuous monitors provision.

The principal challenge implied within this document is maintaining an acceptable balance between data quality and technological progress. The promulgation of the 1987 PM_{10} standards included a performance-based approach to the acceptance of PM_{10} methodology. The current $PM_{2.5}$

monitoring network has achieved relatively high data quality[1] due in large measure to the requirement of design-based methods (i.e., monitors with virtually identical components) and a thorough quality assurance program that followed through on a cycle of planning (data quality objectives), implementation (field/laboratory quality control), data assessment and reporting tasks. Risk in compromising data quality will emerge as an assortment of technologies are accommodated in the network. Consequently, the success of this program will rely not only on the initial data quality objective planning steps, but through a commitment to conducting the remaining quality assurance tasks and retaining the flexibility to take appropriate action in the use of data when systematic failures are encountered within the quality assurance system.

Document Layout

Section 2 examines the available collocated federal reference method and $PM_{2.5}$ continuous monitoring data. This examination illustrates both the successes and challenges of implementing PM continuous monitors. Section 3 and 4 detail the applicability of the correlated acceptable continuous (CAC) monitors and the regional equivalent monitors (REM) including testing requirements and the approval process. Section 5 focuses on network design emphasizing the suggested hybridization of federal reference method and continuous particulate monitors, and proposing a new minimum number of required $PM_{2.5}$ federal reference method sites. Section 6 provides the performance standards for using PM methods and a description of the data quality objective process utilized to derive the goals for precision and bias. The data quality objective process recognizes a number of variables such as measurement precision, population precision, sample bias, sample frequency, a 3-year standard, and sample completeness in order to predict the confidence in a decision around the annual average. Section 7 addresses the use of statistical transformations for each category of continuous methods. The use of such transformations need careful consideration in terms of number of variables, frequency of adjusting, and spatial scale of applicability. Section 8 provides some initial thoughts on developing boundaries for approval of methods across a spatial scale. This section details how a number of inputs such as aerosol composition using both monitored data and modeled data as well as overlaying this output with natural geographic boundaries, such as how State lines or city boundaries may be used. Section 9 provides design guidance on continuous monitoring methods. Section 10 identifies how this effort to enhance a network of continuous particulate monitors is linked to the national monitoring strategy. Section 11 provides a summary of the potential regulatory changes and schedule necessary to implement this plan. Section 12 provides a repository of issues and action items.

Applicability

The scope and intention of this document is focused on addressing continuous particle mass

[1] "CY 2000 Quality Assurance Report of the $PM_{2.5}$ Ambient Air Monitoring Program," U.S.EPA Office of Air Quality Planning and Standards, October 2001.

monitors that provide in-situ sampling/analysis capability producing outputs that can be aggregated upward to one-hour reporting periods (e.g., TEOMs and beta attenuation gauges). The approaches proposed rely on the use of the data quality objective process to produce quantitative performance standards. This process would in concept accommodate alternative particulate matter measurement approaches beyond the more traditional continuous mass methods, assuming performance standards are achieved. Such acceptable examples that might provide a useful alternative to the federal reference method include the use of a continuous speciation monitor alone (e.g., sulfate only) or in combination with multiple speciation monitors (e.g., carbon, nitrate and sulfate), or other filter based methods that do not have current equivalency status (e.g., dichotomous sampler). The principles described in this document are not applicable to measurement systems beyond particulate matter (e.g., utilizing particulate matter measurements to replace ozone or other discrete gaseous measurements).

Section 2. $PM_{2.5}$ Continuous/FRM Relationships

Introduction

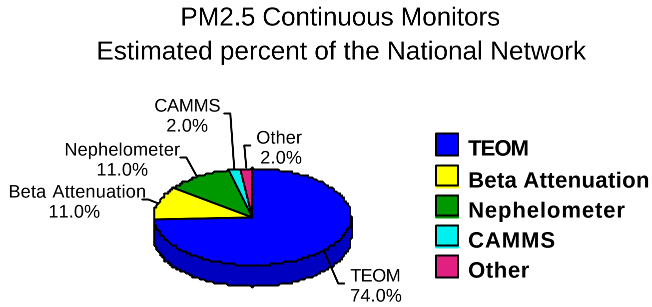
This section represents an initial effort to compile relational analyses between continuous and FRM data. Relationships between $PM_{2.5}$ continuous and FRM monitors are synthesized from a number of sources, including routinely collected data provided by State and local agencies and data from available field studies. The task of comparing $PM_{2.5}$ continuous data with FRMs was accomplished by averaging the hourly continuous mass data between midnight to midnight, to parallel the FRM operations. General information is provided first with a number of analyses presented later in this section. A more detailed set of analysis are presented in Attachment A.

General Summary

Continuous monitors track FRM data with varying degrees of success across the country, with a mix of seasonal and geographical patterns affecting behavior. Analyses to date are somewhat limited by the availability of relatively few formal field studies, and the current (and temporary) situation where only one $PM_{2.5}$ continuous method (the TEOM[2] operated at 50C) has been widely deployed (Figure 2-1). Despite these limitations, there is an emerging understanding that the best $PM_{2.5}$ continuous monitor choice may vary from one monitoring agency to the next. TEOMs operated at 50C appear to predict FRM measurements in locations where volatile losses are minimal. Examples include sites with sulfate dominated aerosols in the Southeast (the Carolinas and Georgia) throughout the year and northeastern and upper Midwest (Iowa and Michigan) locations during the summer. The prevalence of winter month underestimates in certain areas suggests that the TEOM operated at 50 C exacerbates volatile losses during cool conditions when the difference between operational and ambient temperature is greatest. Converting the 50C TEOM to a 30C TEOM with a Sample Equilibration System (SES) should reduce cool season volatile losses. Analyses comparing collocated 50 C and 30 C TEOMs with the SES and FRMs at sites in North Carolina and New York State indicate improved comparability to the FRM for the 30 C TEOM with the SES.

[2]Manufactured by Rupprecht & Pataschnick.

Figure 2-1 Percent of PM2.5 Continuous Methods used Nationally



The beta attenuation monitor (BAM)[3] is operated at several locations (second in number to the TEOM) throughout the western United States with a limited number of new locations in the east. The California Air Resources Board and other organizations sponsored a field study of several major $PM_{2.5}$ commercially available monitors indicating high performance of the BAM conducted during relatively volatile aerosol conditions.[4] EPA's Environmental Technology Verification Program (ETV) included two test sites; one in Pittsburgh, PA in the summer of 2000; and one in Fresno, CA in the winter of 2000-2001. This verification program included a number of $PM_{2.5}$ continuous monitors being deployed by State and local agencies including the BAM, the TEOM operated at 50C, the TEOM operated with the sample equilibration system at 30C, and the CAMMS[5]. While the verification reports do not offer conclusions as to the performance of the monitors, inspection of these reports indicates that the Met One BAM performed consistent at both test sites. The final verification reports from these field studies are available from the U.S. EPA web site.[6]

The Nephelometer is used at many sites in the Pacific Northwest. This monitor can have

---

[3]Manufactured by Met One Instruments.

[4]Reference the CARB report here.

[5]Manufactured by Thermo Andersen.

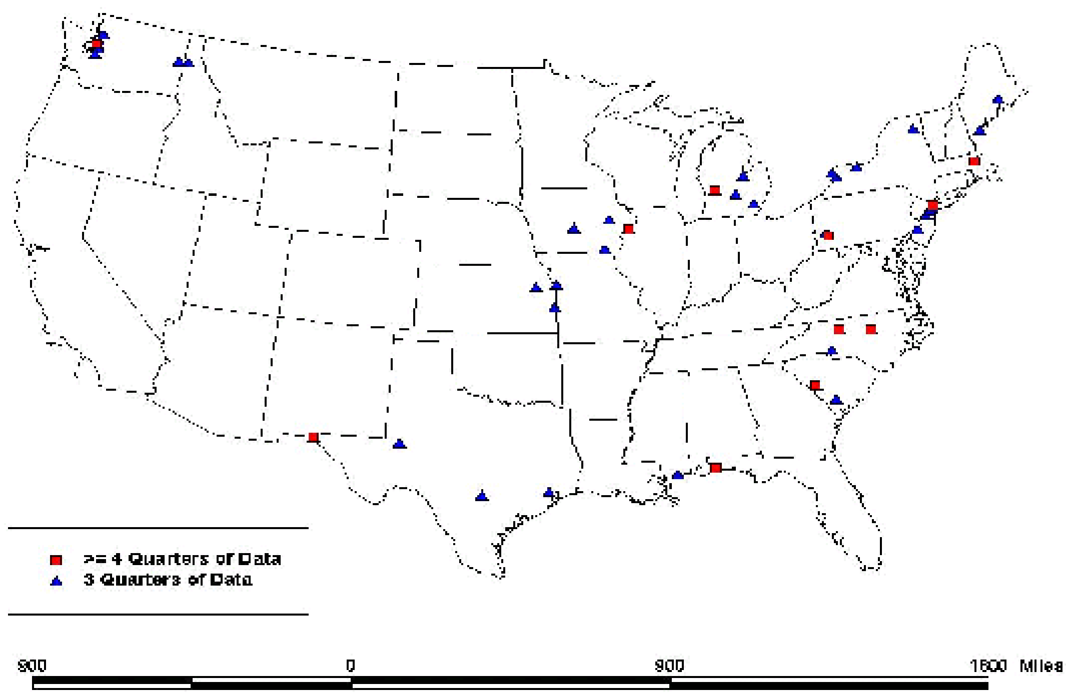
[6]Environmental Technology Verification Statements and Reports:
http://www.epa.gov/etv/verifrpt.htm#07

advantages over PM2.5 continuous methods with respect to its ease of operation. However, Nephelometers can have problems with high humidity and care should be taken to assure sample streams are conditioned so as not to have moisture interfere with the scattering output. There are several manufacturers of Nephelometers, so care also needs to be taken when comparing data from a monitor at one site to another. Although Nephelometers do not provide for a direct output of fine particualte concentration, they can be useful when calibrated against filter based methods to provide for diurnal and day to day signal of fine particulate.

Analysis of the Variety of Relationships for 47 Collocated $PM_{2.5}$ Continuous and FRM Sites

The AIRS database included 11 sites with at least a years worth of collocated $PM_{2.5}$ continuous monitoring and FRM data based on a Spring, 2001 retrieval. An additional 36 sites were included for analyses if they had at least 3 quarters of data with at least 11 valid collocated pairs per quarter for a total of 47 sites (Figure 2-2) forming the basis for the analyses presented in this section.

Figure 2-2 Map of 47 Sites used in $PM_{2.5}$ Continuous Monitors Analyses



Intercomparisons of FRMs and $PM_{2.5}$ Continuous Monitoring Data:

Of the 11 sites with at least 4 quarters of complete data, 8 sites used TEOM monitors with the factory installed correction factor applied for the entire data set. This factory installed correction factor adds 3 ug to the intercept and 3% to the slope for data coming from a TEOM. A table summarizing the range of concentration values from each of the FRM and continuous monitors at these sites is provided below:

Table 2-1 Concentration Ranges for 8 Sites with Collocated PM2.5 FRM and TEOM Monitors

MSA	Site ID	N	Primary Monitor Type	Concentration Range of Data ($\mu g/m^3$)						
				Mean	SD	Min	Q1	Median	Q3	Max
Aiken, SC - Augusta, GA	450370001	144	Continuous	14.50	6.42	1.37	9.85	13.46	18.88	34.75
			FRM	14.49	6.55	2.40	9.75	13.00	18.00	34.20
Davenport, IA - Moline - Rock Island, IL	191630015	453	Continuous	12.00	6.49	2.92	7.26	10.53	15.30	48.81
			FRM	12.81	7.31	2.30	7.30	11.50	16.90	46.70
Winston - Salem, NC	370670022	525	Continuous	16.23	8.05	2.66	10.29	14.45	20.95	64.02
			FRM	16.89	8.70	1.60	10.60	15.00	21.70	69.70
New York, NY	360050110	295	Continuous	15.40	9.26	4.69	8.85	12.85	19.24	85.38
			FRM	15.21	9.17	3.60	8.30	12.30	20.00	53.00
Pensacola, FL	120330004	214	Continuous	14.41	6.74	-17.7	9.90	13.02	17.94	45.83
			FRM	14.03	6.89	1.00	8.60	12.70	18.41	49.30
Pittsburgh, PA	420030064	344	Continuous	16.68	12.00	1.21	7.27	13.19	22.50	68.92
			FRM	20.87	13.39	3.10	11.00	17.20	26.55	78.50
Raleigh-Durham, NC	371830014	389	Continuous	15.02	6.89	2.78	10.00	13.66	18.98	45.88
			FRM	15.59	7.52	3.00	10.10	14.40	20.00	52.80
Seattle, WA	530330057	340	Continuous	13.30	6.39	3.38	9.08	11.87	15.48	44.42
			FRM	12.64	7.25	2.80	7.80	10.95	15.40	46.90

Inspection of Table 2-1 indicates that most of the sites appear to produce similar PM2.5 concentrations regardless of whether an FRM or TEOM is used. Only the Pittsburgh, PA site showed a large discrepancy between the mean of the FRM and PM2.5 continuous monitor. Due to this discrepancy, the Allegheny County monitoring staff were contacted to confirm the operation of the TEOM and use of default corrections factors. While the operation of the instrument was determined to be correctly identified, it was mentioned that the site is located in a community orientated location in close proximity to a large local source.

Scatter plots were produced for each of the 11 sites with at least a years worth of complete data. Data were plotted for each day where both a FRM value and a corresponding average 24-hour continuous $PM_{2.5}$ value were available. Separate plots for linear and log-normal concentrations were plotted for each site. The scatter plots can be separated into several categories: scatter plots with good agreement most of the time - illustrated by most points being on a straight line (Figures 2-3 through 2-6 and 2-9); scatter plots with a small but discernable amount of spread about the best fit line - as illustrated by a mild spread about the best fit line (Figures 2-7 and 2-8); scatter plots with good agreement part of the time and poor agreement in others - illustrated by a large increasing spread with concentration (Figures 2-10 and 2-11); and scatter plots that do not appear to correspond well with any pattern - illustrated by a large spread about the 1:1 relationship regardless of the concentration (Figures 2-12 and 2-13).

These first four figures represent sites in the southeastern United States where the $PM_{2.5}$ continuous monitor appears to track the FRM reasonably well:

Figure 2-3 Raleigh-Durham, NC

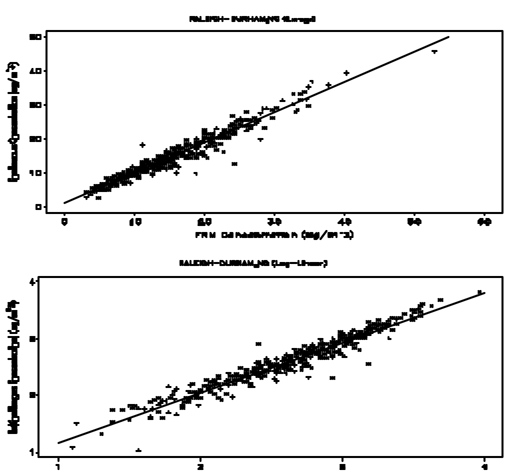


Figure 2-4 Winston-Salem, NC

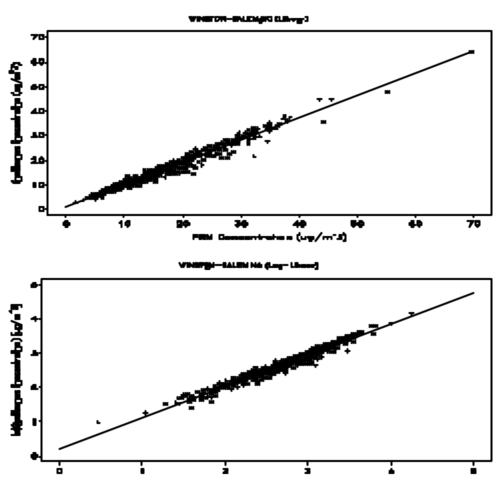


Figure 2-5 Aiken, SC - Augusta, GA

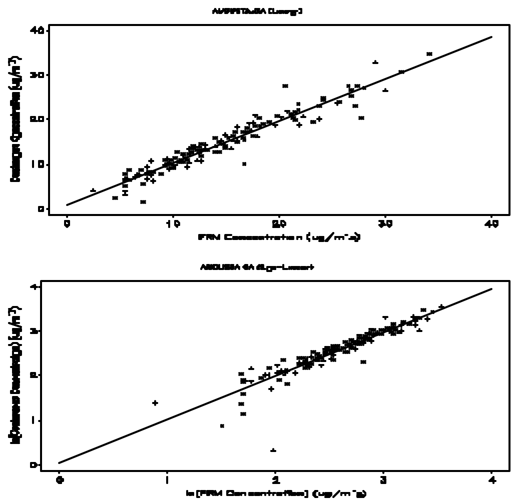
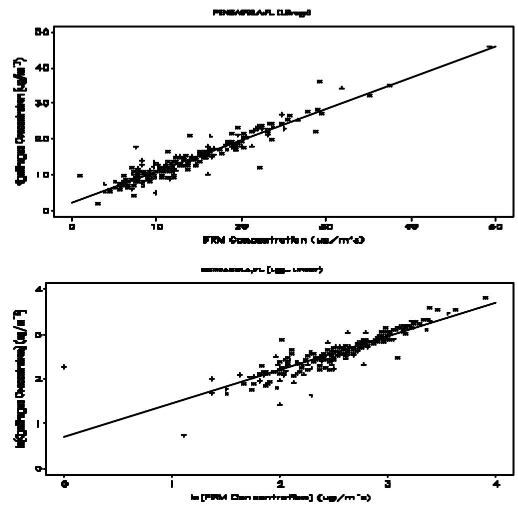


Figure 2-6 Pensacola, FL



The following scatter plots represent cities in the Northeast with some discernable spread about the best fit line, but not severely distorted.

Figure 2-7 New York, NY

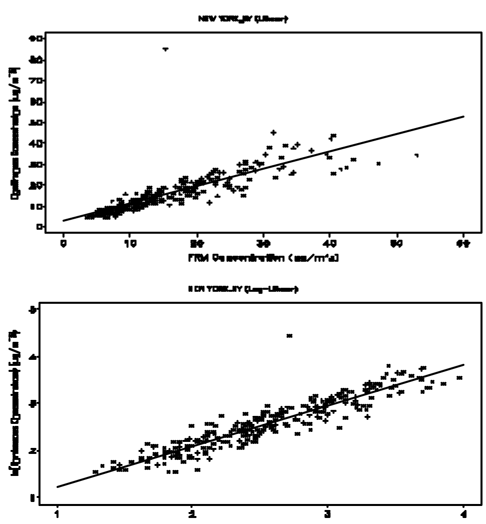
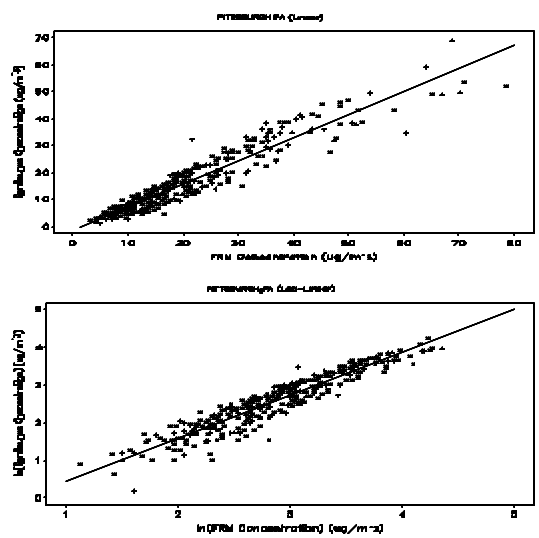
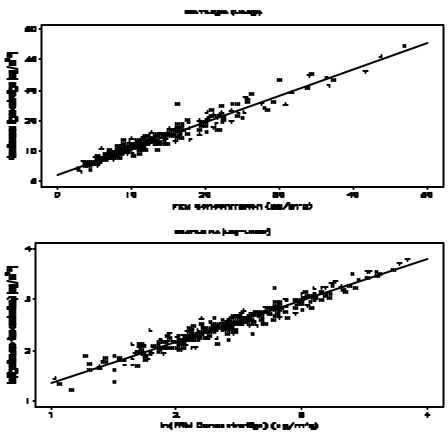


Figure 2-8 Pittsburgh, PA



The following figure is from a northwest site. The scatter plot shows a good fit about the best fit line.

Figure 2-9 Seattle, WA



These figures, using data from sites in the upper mid-west, represent a clear spread with concentration. This is likely an effect of seasonal aerosol changes.

Figure 2-10 Davenport, IA

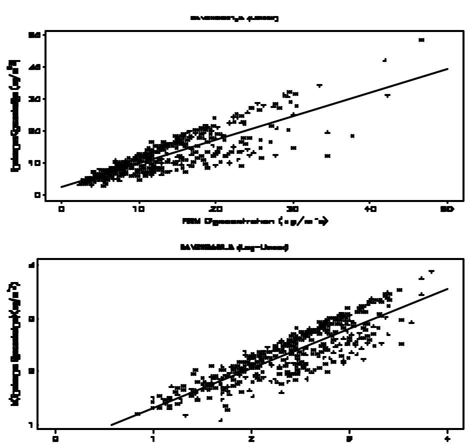
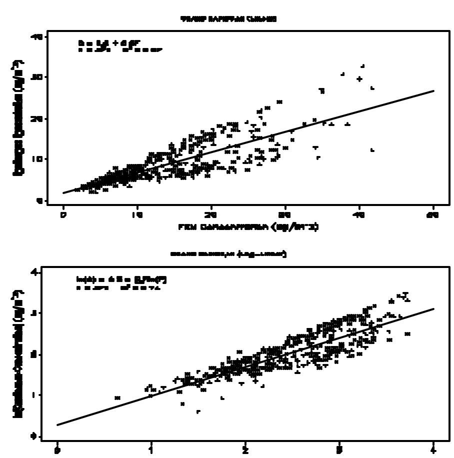


Figure 2-11 Grand Rapids, MI



These figures represent data from air sheds where the TEOM and FRM do appear to correspond well.

Figure 2-12 El Paso, TX

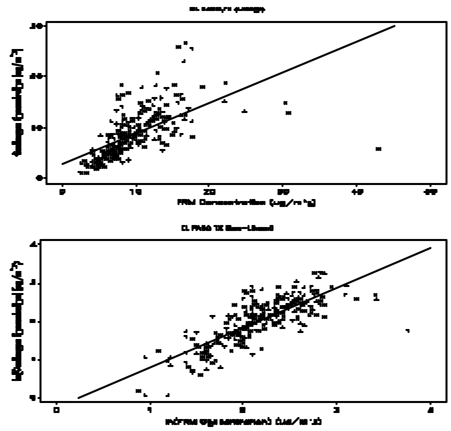
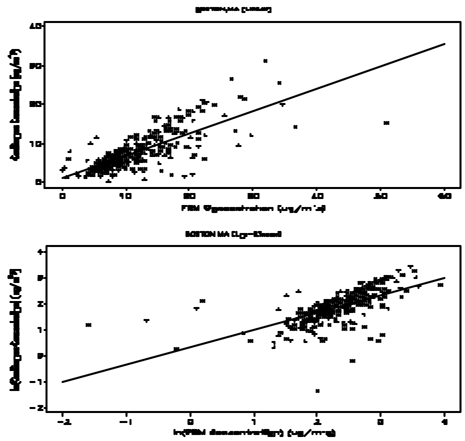


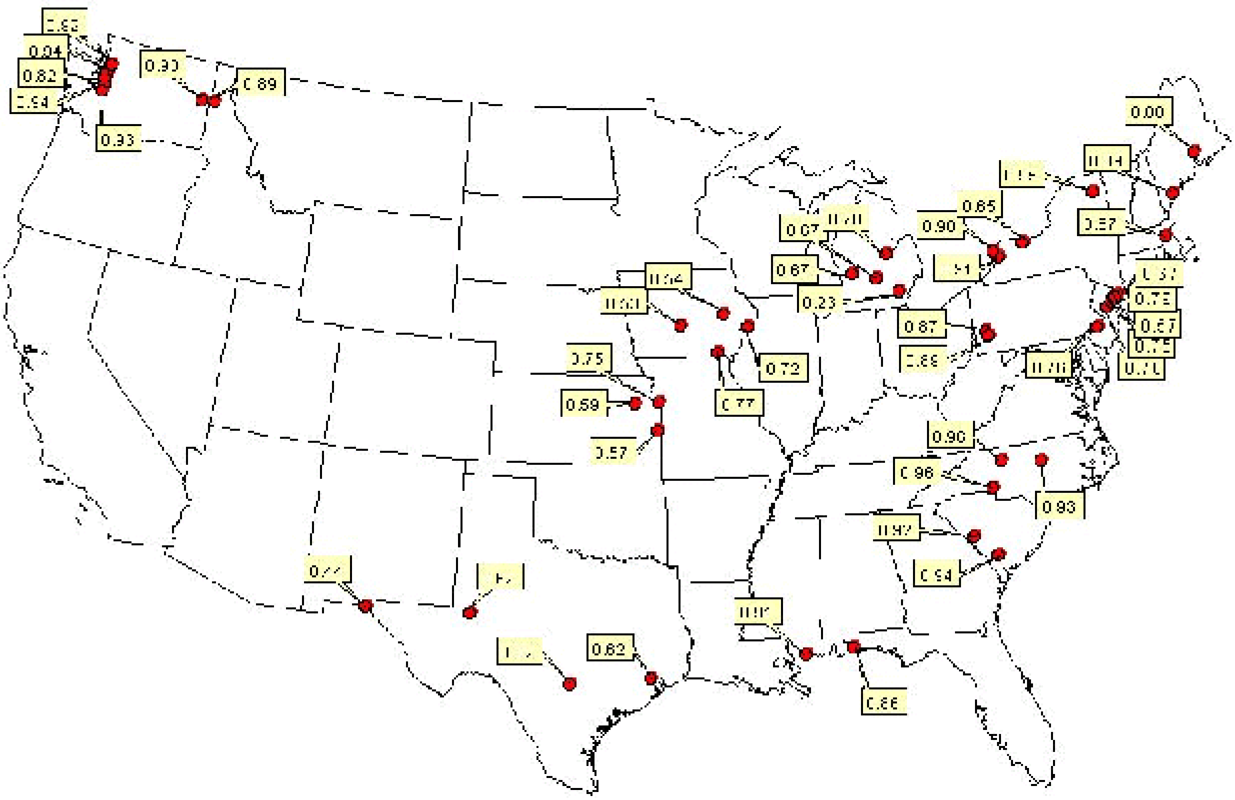
Figure 2-13 Boston, MA



Correlation between $PM_{2.5}$ Continuous Monitors and FRMs

Another way to look at the data is to evaluate the goodness of fit between a model using $PM_{2.5}$ continuous data to explain FRM measurements. The map below (Figure 2-14) illustrates the correlation coefficient (R^2) at each of the available 47 sites. All 47 sites are able to be used because a linear model will not affect the correlation regardless of whether a site specific model is used, the standard correction factors are applied or no model is used at all. The map also indicates that geographical area plays a large role in how high a correlation coefficient is observed. This is likely due to the aerosol encountered at specific sites, the concentration of fine particulate and an effect of the season. Areas exhibiting high correlation include the Southeast, Northwest and selective locations of the Northeast. Areas with poor correlation are likely the result of either regional scale winter time volatilization as demonstrated in Iowa and Kansas or micro-scale to urban-scale influences of local sources such as in Boston and El Paso.

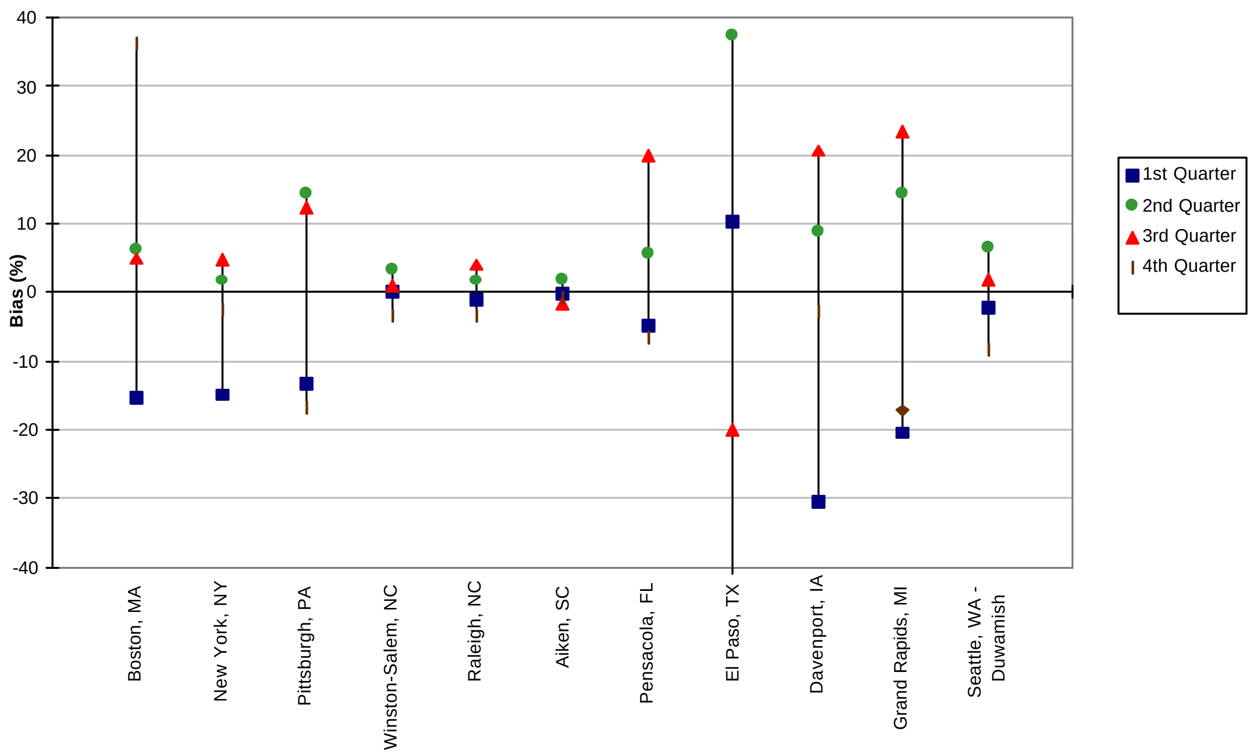
Figure 2-14 Correlation between FRMs and $PM_{2.5}$ Continuous Monitors



Bias by Season

In many air sheds across the United States the species and concentration of the aerosol encountered varies by season. Changes in the species and concentration of the aerosol can lead to changes in performance of a $PM_{2.5}$ continuous monitor. In the illustration below the spread of bias is presented for those sites with at least 4 quarters of complete data. Bias data were calculated by comparing the FRM and collocated continuous monitoring data for days when both instruments produced a valid 24 hour value. Since some monitoring agencies choose to use a standard correction factor in the reporting of their data while others did not, each set of data was first fit to it's own linear model and then the bias were calculated by quarter. Additional graphics depicting the bias by quarter for those sites without 4 complete quarters are available in attachment 1. The tighter the fit between season the better the opportunity to use that continuous instrument to produce FRM-like measurements. Generally, cooler quarters produced the largest negative biases. This is likely due to the larger difference between the operating temperature of the TEOM and the ambient temperature of the atmosphere. The relatively high operating temperature of the TEOM during these cooler months leads to evaporation of a portion of the aerosol that are collected on a filter based sampler.

Figure 2-15 PM2.5 Bias Data for TEOM Monitors by Quarter



Analysis of the Acceptability of the Relationship relative to the Data Quality Objective Process and Class III equivalency.

In the section above, a few of the sites appeared to have $PM_{2.5}$ continuous monitors that are replicating the FRM measurements very well with other sites not performing well and many sites in between. A site may be expected to replicate the FRM very well by virtue of having a scatter plot close to unity, a high correlation coefficient and a low bias. But with a variety of performances across sites, at what level should a site be considered acceptable? In this section data from 160 collocated FRM/FRM sites and 47 collocated $PM_{2.5}$ continuous/FRM sites are compared to various levels of the Data Quality Objective (DQO) process and the equivalency criteria. For the DQO criteria, precision and bias statistics are determined for each site and results are presented as a function of the percentage of sites that satisfied the criteria. For the equivalency criteria, linear regression is performed for each site and results are presented as a function of the percentage of sites that satisfied the criteria.

Table 2-2 Percentage of Collocated Sites meeting individual DQO and Equivalency Criteria

Criteria	160 Collocated FRM/FRM (% of sites meeting criteria)	47 Collocated FRM/Continuous Sites (% of sites meeting criteria)
Data Quality Objective		
Bias 5%	86.9	34.0
Bias 10%	97.5	53.2
Precision 5%	28.1	0.0
Precision 10%	68.8	12.8
Precision 20%	NA	61.7
Equivalency		
Slope (1±0.05)	77.5	91.5
Intercept (±1 µg)	82.5	97.9
Correlation ($0.97)	66.2	10.6

Interpreting Table 2-2 leads to several observations:

C Evaluations of the collocated FRM/FRM sites against the existing goals of ±10% bias and ±10% precision, indicate that precision is the limiting factor. Most (97.5%) of the sites meet the bias goal and 68.8 % meet the precision goal. As will be demonstrated in section 6, bias strongly influences the uncertainty of a 3 year mean, while precision has little effect due to the large number of samples in 3 years of data. Therefore, we have confidence that the FRM network is performing well, as indicted by 97.5% of the sites meeting the bias statistic.

C Evaluating the FRM/FRM sites against the existing criteria for Class III equivalency[7] indicates that correlation is the limiting factor with 66.2% of the sites passing. That's

[7]40 CFR 53

important since we believe we have a well-operating $PM_{2.5}$ FRM network; however, over one-third of the sites would fail the Class III equivalency testing criteria. If a collocated network of FRM cannot largely meet the equivalency criteria, it will be very difficult for a network of FRMs collocated with $PM_{2.5}$ continuous monitors to meet this criteria.

- Evaluations of the collocated FRM/continuous sites against the existing goals of ±10% bias and ±10% precision indicate that precision is also the limiting factor with 53.2 % of the sites meeting the bias goal and only 12.8 % meeting the precision goal. As mentioned above and demonstrated in section 6, bias strongly influences the uncertainty of a 3-year mean, while precision has little effect due to the large number of samples in 3 years of data. If the precision goal could be reduced to ±20%, then 61.7% of the sites in the analysis would have satisfied this criteria. Although an even less stringent precision goal could potentially be chosen, bias has now become the limiting factor for performance of the continuous monitors. While precision could potentially be relaxed and we would still have a high degree of confidence in the 3 year annual mean, the need to monitor for other monitoring objectives necessitates controlling precision to some degree. A detailed explanation of the DQO process will be explained in section 6.

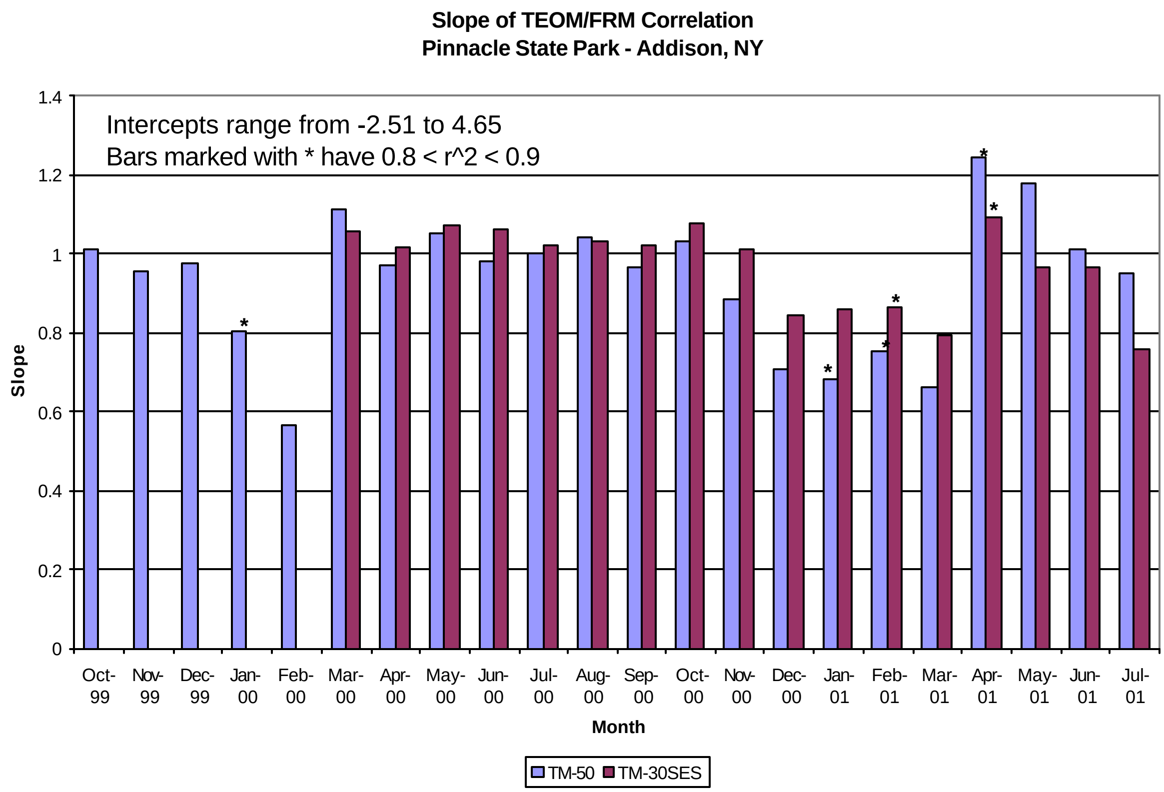
- Evaluating the FRM/continuous sites against the existing criteria for Class III equivalency indicates that correlation is the limiting factor with 10.6% of the sites passing. If it can be demonstrated that the continuous monitors are producing FRM-like measurements that meet the goals established in the DQO process rather than the equivalency criteria, than the correlation criteria becomes irrelevant.

Note: In addition to this analysis the EPA has produced assessments of the quality of the $PM_{2.5}$ monitoring program for the currently operating FRMs for calendar year 1999 and 2000. The calendar year 1999 report is final and can be reviewed on-line at the EPA web site: http://www.epa.gov./ttn/amtic/. The calendar year 2000 report is in review and a draft copy can be obtained from the same web address.

Analysis of Collocated TEOMs with a FRM

In New York State two sites have operating collocated TEOMs with a FRM. Additionally, a site in Raleigh North Carolina also has two TEOMs and a FRM. At each site one of the TEOMs is run with an operational temperature of 50C, while the other is operated at 30C and utilizing a Sample Equilibration System (SES). Data are compared to the operating FRM at the sites, which for all 3 locations is a R&P 2025 FRM. The site with the longest record of data is located at Pinnacle State Park in Addison, NY. This site is located in a rural area of New York's Southern Tier. The illustration below provides some indication of the improvement a TEOM operated at 30 degrees C with a SES can have over operating the conventional TEOM at 50 C. The improvement is most pronounced in the cold weather months of November through March. A table summarizing regressions for all 3 sites by month is available in attachment A.

Figure 2-16 Slope of TEOM/FRM at Pinnacle State Park, NY



Data courtesy of New York State Department of Environmental Conservation and University of Albany, Albany NY.

Conclusion

Although this analyses is very limited it's becoming clear that some areas of the country may already be operating PM continuous monitors that produce data with similar quality to that of the FRM. If a mechanism to approve the use of these continuous monitors could be made where the performance of the instrument is defined to be acceptable than a large resource savings may be gained by divesting of some of the FRM operations. Other areas of the country may not be producing PM2.5 continuous data that could be used to replace the FRM. For these areas, agencies may need to pursue improvements to their instrumentation or new technologies altogether. Comparing the performance of sites that have a collocated FRM/FRM pair with a collocated FRM/continuous pair to the expected equivalency criteria revels that the correlation statistic (r^2 $0.97) would be the limiting factor for either FRMs or continuos monitors to meet equivalency. If this is the case than an evaluation of the expected statistical criteria for equivalency of a continuos monitor should be made. Section 6 of this document examines the performance standards of PM2.5 continuous monitors in detail.

Section 3. Enhanced Correlated Acceptable Continuous Methods (CAC)

A provision to enhance the existing provision for Correlated Acceptable Continuous (CAC) monitors is being proposed in concert with a new Regional Equivalent Monitor (REM) program to provide agencies with options to enhance their network of PM continuous monitors. Rationale based on data comparability for selecting the CAC or REM vehicle is discussed in Section 5 and 6. The basic premise of a revised CAC is to provide flexibility in method selection for PM monitoring sites that are not needed for direct comparison to the NAAQS and for sample frequency relief. These sites would be allowed to use CAC monitors if they meet specified performance criteria. While the current provisions for CAC(s) only allow for a reduction in sample frequency of the accompanying FRM/FEM, the provision under consideration would also allow for a continuous monitor to be approved for use without the collocation of a FRM at sites that are not required for the NAAQS. This additional flexibility is being considered for CAC monitors since no agencies have yet to have a CAC approved and it would be better to enhance the usefulness of CACs rather than to have another provision in the regulation. This approach would potentially be targeted for those agencies that need to monitor for a number of monitoring objectives other than NAAQS attainment decisions. Thus while the CAC cannot be used for attainment decisions - it can be used to meet all other applicable monitoring objectives such as: public reporting, trends, mapping, and exposure. By allowing a portion of the currently required FRM sites in a network to be substituted with continuous monitors meeting performance based criteria, the monitoring agencies can realize a reduction in resource requirements while maintaining data delivery with an acceptable defined level of quality. Also, some of the remaining FRM sites would be collocated with the same continuous methods as the CAC's to provide the performance data for ongoing assessment of the continuous method. These revised CACs would be different than the conventional Federal Equivalent Methods (FEMs) in that they could only replace a limited number of sites and the CAC met the performance criteria specified in Section 6 - Performance Standards for Continuous Monitoring. CACs would be different from REMs in that they could not be used for direct attainment decisions and there would be much more flexibility in the use of data transformations as described in Section 7 - Data Transformation Policy and Guidance. This section describes the current provisions for CAC monitors and lays out the potential scope of using CACs in a revised network.

Performance Criteria

There are two types of performance criteria to consider. The first criteria to consider are the performance standards for acceptance of a method. These criteria are provided for in section 6 and are primarily based upon the goals for measurement uncertainty as developed in the data quality objective process for the $PM_{2.5}$ monitoring program. Since the CAC is not used for regulatory decision making the specific criteria for precision and bias at a site or network of sites will remain "goals" and not requirements. The second type of criteria are for on-going evaluation that the method is providing data of sufficient quality for its intended monitoring objective. These criteria are the same performance standards developed for measurement uncertainty in the $PM_{2.5}$ monitoring program and are also presented in section 6 of this document.

Testing Requirements

There are a number of testing requirements that need to be considered. These testing requirements are intended to be designed so that State and local agencies can readily implement a field testing program to pursue a CAC for use in their network. The table below identifies the suggested criteria and rationale for CACs:

Table 3-1 Test Specification for $PM_{2.5}$ CACs

Testing Requirement	Suggested Criteria for CACs	Rational for Criteria
Number of Test Sites	1 on a site by site basis or minimum of 2 for a network (see Table 3-2 below)	Need to demonstrate that the method can meet performance criteria at a specific site or multiple locations in a State or local network.
Number of FRMs per site for generating baseline data in testing	1 - However strongly suggest locating test sites at collocated FRM precision sites to assure control of FRMs and to have high sample completeness	Precision of FRM can be assumed from FRM network precision statistic
Number of Candidate Samplers	2 for first CAC site, 1 each for each additional site tested.	Need to have collocated candidate CACs in order to calculate measurement precision of the continuous method for at least one site in the network.
Number of hours to make a valid 24 hour sample for comparison to the FRM	18	75% completeness of the 24 hour period
Length of testing	All 4 seasons - however testing can begin and end at any point during the year	Need to assure that changes in aerosol or meteorology related to changes in season can meet performance requirements.
Number of data pairs - Primary Monitors, both the FRMs and the candidate CACs	90 per site with at least 20 per season See reference in section 7	Expected to be similar to 1 in 3 day sample frequency at 75% completeness for four seasons
Number of data pairs - Collocated FRMs	As found in network	Use existing collocated FRM precision sites
Number of data pairs - Collocated candidate CACs	- 60 sample pairs - At least 15 sample pairs per season	Based upon 90% confidence that the precision statistic is within 15% of the true precision. Since these are continuous methods may expect to have a substantially large data set.

Range of concentrations for siting	As found in the area of consideration.	Need to evaluate method under the conditions in which it will operate.
Range of concentrations for use in data set when determining performance of methods	May (but not required to) exclude values where the FRM concentration is below 6 ug/m^3. Exclusion of values due to low concentrations does not result in failure of completeness requirements	As concentration values approach 0, biases can appear large. By focusing on the values that are above 6 ug/m^3 estimates of the performance of the candidate methods are more stable.

Guidance for Developing Boundaries for Applicability of CAC

Section 8 of this document provides the detail for how the appropriate geographic size is determined for use of an approved CAC.

Number of test sites for Collocated Acceptable Continuous monitors

The number of test sites for CACs depends on a number of factors such as whether one site or a network of sites is being considered for approval of a CAC and the homogeneity of the aerosol across the area of consideration. At a minimum, 2 sites are to be tested to support a candidate CAC across a network. The following table details how many sites are to be tested assuming the aerosol is homogeneous across an area in which it is being tested:

Table 3-2 Test Site Specifications for $PM_{2.5}$ CACs

Geographical Area of Consideration for CAC	Number of Test Sites
One MSA	2
Multiple MSA's in the same air district or State	1 for each MSA up to the first 3 MSAs, plus at least 1 site in a rural county.
Multiple States	1 for each MSA up to the first 2 MSAs, plus at least 1 site in a rural county. For each additional State add 1 urban and 1 rural site.

Note: if the aerosol is expected to vary according to the guidance provided for in section 8, then apply test sites as if each State or air district were performing testing separately. This will ensure that for each type of aerosol encountered a minimum number of sites are tested.

Review Procedures

Since the monitoring objectives for CACs do not include direct comparison to the NAAQS, the approval procedures for use of a method should be streamlined. Thus the review procedures should be included in the annual network review that is submitted by the State, local or Tribal Agencies to the Region. The Region would work to determine that the performance criteria have been appropriately addressed and the continuous method is suitable for inclusion in the network. Since many agencies potentially seeking the CAC approach for relief from FRM sampling are expected to be substantially below that standard, the Regions should work towards approval of the CACs where they make sense and not prevent their approval if a specific goal is not met. For instance, one way for Regions to make a good decision on the approval of a CAC is to utilize the DQO tool that has been developed with inputs of a number of variables and see if the uncertainty around the NAAQS would be worse or better. If the goals for measurement uncertainty are ± 10% bias and 20% CV and the agency has a bias of 5% and CV of 23% with their continuous method, then the uncertainty around the NAAQS may actually be better.

Ongoing Evaluation of Method Performance

Since the CAC is not to be used for direct comparison to the NAAQS, the specific QA/QC requirements of the $PM_{2.5}$ quality system do not apply in a strict sense. However, since the data are to be used for a number of other important monitoring objectives the $PM_{2.5}$ quality system does apply in a qualitative sense. This means that agencies must develop appropriate measures to determine precision and bias estimates for the CAC monitors used in their network, but they are not held to specific numbers as if they were regulatory monitors. Additionally, the CACs should be appropriately addressed in the monitoring agencies Quality Assurance Project Plan (QAPP). Agencies should be evaluating the quality of their network on an ongoing basis and work to resolve problems as they are encountered.

Potential Use of CACs in $PM_{2.5}$ Monitoring Networks

The expected outcome of having a CAC approved for use at a site or in a monitoring network is that it can be used in combination with a limited number of FRMs as part of a “hybrid” network. Section 5 of this document lays out the detailed network design of a potentially revised network.

Section 4. Regional Equivalent Monitors

A provision to allow for Regional Equivalent Monitors (REMs) is proposed to enhance the network of PM continuous monitors. The basic premise of a REM is that when a $PM_{2.5}$ continuous method meets the precision and bias performance criteria identified in section 6 and the testing specification described below within the geographic area that it is used, then this method may be used anywhere in the network for which it is approved. The spatial extent of the approval of the method would be based upon a number of factors such as number and location of sites tested and homogeneity of the aerosol in the network. This flexibility is being considered since some methods are expected to work well in replicating FRM measurements across specific agencies networks across all seasons, but not in every network in the country. Approved REMs would be allowed to be used for attainment decisions as part of a "hybrid" network of $PM_{2.5}$ FRMs and continuous monitors as described in section 5 - Network Design. For implementation purposes REMs are different than the conventional Federal Equivalent Methods (FEMs) in that they are only for use in the specific geographic area of approval and a minimum network of FRMs must be retained for operation in each network in which they are used. When FRMs are collocated with REMs, the FRM is identified as the *Primary* monitor, meaning it is the monitor to be used for comparison the NAAQS at that site. REMs are different from Correlated Acceptable Continuous (CAC) monitors in that data from REMs are used for direct comparison to the NAAQS, while data from CACs are not. Since the data from REMs are used for comparison to the NAAQS, there is much more control on the approaches for data transformations, as described in section 7. This section describes the test specifications and approval process for REMs.

Performance Criteria

There are two types of performance criteria to consider. The first criteria to consider are the performance standards for acceptance of a method. These criteria are provided for in section 6 and are primarily based upon the goals for measurement uncertainty as developed in the data quality objective process for the $PM_{2.5}$ monitoring program. The second type of criteria are for on-going evaluation that the method is providing data of sufficient quality for its intended monitoring objective. These criteria are the same performance standards developed for measurement uncertainty in the $PM_{2.5}$ monitoring program and are also presented in section 6 of this document.

Testing Requirements

There are a number of testing requirements that need to be considered. These testing requirements are intended to be designed so that State and local agencies can readily implement a field testing program to pursue a REM for use in their network. The table below identifies the required criteria and rationale for REMs:

Table 4-1 Test Specification for $PM_{2.5}$ REMs

Testing Requirement	Suggested Criteria for REMs	Rational for Criteria
Number of Test Sites	Minimum of 2 (see Table 4-2 below)	Need to demonstrate that the method can meet performance criteria at multiple locations in a State or local agency network.
Number of FRMs per site for generating baseline data in testing	1- However strongly suggest locating test sites at collocated FRM precision sites to assure control of FRMs and to have high sample completeness	Precision of FRM can be assumed from FRM network precision statistic
Number of Candidate Samplers	2 for first REM test site, 1 for each additional site	Need to have collocated candidate REMs in order to calculate measurement precision of the continuous method for at least one site in the network.
Number of hours to make a valid 24 hour sample for comparison to the FRM	18 - valid hourly values within the midnight to midnight period.	75% completeness of the 24 hour period.
Length of testing	All 4 seasons - however testing can begin and end at any point during the year.	Need to assure that changes in aerosol or meteorology related to changes in season can meet performance requirements.
Number of data pairs - Primary Monitors, both the FRMs and the candidate REM	90 per site with at least 20 per season. See reference in section 7	Expected to be similar to 1 in 3 day sample frequency at 75% completeness for four seasons
Number of data pairs - Collocated FRMs	As found in network	Use existing collocated FRM precision sites
Number of data pairs - Collocated candidate REMs	- 60 sample pairs for the REM - At least 15 per season for the REM.	Based upon 90% confidence that the precision statistic is within 15% of the true precision. Since these are continuous methods may expect to have a substantially large data set.
Range of concentrations for siting	As found in the area of consideration.	Need to evaluate method under the conditions in which it will operate.
Range of concentrations for use in data set when determining performance of methods	May (but not required to) exclude values where the FRM concentration is below 6 ug/m^3. Exclusion of values due to low concentrations does not result in failure of completeness requirements	As concentration values approach 0, biases can appear large. By focusing on the values that are above 6 ug/m^3 estimates of the performance of the candidate methods are more stable.

Guidance for Developing Boundaries for Applicability of Regional Equivalent Monitors

Section 8 of this document provides the detail for how the appropriate geographic size is determined for use of an approved REM.

Number of Test Sites for Regional Equivalent Monitors

The number of test sites for REMs depends on a number of factors such as the area of consideration for approval of a REM and the homogeneity of the aerosol across the area of consideration. At a minimum, 2 sites are to be tested to support a candidate REM. The following table details how many sites are to be tested assuming the aerosol is homogeneous across an area in which it is being tested:

Table 4-2 Test Site Specification for $PM_{2.5}$ REMs

Example Geographical Area of Consideration for REM[8]	Number of Test Sites
One MSA	2
Multiple MSA's in the same air district or State	1 for each MSA up to the first 3 MSAs, plus at least 1 site in a rural county.
Multiple States	1 for each MSA up to the first 2 MSAs, plus at least 1 site in a rural county. For each additional State add 1 urban and 1 rural site.

Note: if the aerosol is expected to vary according to the guidance provided for in section 8, then apply test sites as if each State or air district were performing testing separately. This will ensure that for each type of aerosol encountered a minimum number of sites are tested.

Review Procedures

The approval of a "Regionally" equivalent monitor should follow the same process for review and approval of other federal equivalent methods. This process works through the Office of Research and Developments National Exposure Research Laboratory (NERL) Reference and Equivalency program. That program receives, reviews and provides feedback to vendors and other parties that have applied for equivalency. Once all the criteria have been appropriately addressed and the candidate method has been determined to meet the appropriate performance criteria the Reference and Equivalency program makes a recommendation that the method be approved as "equivalent". Once approved by EPA management as "equivalent" a notice is published in the Federal Register indicating

[8] The example presented at best reflects a minimum requirement. Definition of extent of regional applicability is addressed more completely in section 8 and is a topic requiring significant development.

that this status has been achieved. Any geographic limitations to a methods approval would also be included in this notice.

Ongoing Evaluation of Method Performance

Since the REM is to be used for NAAQS decision making all applicable elements of the $PM_{2.5}$ quality system are to be applied to its use. This means that REMs are to be collocated with both FRM and the same continuous method as the primary monitor as well as being subject to performance evaluation audits defined in Appendix A of Part 58. Additionally, the CACs should be appropriately addressed in the monitoring agencies Quality Assurance Project Plan (QAPP). If for three consecutive years the REM does not meet the DQOs and a examination of the data indicates that the uncertainty in decision errors is increasing, then the monitoring agency should - **NOT SURE WHAT THE CONSEQUENCE SHOULD BE.** Would like to have agencies work through a solution.

Potential Use of Regional Equivalent Monitors in $PM_{2.5}$ Monitoring Networks

The expected outcome of having a REM approved for use in a monitoring network is that it can be used in combination with a limited number of FRMs as part of a “hybrid” network. Section 5 of this document lays out the detailed network design of the a potentially revised network.

Section 5 - Network Design

Introduction:

The $PM_{2.5}$ monitoring program has been implemented with a heavy emphasis on Federal Reference Method (FRM) samplers in order to support comparing mass data to the National Ambient Air Quality Standards (NAAQS). Approximately 1143 (July 11, 2001 AIRS) monitoring sites in the United States are now operational with FRM samplers. The entire $PM_{2.5}$ network includes components for chemical speciation and advanced measurements (Attachment B). Only the FRM or Federal Equivalent Method (FEM) can be used for direct comparisons to the NAAQS. This plan proposes a more balanced hybrid network of filter based and continuous mass samplers, assuming that data analysts would incorporate filter based and continuous methods (seamlessly) when utilizing network data for broad scale spatial applications such as positive matrix factorization (PMF) and air quality model evaluation. This hybrid network would include a reduced number of existing FRM samplers for direct comparison to the NAAQS and continuous samplers that meet specified performance goals related to their ability to produce sound comparisons[9] to FRM data. Two approaches described in sections 3 and 4 for integrating continuous mass monitors are proposed to maximize flexibility for agencies; an expanded use of Correlated Acceptable Continuous Monitors (CAC), and Regional Equivalent Monitors (REMs). The CAC approach would enable agencies to address any monitoring objective, other than ***direct***[10] comparisons to NAAQS for attainment and non-attainment designations, while the REM approach would serve any objective.

There is an unknown amount of degraded data quality risk associated with moving from the current design based system to one relying on performance based specifications. Therefore, this hybrid network will maintain a core of FRMs to maintain an ability to quantify the relationship between FRMs and continuous samplers for continuity to both the historical record as well as ongoing and prospective use of continuous methods. The remaining network of FRMs might constitute 30% to 50% of the current network. A large network of continuous monitors meeting performance criteria would eventually be in place to improve the data base for: public reporting of Air Quality indices (AQI) and mapping through AIRNow; supporting health effects and exposure studies addressing short term exposures; evaluating air quality models and emission inventories, and supporting compliance needs related to direct comparisons with the NAAQS and delineating the spatial extent of attainment/nonattainment areas.

---

[9] Comparability between FRMs and continuous samplers is desired, based on the extensive FRM network available. This practical need also recognizes inherent differences between measurement principles of integrated and continuous methods and does not assume any one type of measurement best represents true atmospheric aerosols conditions. Further discussion on incommensurabilities between measurement systems is provided in section 7.

[10] Data from CACs would be expected to be incorporated in as yet undetermined weight-of-evidence analyses to define boundaries of non-attainment/attainment areas.

Minimum number of Federal Reference Method (FRM) Samplers

A separate but parallel effort is underway to better identify redundant monitoring for all pollutants as part of the National Monitoring Strategy (see section 10). This national strategy supports an investment in continuous PM monitors balanced by a divestment in $PM_{2.5}$ FRM sampling. Progress in enhancing PM continuous monitoring requires a burden reduction in FRM sampling. Currently, nearly 1100 FRM samplers operate across the United States, and an additional 200-300 IMPROVE and continuous samplers. The spatial richness this network should not be severely compromised; however, areas of redundancy are evident based on a variety of national and regional based assessments that illustrate broad expanse of homogeneous aerosol behavior. A reduction in required FRM samplers is possible using network assessment processes to determine effective numbers of samplers across regional and urban spatial scales. Nationally, we are suggesting that a minimum of 300-500 FRM/FEMs be retained to ensure consistency with the existing network, and provide the primary regulatory base of data. A total of approximately 600 equivalent samplers (including FRMs, FEMs and REMs) for direct comparisons to the NAAQS are recommended. The network size of approximately 600 is based on several data analyses. One analysis shows that the large spatial patterns in $PM_{2.5}$ are nearly identical whether using 300 or 1200 monitoring locations in an area that covers much of the eastern United States. A second analysis indicates that several urban areas are likely over-sampled by approximately 25-35%. Perhaps as many as 1000 (or more) $PM_{2.5}$ mass (FRM and continuous) sites nationally are needed for spatial characterization, but request that actual number of sites be a function State/local agency discretion as agencies must balance several competing monitoring priorities. Note, that this approach while increasing flexibility could have unintended negative consequences by accommodating too many diverse methods that do not relate well with each other. Agencies are encouraged to strive for consistency in deploying their continuous PM network and consider not only consistency of methods within an agency, but attempt to harmonize technology across regional areas.

Table 5-1 summarizes the applicability of each monitoring method category to the type of site in the network. Tables 5-2 and 5-3 include examples of revised network requirements for $PM_{2.5}$ samplers. Specific modifications to the $PM_{2.5}$ monitoring regulations are being addressed through a workgroup of state/local agency, Tribal nation, and EPA representatives (see section 11).

Table 5-1 PM Method Applicability

Method	Required Sites for NAAQS						
	< 80% of NAAQS	80% to 120% of NAAQS	>120% of NAAQS	Sites that are currently required but are not required in a future network.	Current Supplemental Sites	Background and Transport Sites	Speciation and IMPROVE
FRM/FEM	T	T	T	T	T	T	
REM	T • With 30% FRM collocation in network	T • With 30% FRM collocation in network	T • With 30% FRM collocation in network	T	T	T	
CAC	T • With 100% FRM Collocation in network FRM operates 1-6		T • With 100% FRM Collocation in network FRM operates 1-6	T	T	T •	
IMPROVE						T	T
Speciation						T •	T
Existing Continuous mass PM					T		

T The method category in the row is applicable for the monitoring objective in the column.

• This symbol indicates a change to the monitoring regulation is needed

Method Applicability Summary

FRM/FEM/REM - These methods can be used at a required site, regardless of the concentration; at any current or future supplemental sites; and at any background or transport sites. REMs would be required or have at least 30% collocation with FRMs when they are sited at required sites. The FRM would be the primary sampler when collocated with an REM.

CAC - This monitor could provide relief up to 3 new ways:

To convert a site from a filter based sampler to a CAC:

1.) At current supplemental, background, and transport sites CAC monitors may be used as the primary monitor. Collocation at these sites would follow the provisions of Appendix A; which is expected to be 15% collocation with the first collocated monitor being an FRM and the next one being of the same make and model as the CAC.

2.) The minimum number of required sites is to be reduced in Appendix D of Part 58. There is an expectation that there will be more sites operating than the minimum number required. For sites that are no longer required to be operated; but the agency still intends to operate the site to meet other monitoring objectives, the agency may choose to operate the site with a CAC. Appendix A collocation requirements would apply for these CACs. Moreover, it is feasible that revised monitoring regulations may require a similar total number of monitors currently required or operating (i.e., 850 to 1100) with a subset required to have reference/equivalent status and the remainder being satisfied by an equivalent/reference or CAC designation.

To provide additional sample frequency relief:

3.) For required reference/equivalent sites (current or future) that are either substantially above or below the NAAQS, the CAC may be operated to provide a signal of PM provided it is collocated with a FRM operating on at least a 1 in 6 day schedule. The FRM maintains the status as the primary monitor.

The conventional sample frequency relief for a CAC would still apply:

A FRM site that is required to operate daily may have its sample frequency reduced to 1 in 3 provided it is collocated with a CAC regardless of the concentration or NAAQS status.

Table 5-2 Network Design Criteria for $PM_{2.5}$ Required SLAMS

Network Design Criteria	Current Network	Example Revised Network
Required minimum number of sites at State and local Air Monitoring Stations (SLAMS)	Approximately 850* * 100 are for background and can use IMPROVE samplers	Assuming ~600 sites are reasonable we envision a hybrid network of FRM and continuous methods meeting acceptable performance standards. A minimum of ~ 30% of each monitoring agencies future network would be required to remain as FRMs.
Scale of representativeness - Annual Average	Neighborhood or Urban Scale with FRM or FEM	Neighborhood or Urban Scale with FRM/FEM or hybrid network of FRM and continuous monitors meeting performance based criteria
Scale of representativeness - Daily Average	Micro, Neighborhood, Urban Scale	Micro, Neighborhood, Urban Scale. For sites that are expected to only have a violation of the daily standard, but not the annual average, site with a FRM/FEM. Collocated with a Continuous monitor, if needed
Community Monitoring Zones	Optional	Consider deleting this provision since no agencies are using

Section 6 - Performance Standards for Continuous Monitoring

Introduction

The current paradigm for a $PM_{2.5}$ continuous monitor to receive an federal equivalent monitor designation requires field tests at multiple locations over an entire year with the field data being able to meet conservative test specifications that include slope, intercept, and R^2. If a candidate method meets all the criteria, then it receives an “equivalency” designation for use anywhere in the national network, even if it has not been tested in all areas. The assumption is that the method will perform as intended in all areas if it meets strict test specifications at a limited number of sites covering a range of environmental and aerosol conditions. Also, once a method receives an equivalency designation, no additional field tests are required to ensure that the equivalency holds through time.

The approach presented in this section is to link the testing requirements and the ongoing performance requirements to the Data Quality Objectives (DQOs). The DQOs provide a level of uncertainty in the data that is acceptable, given the intended use of the data. Methods that meet or exceed the DQOs can be used in the networks in which they were tested, provided they continue to meet the DQOs through time.

The PM2.5 Data Quality Objective was developed for comparison of values around the 3-year annual average NAAQS since it was found to be the more restrictive standard (i.e. any violation of the daily standard would in almost all cases be in violation of the annual standard). Therefore, use of the DQO for continuous monitoring, at present, is limited to comparisons against this objective. OAQPS is pursuing development of a DQO controlling data quality around the daily standard.

Background and Rationale

$PM_{2.5}$ DQO Process

DQOs are qualitative and quantitative statements that clarify the monitoring objectives, define the appropriate type of data, and specify the tolerable levels of potential decision errors that will be used to determine the quality and quantity of data needed to support decisions (i.e., NAAQS comparisons). A more complete description of the $PM_{2.5}$ DQOs and how they were derived is presented in Attachment B.

DQOs for $PM_{2.5}$ were developed during the months from April to July of 1997. A number of assumptions were made in order to generate realistic error rates. Table 6-1 provides a listing of these assumptions. In 2001, EPA reassessed the assumptions underlying the 1997 DQOs. In almost all cases, the assumptions made in the 1997 process held true in the 2001 evaluation.

The $PM_{2.5}$ DQOs were generated using conservative but realistic assumptions. For example, the DQOs were generated assuming a sampling frequency of every 6 days with 75% completeness. This is the lowest sampling frequency allowed in the Code of Federal Regulation. A 95% confidence limit around the annual mean at this sampling frequency would be "wider" than a 95% confidence limit for an every day sampling frequency at 90% completeness. In all cases, the assumptions in Table 6-1 are close to the extremes of realistic and allowable data. Assumptions in bold are variables that will be discussed later in this section.

Table 6-1 2001 DQO Assumptions

1.	**Bias is -10% or + 10%**
2.	**Precision is 10%**
3.	Annual NAAQS is controlling standard
4.	No spatial uncertainty and each monitor stands on its own (no spatial averaging)
5.	**1 in 6 sampling with 75% completeness (144 days)**
6.	3-year annual average is truth, (every day sampling and 100% comp.) up to bias and measurement variability
7.	**Lognormal distribution for population variability, 80% CV**
8.	Normal distribution for measurement uncertainty
9.	**Seasonal ratio (ratio of avg conc for highest season to lowest season) = 5.3**
10.	**No auto correlation in daily concentrations**
11.	Bias and measurement variability (precision) applies to entire 3 years
12.	Type I and type II decision errors set to 5%

Figure 6.1 provides the power curve based on the 2001 assumptions shown in Table 6-1. A power curve is an easy way to display the potential of decision errors based upon the choice of various assumptions that affect data uncertainty. The gray zone is the range of concentrations for which the decision errors are larger than the desired rate of 5%.

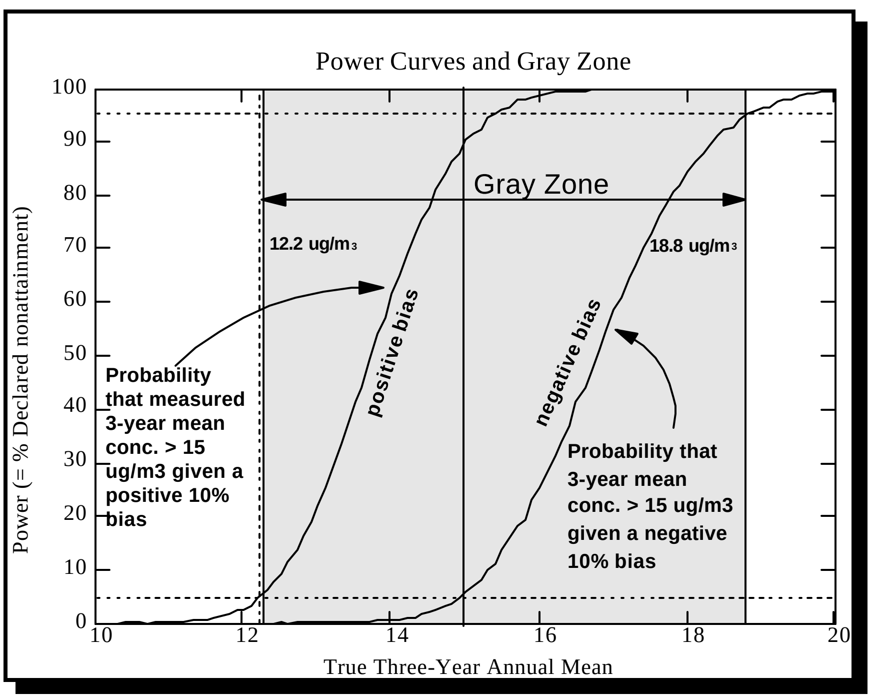


Figure 6.1 Power curve based on 2001 assumptions

Based on the 2001 assumptions, the gray zone is 12.2 to 18.8 µg/m^3. This means that if all the 2001 assumptions hold, the decision maker has a 5% chance of observing a 3-year mean concentration that is greater than 15 µg/m^3 even though the true mean concentration is 12.2 µg/m^3. As has been mentioned, the 2001 assumptions are realistic but conservative. For example the CY00 $PM_{2.5}$ QA Report demonstrates that the precision and bias estimates at a national level are well within the DQOs. Assumptions that are "better" than those listed in Table 6-1 will tend to decrease the width of the gray zone. Figure 6.2 provides an example of the power curve/gray zone changes for a simple change in sampling frequency from 1 in 6 day (green/solid) to 1 in 3 day (blue/dots) to every day (red/dashed); all the other 2001 assumptions remain the same. Higher sampling frequencies result in narrower gray zones, meaning that decision errors are reduced.

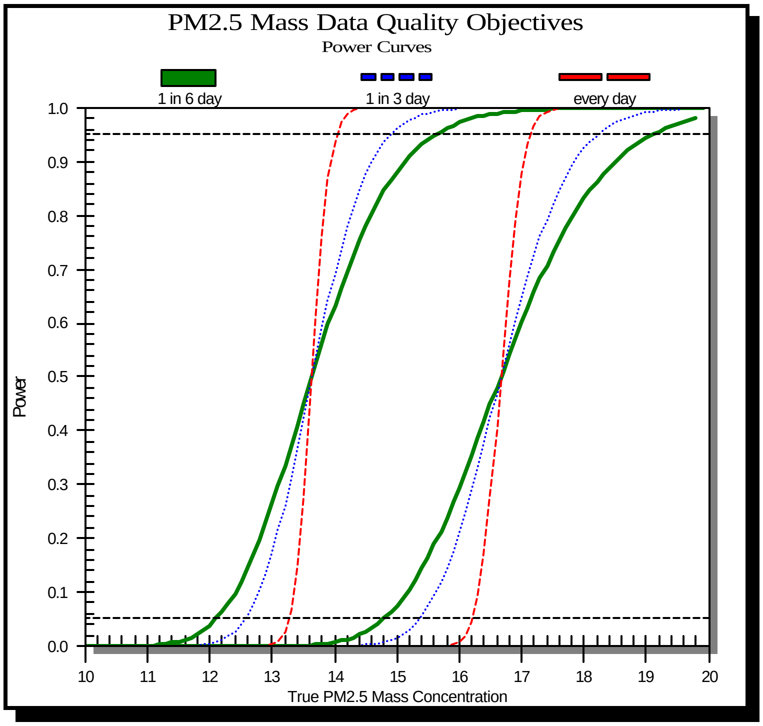


Figure 6.2. Power curve changes due to changes in sampling frequency

Because there is potential for the assumptions to vary, OAQPS commissioned the development of a software tool to help Headquarters and State, local and Tribal organizations determine the potential for decision errors based on assumptions relevant for sites within their network. Figure 6.2 is generated using this tool and allows for multiple scenarios (power curves) to be reviewed on one table. The assumptions listed in bold in Table 6-1 can be changed to suit a particular network. This tool is being finalized and should be available by December, 2001. Furthermore, the tool will be useful for making decisions about the acceptability of REMs or CACs within a network.

The DQO evaluation showed that sampling frequency, population variability (assumed to be lognormally distributed with a CV of 80%), and measurement bias play a significant role in the width of the gray zone. Measurement precision did not have a significant effect on the gray zone which suggests more imprecision could be tolerated with little effect on decision errors (when evaluating an annual mean developed with 3 years of data).

CONCLUSIONS FROM DQO TOOL

The $PM_{2.5}$ mass DQOs were developed for making good decisions about the 3-year average of annual means, since it was assumed that the annual standard was the controlling standard. In particular, they were developed to evaluate the chance of concluding an average concentration was above 15 :g/m^3, when in truth it was not, and the chance of concluding an average concentration was below 15 :g/m^3, when in truth it was not. Due to the number of measurements that go into the 3-year average of annual means (at least 144), it is easy to see why measurement precision does not have a large influence on the size of the gray zone of the power curve. If, however, the DQO tool displayed the power curves for the daily standard (the 3-year average of the annual 98^{th} percentiles), it is likely that measurement precision would be important for the decision errors, since the extremes of distributions are less robust than the centers. Recent evaluations of the continuous monitors have shown precision estimates comparable to the FRMs.

Data uses that involve no averaging, such as real-time reporting, are even more sensitive to measurement imprecision. Thus, caution should be exercised in drawing conclusions from the DQO power-curve tool. The tool has been designed for specific data uses, namely, evaluating decision errors associated with the $PM_{2.5}$ standards and is based on specific assumptions. If the assumptions are not appropriate or if the data use is different than comparison to the standards, the power curves and gray zones likely do not reflect the true decision errors.

The DQO tool is being enhanced to present both forms of the standard to ensure that decision errors are acceptable for both standards. This tool will be available for monitoring agency use in CY02. In addition, we hope to be able to develop a report in AIRS that would automatically generate the DQO assumptions listed in Table 6-1 by a variety of data aggregation schemes (i.e., reporting organization, by a collection of sites etc.)

Acceptable Performance Criteria for Continuous Monitoring Using Power Curve Tool

Figure 6.1 set up the most extreme case that is tolerated in the $PM_{2.5}$ DQO, based on the assumptions in Table 6-1. The DQOs have associated with them a gray zone which will be used to develop acceptable bounds for the quality of the data required (REM) or recommended (CAC) for the continuous monitoring program. An important note is that the data for which the quality is being evaluated is not the raw data produced by the continuous monitors. Rather it is the continuous data that has been transformed, using a statistical model, to be FRM like.

Subsequent discussions will include the terminology of “simple” transformations and “complex” transformations. For this document, the definition of a simple transform is one in which the FRM data are the response variable (also called the dependent variable) and the only explanatory variable (also called the independent variable) allowed is the continuous data, summarized to the daily level. Thus, simple transforms are of the form Y = a + bX, where Y is the FRM data and X is the continuous data. The transformation is still considered to be simple if the natural logarithms of X and Y are used instead of the raw data. The definition of a complex transformation is one in which the FRM data are the response variable and any variable is included as an explanatory variable. Minimally, the continuous data are an explanatory variable. Again, complex models may be based on the raw data from the monitors or based on their natural logarithms.

The following table describes some of the fundamental differences between a REM and CAC, as pertains to data use, allowable transformations, and data quality requirements or goals.

REM	CAC
Assumptions: Will be used in comparison to NAAQS Must have FRMs in Network Can only include simple transformations Must meet **1-3 day** DQO (gray zone) but specifically meet 10% bias DQO	***Assumptions:*** Will not be used in comparison to NAAQS Must have FRMs in Network Can include complex transforms Should meet **1-6 day** DQO (gray zone) but specifically meet 10% bias DQO

Developing performance criteria using the power curve tool is a multi-step process. The first step is to collect information from the CAC/REM network. The second step is to develop a transformation that produces FRM-like data from the CAC/REM (details of which are provided in Section 7). The third step is to determine the spatial extent for which the transformation is appropriate (details of which are provided in Section 9). The fourth step is to determine reasonable values for the highlighted parameters in Table 6-1. The values should be reflective of the entire spatial extent of the CAC or REM network being evaluated. The last step is to use the DQO software tool to determine the gray zone that results from the values from the previous step. If the bias is within -10% and +10% and the gray zone is within 12.7 and 18.1 :g/m^3 (the gray zone for an FRM that operates every third day), then the continuous sampler meets the requirements for being a REM. If the bias is within -10% and +10% and the gray zone is within 12.2 and 18.8 :g/m^3 (the gray zone for an FRM that operates every sixth day), then the continuous sampler meets the goals for being a CAC.

Figure 6.3 provides an example of the power curve for a 3-year mean based on the following data quality input parameters

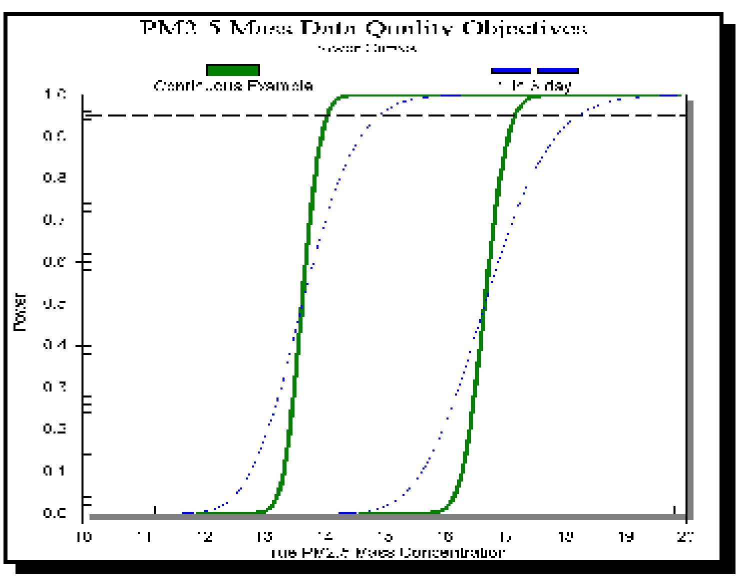


Figure 6.3. Example continuous monitoring network power curve relative to 1 in 3 day power curve.

< bias 10%
< completeness 75%
< sampling frequency every day
< measurement CV 30%
< population CV 80%
< Seasonal ratio 5.3

The resultant gray zone is 13.2 µg/m^3 (lower left line green solid) and 17.1 µg/m^3 (upper right line green solid which is within the 1-3 day DQO of 12.7 (lower left blue dashed) and 18.1 (upper right blue dashed). Therefore, this example continuous monitoring network could be considered acceptable for CAC or REM designation.

Simplified Performance Criteria for Continuous Monitoring

Organizations may use the DQO process described above to determine levels of measurement imprecision that can be tolerated but still provide data of a quality to support decisions about comparison to the NAAQS. For organizations not interested in using the DQO tool to develop gray zones applicable to specific areas, the DQOs are set to 20% CV and bias within -10% and +10%. REMs are required to meet these objectives whereas it is highly recommended that CACs meet these objectives.

Summary of Performance Criteria for $PM_{2.5}$ Methods

When discussing performance criteria, it's important to clarify the difference between acceptance of a method in the designation process and the on-going performance based goals. The acceptance of a method in the designation process is associated with the Reference and Equivalency program defined in 40 CFR Part 53. This process is purposely strict in order to assure the quality of data when subsequently designated methods are used throughout the country. Table 6-2 summarizes each category of existing and potentially revised methods with criteria for acceptance of the method and criteria for the on-going evaluation of the performance of that method.

Table 6-2. Performance Specifications for PM2.5 Methods

Category of Method	Requirements for Acceptance of Method	Existing Performance Goal for Acceptable Measurement Uncertainty	Future Performance Goal for Acceptable Measurement Uncertainty
FRM	Many design and performance criteria. Precision for field testing: < 2 $\mu g/m^3$ when concentration is <40 $\mu g/m^3$ (24 hour sample) or <30 $\mu g/m^3$ (48 hour sample); Rpj <5% for concentration > 40 $\mu g/m^3$ (24 hour sample) or >30 $\mu g/m^3$ (48 hour sample).	10% coefficient of variation (CV) for total precision and +/- 10 percent for total bias.	No Revision
FEM	Across a limited number of field test sites depending on class of equivalency: Slope of 1+/- 0.05 Intercept of 0 +/- 1 µg R^2 $ 0.97	10% coefficient of variation (CV) for total precision and +/- 10 percent for total bias.	No Revision.

REM	Within each network that is being considered: 20 % coefficient of variation (CV) for total precision and +/- 10 percent for total bias.	NA	Utilize 1 in 3 day DQO/Powercurve or simplified approach of 20% coefficient of variation (CV) for total precision and +/- 10 percent for total bias.
CAC	Within each network that is being considered: 20 % coefficient of variation (CV) for total precision and +/- 10 percent for total bias. (Goal, not requirement.)	NA	Utilize 1 in 6 day DQO/powercurve or simplified 20% coefficient of variation (CV) for total precision and +/- 10 percent for total bias. (Goal, not requirement.)

Section 7. Data Transformation Policy and Guidance

Variations in $PM_{2.5}$ measurements attributed to methodological differences should be minimized to support consistent data analysis across temporal and spatial regimes. For example, it would be erroneous to infer that 20% of the $PM_{2.5}$ measured in an urban area is due to local sources based on a comparison of the concentrations measured by monitors in an urban area to concentrations from upwind sites, if the instrumentation at the upwind sites are biased low by 20%, relative to the instrumentation used in the urban area. Realistically, $PM_{2.5}$ measurements should "look" like measurements taken by an FRM. This is because of the richness of the available FRM data base and due to the difficulty in ascribing a "reference" check for aerosol measurements.

Non-FRM samplers generally operate at a higher temporal resolution than FRMs and many will operate where there is no FRM, thus helping to fill spatial gaps in the FRM network. However, to be able to use data from multiple types of $PM_{2.5}$ mass monitoring networks (FRM, non-FRM) in the same analysis, the data must be comparable. Comparable means that if the various samplers were spatially and temporally interchanged, approximately the same concentrations would be measured. To achieve comparability, it is possible to transform, using statistical models, non-FRM data to look like FRM data or vice versa. Due to the interest in FRM-like concentration surfaces, the remainder of this section will only address transforming data from non-FRM samplers to produce FRM-like measurements.

Due to the inherent differences in measurement principles between FRM and PM continuous monitors there may be biases between the measurements obtained from an FRM and continuous monitor. If the bias is consistent through time and across space, a standardized correction factor could be used to produce FRM-like measurements from the continuous monitors. However, since mass concentration and composition and environmental conditions vary, a standard correction may not be practical on a national scale but may be achievable on a more regional scale. This section provides information about the development of transformations to produce FRM-like measurements from continuous measurements.

Based on preliminary analyses summarized in Section 2, developing a statistical model to relate concentrations from continuous samplers (predominantly TEOMs) to FRM samplers is achievable, although the complexity of the model varies by location and may vary through time. The complexity likely is a function of the stability of the composition of the aerosol, the stability of the meteorology (temperature and humidity), and the continuous monitoring methodology. The following guidance for developing transformations is based on the experience gained in analyzing the limited collocated FRM/continuous database to date. The database is limited due to temporal representativeness (at best 2 years since the FRM network was deployed in 1999), spatial representativeness (continuous samplers have been and continue to be deployed predominantly in large urban areas), and non-FRM sampling techniques. The database is predominantly based on data reported to AIRS. Prior to 2000, it was not possible to determine whether the data from a continuous monitor was reported after being adjusted by "correction" factors. Beginning in 2000, AIRS method codes were expanded so that it would be possible to determine whether correction factors had been applied, although it is not possible to specify the form or parameter estimates of the adjustment. These new method codes appear not to be accurate for all sites, as seen in Section 2, making it a further challenge to determine appropriate transformations.

A balance between forcing a particular measurement principle to mimic another (i.e., the FRM) is a significant complication that must be recognized in this task. The practical needs for data analysts demand some level of comparability. However, there is intrinsic value in the very differences that emerge between measurement systems due to the complex character of aerosols. The intention clearly is not to define the FRM as truth, but rather to recognize the practicality of the existing network. These considerations of basic measurement principles are embodied in this transformation guidance. Where relationships between two measurement systems exhibit simple linear and constant character, one can probably assume the difference in measurement approach does not result in a significantly different indicator of ambient aerosol. Such simple relationships are the foundation for accommodating REMs that can be compared to the NAAQS. On the other hand, more complex relationships between a candidate system and the FRM suggest that a significantly different aerosol property is being accounted for (likely varies over time or space) in one system relative to the other. This does not mean one system is superior to the other, but reasonable judgement suggests a limit to forcing a system to mimic the FRM for regulatory use, but to accommodate the system for other data uses within the limits of data comparability guidelines. This latter approach reflects the concept underlying the expanded use of CACs.

The guidance on transformations will be broken into two sections, one for the CAC and one for the REM. The guidance for acceptable transformations for REMs will be very strict and limited to simple transformation models. Acceptable transformations for CACs will be less strict. For either case, recall that the performance criteria presented in Section 6 is based on the transformed continuous measurements. Note that if the performance criteria are met with the raw continuous measurements, then no transformation is required. That is, transformations need not always be developed.

Regardless of whether the continuous sampler is a CAC or REM, measurements should be reported to AIRS. Given that data users might not understand the differences in the sampling methodologies, it is recommended that the data be entered AFTER applying a transformation to produce FRM-like measurements. However, it will be important for other data uses to know what transformations have been applied. EPA will be investigating possible ways to include the transformation information in AIRS so that it will be possible to “back out” the transformation and have the original, non-FRM measurements.

Transformation Guidance for CAC

Even though the data from a CAC will not be used for direct comparison to the NAAQS, they should meet the DQOs, as described in Section 6. Although this is not a requirement, it is strongly recommended for comparability of measurements across the network. The data used for evaluation in the DQO process are those that have been transformed to be FRM-like; that is, the DQOs are not necessarily based on the raw data from the non-FRMs. This section describes the process for developing the transformations for CACs. The rationale and details for the selection of many of these criteria are included in the EPA document *Reporting an Air Quality Index (AQI) Using Continuous $PM_{2.5}$ Data: Data Quality Objectives (DQOs) and Model Development for Relating Federal Reference Method (FRM) and Continuous $PM_{2.5}$ Measurements (Attachment C).*

Step 1. Create daily non-FRM measurements. If the non-FRM data are collected more frequently than daily, the sub-daily intervals should be averaged before comparing to the FRM data. At least 75% of the sub-daily intervals should be valid to consider the average to be valid. Also, the sub-daily intervals to be averaged should be those that most closely span midnight to midnight, the operating interval of the FRMs.

Step 2. Determine if there are sufficient data to develop statistical model. The model to relate the non-FRM and FRM data should be based on data from all four seasons and have at least 104 valid pairs of data, approximately evenly distributed through each season. It is recommended that each season have at least 20 valid pairs. If there are not more than 100 valid pairs approximately evenly distributed through the seasons, it is recommended that additional data be collected. The 100 pairs need not be from only one year.

3. Develop a statistical model. The statistical model relating the non-FRM and FRM data should have the FRM data as the response variable (also called the dependent variable) and minimally must include the non-FRM measurements from *Step 1* as an explanatory (independent variable). The number and type of explanatory variable allowed in unlimited. The model can be based on the data as is or can be based on the natural logarithms of the data. The final R^2 between the measured and predicted FRM measurements should be 0.80 or greater.

4. Spatial extent for use of one transformation. Section 8 describes the process for determining the area within which one transformation may be used for all of the continuous samplers, regardless of whether the continuous sampler has been previously collocated with an FRM.

5. On-going evaluation of transformation and its spatial extent. The statistical model should be revisited every 3 years, or more frequently if there is reason to believe a change in the relationship between the non-FRM and FRM may have occurred. Possible reasons for such changes include, but are not limited to, a change in sampling methodology, change in aerosol composition due to control strategies, or different meteorological regimes than what was observed during the development of the statistical model. If a new statistical model is more appropriate, that model should be used from that date forward. That is, one model would be used up to one date and the next model would be used for subsequent dates.

Transformation Guidance for REM

The data from REMs must meet the DQOs, as described in Section 6. The data used for evaluation in the DQO process are those that have been transformed to be FRM-like; that is, the DQOs are not necessarily based on the raw data from the non-FRMs. This section describes the requirements for the transformations. Because the data are intended to be used for NAAQS comparisons, the allowable statistical models and parameter estimation will be explicitly defined. The reason for this specificity is to ensure that two independent data analysts will produce the same transformation and hence will produce the same FRM-like concentrations. Most of the details of this guidance are unknown at this time, due to limited data, but the issues that need to be addressed and a time line for addressing them is included. The guidance components are as follows.

Step 1. Manipulation of non-FRM data, prior to development of statistical model. This section will detail how data are aggregated to produce a daily number to be used to compare to the FRM. Issues to address will include handling of missing data, producing averaging periods that are approximately midnight to midnight, and handling of negative or zero concentrations prior to aggregation. Data completeness will also be addressed. Likely, at least 75% of the sub-daily intervals should be valid to consider the average to be valid.

Step 2. Identification of pairs to use in development of transformation model. This section will address the number of required valid pairs, temporal representativeness of those pairs (e.g., whether highest and lowest seasons are sufficient or if every season must be represented), range of concentrations spanned by the pairs (need a good spread so that the model is appropriate and can be used for prediction through wide range of concentrations), handling of negative or zero concentrations, handling of concentrations less than some cutoff value (e.g., minimum detection limit), identifying and handling influential pairs.

Step 3. Development of statistical model. This section will detail how the statistical model relating FRM and aggregate non-FRM collocated data is to be developed. The only model allowed will be one for which the aggregate non-FRM data is the only explanatory variable and the FRM data is the response variable, that is, only a slope and intercept will need to be estimated. Issues include whether the raw or natural logarithm of the raw data are to be modeled, the required R^2 between the measured and predicted FRM measurements, the equations for estimating the slope, intercept, and R^2 especially if seasons are not equally represented in the data set (that is, should the estimates be weighted).

Step 4. Inferences to be drawn from the statistical model. This section will discuss how to determine the spatial representativeness of the model (what is the area that can use the same transformation, which will be discussed in Section 8) and the temporal representativeness of the model (for how long is the model valid). When a transform is found to be no longer appropriate, what is done with the previously transformed data? Is the old transformed used until up to one date and then the new transform used for subsequent dates?

Step 5. On-going evaluation of statistical model and its spatial extent. At least 30% (rounding up) of the non-FRM sites must be permanently collocated with FRMs to provide the data needed to evaluate regularly the reasonableness and consistency of the transforms. The collocated sites should be distributed to represent different composition and meteorological regimes. Issues to cover include the frequency at which the transformation is formally evaluated to determine whether it is still appropriate. For example, it would not be practical to have a transformation that is changed every month or quarter, but the transformation should be reviewed at some frequency.

Due to the numerous issues to developing statistical models, EPA will establish a panel to recommend solutions to the various issues listed above. The panel will be comprised of people conversant in statistics, ambient air monitoring, and air quality management. Solutions to these issues and final guidance on the development of transformations for regulatory data use are expected to be completed by the end of calendar year 2003.

Section 8. Defining Regional Applicability

The basic relationship between a continuous monitor and an FRM should be similar throughout a given "region" of application, especially with respect to bias. This reasoning is the foundation for the new regional equivalent method approach which assumes consistent monitor behavior can be achieved within a "region" despite inconsistencies nationally. Sections 3 and 4 provided example requirements (2 sites per MSA) for demonstrating consistency. The determination of regional applicability should be based first on technical considerations related to the homogeneity of aerosol composition and meteorology. This section addresses candidate approaches to determine regional applicability, and is intended to raise the understanding of this topic for further development of applications guidance.

Operationally, only one transformation model would be applied within the region of consideration. Determining the region in which the use of one transform is appropriate, meaning that all the sites within the region will meet the bias and precision requirements (REM) or goals (CAC), can be approached in two ways. One approach is to establish regions a priori where the regions explicitly cover specific land masses in the United States. For example, regions may be the interior southeast, the east coast, Florida, the industrial belt, the Midwest, the western coast, the arid southwest, Alaska, the Rocky Mountain states, and the humid northwest coastal area. The testing requirements for a candidate method would have to be met throughout one or more of these previously established regions. If one of the sites does not meet the testing requirements, then the method can not be used within that region. Such an approach implies knowledge about areas in which a particular type of continuous methodology and the FRMs have similar relationships. As shown in Section 2, knowledge based on the analysis of ambient measurements does not currently exist due to lack of data, especially data from emerging continuous monitoring methodologies. However, as more ambient measurements are collected for the various continuous monitoring methodologies, environmental conditions, and particulate composition and size distributions, such regions may become more clearly defined. ***This approach could be acted on by establishing a panel of experts charged with developing these regions.*** EPA and other organizations (monitoring agencies, Tribal nations, Regional Planning Organizations, RPO's) would address logistical and administrative complications associated with multiple monitoring organizations operating in a defined "region".

A second approach is to allow any size and shape of region. The State/local/tribe, RPO, or vendor interested in using a particular type of continuous instrument would specify the boundary of the region and then follow the testing requirements or goals to prove whether one transformation would be adequate for the entire region. That is, the domain of the region is flexible. However, once the testing has been completed for a specific domain, the domain remains fixed until on-going evaluations indicate the performance criteria are no longer being met throughout the region.

Both approaches will be pursued. It will be strongly encouraged that potential continuous monitoring methodologies be deployed at a core set of sites where the data from these sites will help to determine potential regions for the first approach. Until there are sufficient data to determine appropriate regions, the second approach will be used.

Definition of Regionality of Transformation

For a specific type of continuous monitoring methodology, given a sufficiently dense monitoring network of these monitors collocated with FRMs, it would be possible to develop a surface of the bias between the two types of instruments. In some places the bias might be small while in other places the bias might be large. In some places the bias might be negative and in others, it might be positive. Hopefully, the surface of biases would be smooth, that is, it would gradually change from one location to the next. Given such a smooth surface, it would be possible to produce FRM-like measurements at any location, even if there were no collocated FRM.

A difficulty with this construct of a surface of biases is that there is not a sufficiently dense network with which to build a surface for any large geographical area, especially for each type of continuous monitoring methodology. However, understanding this surface is the basis for being able to know sizes of regions. Collection of data with which to build such a surface is an important step to understanding regionality and is described below.

A surface of biases implies that the transformation to generate FRM-like measurements from continuous data would vary from site to site. Implementing site-specific transformations likely would prove to be intractable for a large number of sites, especially if the transformation is considered to be part of the method. One way around this problem is to use one transformation over an area where the biases are "similar." Specifically, the definition of the regionality of a transformation is that geographical area in which it is possible to use one statistical model to estimate FRM-like measurements and those FRM-like measurements meet the performance criteria specified in Section 6. Determining regions for which biases are "similar" also hinges on a dense data base of collocated FRMs and continuous instruments.

Data Collection to Support Definition of Regionality of Transformations

At least 100 sites of collocated continuous monitors and FRMs will be established as part of a National Core (NCore) network. These collocated sites will provide the data necessary to understand and monitor the temporal and spatial relationships between FRMs and continuous samplers. Characteristics of the sites include: (1) FRMs should operate at least every third day, (2) monitors should operate year-round and every year, (3) speciation trends sites are ideal given that the speciated data may help better understand the relationships, and (4) sites upwind of the speciation trends sites are also ideal, as the upwind sites likely have different compositions due to urban/rural gradients. The database generated by these sites will be regularly analyzed to determine if and how the FRM-continuous relationships vary spatially and temporally and how those relationships may change over time as compositions change due to implemented control strategies.

Until such time that a priori regions are defined, the regions may be any size and shape and the following guidance is applicable.

Regionality of Transformations for CAC

If the data from the non-FRMs are intended to be used for non-regulatory purposes, it is important that the data be comparable to the data produced by FRMs. However, since the data will not be used for direct comparison to the NAAQS, there is more flexibility in determining the regions within which one transformation is applicable.

Step 1. Develop transformations for each collocated site within the region of interest, based on the guidance provided in Section 7.

Step 2. Determine whether the transformations are statistically equivalent. For the sites that are equivalent, pool their data together to estimate one transformation. This one relationship should be used at each of the sites that was considered equivalent and may be used at other continuous sites for which there is no collocated FRM, provided that the sites operate the same type of non-FRM sampler using the same standard operating procedures, have similar chemical composition, and are exposed to similar meteorology. For examples, it would be inappropriate to apply a relationship established at a site running a TEOM to a site running a BAM, to apply a relationship established at a population-oriented site without any nearby sources to a site impacted by a large local source, or to apply a relationship established at an inland site to a coastal site. Sites that are not statistically equivalent to others should be considered unique, meaning that the transformation for the site should not be applied to any other site.

Step 3. On-going evaluation. It is recommended that at least 10% of the non-FRM sites be collocated with FRMs for at least 1 year of every 3 years and that the regionality be re-evaluated every 3 years. This recommended level of collocation on a permanent basis generally is met or exceeded in current networks.

Regionality of Transformations for REM

Following the approaches for CAC and REM discussed in Sections 3-7, the approval process for regional applicability for REMs would incorporate an as yet undetermined independent review procedure and more formalized demonstration of meeting performance and test requirements. The development of a review panel or board was raised above, and such an approach might be necessary given the probability of several unique cases and the desire to maintain equity in approval nationally.

Alternatively, due to the lack of understanding of the regionality of a relationship between data produced by FRMs and non-FRMs, the size of the region within which a continuous monitor can be considered for equivalency will be no larger than a site. As the data from the core sites becomes available and some understanding of the relationships grows, the size of potential regions will be reconsidered. Given that the continuous monitoring technology is changing and as a result few locations have at least a year of collocated measurements collected using the most-current SOPs, it is premature to propose an approach for using one *transformation over an area larger than an site. Too little is known about the potential gradients in the bias surface at this time.*

Section 9. Monitoring Methods Guidance and Support

Despite a substantial allocation of resources in overall PM monitoring implementation, very little methods development work has been performed in the area of $PM_{2.5}$ continuous monitors. This lack of development combined with requirements for lengthy field testing in multiple sites and high statistical correlations for designation as a $PM_{2.5}$ Federal Equivalent Method (FEM) have resulted in no applications for designation of continuous $PM_{2.5}$ monitors as FEM.

Introduction:

During the planning stages of the $PM_{2.5}$ monitoring program there was little emphasis on development of PM continuous methods by EPA. There were no nationally coordinated field testing programs to assess the usefulness of continuous methods over a variety of locations and aerosols. A guidance Document was written in 1998 compiling the available field testing on PM continuous methods to date; however, this document offered little insight on planning a long term strategy of using $PM_{2.5}$ continuous methods for regulatory purposes. Additionally, EPA never actually proposed or promulgated Class III equivalency criteria that would provide the testing requirements for $PM_{2.5}$ continuous methods. Since no criteria have ever been proposed there has never been an opportunity for the various stakeholders in the monitoring community to provide comments on the usefulness of the Class III equivalency testing criteria. There is an expectation that the equivalency criteria for Class III designations would be at least as strict as the Class II criteria. But since these criteria have never been published by EPA there is no clear path for acceptance of $PM_{2.5}$ continuous methods. Without EPA directly involved in developing PM continuous methods, vendors have been left to pursue improvements on their own. While some vendors have been successful at improving their methods by working directly with the States, these methods have not been appropriately tested on a national scale. For instance, the California Air Resources Board (CARB) has been working with the Met One Beta Attenuation Monitor over the last few years, yet very little data exists on this method collocated with FRMs in any east coast States. Also, Rupprecht & Patashnick Company have commercialized the Sample Equilibration System (SES) as an add on to the TEOM PM continuous monitor to allow for operation of this instrument at lower temperatures; however, little information is known about the long-term usefulness of the SES. The result of all of this is that there are no designated equivalent methods for $PM_{2.5}$ continuous monitors. Also, little information is available in the form of peer reviewed field studies over a variety of methods and locations. Despite all these issues there is still a great deal of information to glean from monitoring agencies and vendors on how these methods may be best suited for implementation in routine regulatory networks. This section attempts to summarize a number of points in how to best set-up and operate $PM_{2.5}$ continuous monitors. Many of these suggestions have already been incorporated into commercially available monitors. None of the suggestions should be considered as "required" since ultimately the best measures of success are performance of the $PM_{2.5}$ continuous monitor with respect to its ability to reproduce itself (measurement precision) and comparison to a FRM (bias).

Recommendations for Design and Operation of $PM_{2.5}$ Continuous Methods

In order to design an appropriate configuration for a $PM_{2.5}$ continuous monitor many issues need to be addressed. This section attempts to provide the general specifications for $PM_{2.5}$ continuous methods. A detailed accounting comparing the FRM design and performance specifications with applicability to a generic $PM_{2.5}$ continuous monitor follows.

Comparing FRM and Continuous Methods for Design and Performance Criteria

The Federal Reference Method is based upon both design and performance criteria as identified in 40 CFR, Part 50, Appendix L. Design criteria are applicable to components of the reference method such as the inlet and second stage separation device. Performance criteria are applicable to things such as the control of flow rate and maximum allowable temperature difference between the filter and the ambient temperature. For any potential continuous method to be used in the routine regulatory network only performance criteria with respect to the comparison of collocated FRM and continuous data are to be used. However, the performance of a continuous method may be expected to be optimized by adhering to as much of the reference method as practical. In reality, many aspects of the design and performance of the FRM will not be included in a continuous methods operation due to the measurement principle of the instrument or other factors. For instance, much of the laboratory FRM criteria are not practical since there is not expected to be any pre or post-sampling gravimetric analyses in the traditional sense. This section discusses the current understanding of the FRM design and performance criteria that may be applicable to a potential continuous method for use in a regulatory network. Also, where applicable, alternatives to the design and performance criteria of the FRM are included as may be appropriate for use with continuous methods. This section is intended to provide information on how a continuous method might best be designed so that resulting data mimic that of the FRM. Due to the inherent operation of any one continuous method, many of the FRM design and performance criteria may not be suitable for inclusion in its design; therefore, none of the FRM criteria are required. Also, improvements to a design or performance criteria of the FRM are encouraged where appropriate in order for resulting PM continuous data to match that of the FRM.

General Specifications

There are many specifications listed in the FRM as detailed in 40 CFR, Part 50, Appendix L. Among the general specifications, a number of items may be applicable to PM continuous monitoring. This section details those general provisions of the FRM that should be included in the design of a PM continuous method:

- *Pollutant* - Fine particulate matter having an aerodynamic diameter less than or equal to a nominal 2.5 micrometers in the ambient air. Surrogates of this are possible if they result in meeting the necessary performance standards identified in section 6 of this document.

- *Units* - Provide for data to be reported in units of micrograms per cubic meter. This may be calculated directly or indirectly through use of other inputs.

- *$PM_{2.5}$ measurement range* - Provide for a lower and upper concentration limits that allow for meaningful comparison to the FRM. While the FRM is estimated to have a lower concentration limit of at least 2 ug/m^3 and upper concentration limit of at least 200 ug/m^3, continuous methods may be able to operate over an even wider range of concentrations. Most importantly, PM continuous methods need to provide concentration values in the environments they operate in. For instance, in an extremely dirty environment, a continuous method may be able to operate above 200ug/m^3, if designed appropriately. Similarly, when a continuous method is operated at a very clean site the performance of the instrument should be able to discern changes in ambient $PM_{2.5}$ even over very low concentrations.

- *Sample Period* - Provide for a sample period that can be used to calculate the midnight to midnight 24-hour average $PM_{2.5}$ concentration. For all other criteria pollutant continuous data the reported averaging period is usually 1-hour. Depending on the precision of the PM continuous instrument shorter or longer averaging periods may be necessary in order to have a meaningful averaging period. Therefore, 1-hour averages should be capable of being reported; however shorter or longer averaging periods may be necessary depending on the measurement precision of the instrument.

- *Accuracy and Precision* - Because of the size and volatility of the particles making up ambient PM vary over a wide range and the mass concentration of particles varies with particle size it is difficult to define the accuracy of $PM_{2.5}$ measurements in an absolute sense. The accuracy of $PM_{2.5}$ measurements is therefore defined in a relative sense, referenced to measurements provide by the FRM. Section 6 defines the performance standards for $PM_{2.5}$ continuous methods.

Design Criteria

Design criteria for the FRM are largely associated with the inlet and separation device to obtain the desired size selection of aerosol in the sample stream. Many of these criteria can be applied to a potential continuous method. Most of the commercial vendors of PM continuous methods have already incorporated these design criteria into their instruments. The table below describes the various design criteria for the FRM and their applicability to PM continuous methods. Also, where appropriate, alternatives to the FRM design criteria are offered:

Design Areas	Section of Appendix L	FRM specification	Applicability to Continuous Methods
Inlet Assembly	7.3.2	PM_{10} head with dimensions as described in figures L-2 through L-18	This should be applicable to most PM continuous methods
Downtube	7.3.3	With dimensions as described in figure L-19	This may or may not be applicable to a PM continuous method. A downtube may not be needed if there is sufficient clearance for the PM 10 head above the monitor. Also, there needs to be a provision for a leak check adapter to be attached at the point where the PM_{10} heads attaches if the downtube is not utilized.
Impactor	7.3.4	WINS with dimensions as described in Figures L-20 through L-24.	The WINS may be used or alternatively the Sharp Cut Cyclone (SCC) or newer generation of SCC or other cyclone providing an appropriate $PM_{2.5}$ separation may be used. The SCC is expected to maintain an appropriate separation of coarse and fine particulate over a longer period of time than the WINS making it more suitable for use with $PM_{2.5}$ continuous monitors.
Filter Holder Assembly	7.3.5	Many specifications as described in the text and with dimensions as detailed in Figures L-25 through L-29.	Most of the filter holder assembly design specifications will not be applicable to PM continuous monitors. Some of the important areas to strive for in the design of a PM continuous method include: - providing for a uniform face velocity of the sample stream during sample collection. - preclude significant exposure of the filter (or surrogate collection device) to possible contamination.
Flow Rate Measurement Adapter	7.3.6	As described with the dimensions in Figure L-30	Ideally, this would be the same so that flow rate adapters would be interchangeable between FRMs and continuous methods.

Surface Finish	7.3.7	Anodized aluminum for all internal surfaces exposed to sample air prior to the filter.	Ideally continuous methods will also have anodized aluminum for all internal surfaces exposed to sample air prior to the filter or surrogate collection device. This is especially important to note for the Sharp Cut Cyclone; if used, since it is not part of the FRM.
Sampling Height	7.3.8	2 meters ± 0.2 meters	Ideally, the sample inlet on a continuous method would meet this.

Performance Specifications

Performance specifications for the FRM are largely associated with maintaining the flow rate within an acceptable range and the operational conditions for which the instrument should be capable of operating in. Most of the flow rate performance specifications for the FRM should be applicable to continuous methods; however, the operational conditions for which an instrument should be capable of operating in may or may not be applicable to any one continuous method. Many of these performance criteria can be applied to a potential continuous method. Most of the commercial vendors of PM continuous methods have already incorporated these performance criteria into their instruments. The table below describes the various performance specifications for the FRM and their applicability to PM continuous methods. Also, where appropriate, alternatives to the FRM performance specifications are offered:

Performance Specification Area	Section of Appendix L	FRM specification	Applicability to Continuous Methods
Sample Flow Rate	7.4.1	16.67 L/min measured as actual volumetric flow rate at the temperature and pressure of the sample air entering the inlet.	Generally applicable with the exception of any potential use of nephelometers. This flow rate is necessary if a PM_{10} size selective inlet is used as well as for most second stage separators.
Leak Test Capability	7.4.6	Provide for an convenient external leak test capability	Generally applicable.
Range of Operational Conditions	7.4.7	Ambient Temperature -30 to +45 C Ambient Relative Humidity 0 to 100 percent Barometric Pressure 600 to 800 mm Hg	Generally applicable as a starting point for design of an instrument; however, some continuous instruments may need to be located in an environmentally controlled shelter in order to have operate correctly. Some instruments may not meet all of these specifications which may limit their use geographically.
Ambient Temperature and Barometric Pressure Sensors:	7.4.8 and 7.4.9	Capable of operating over the range of operating conditions	Applicable for the operation of the continuous instruments in the range of environmental conditions they will encounter.
Filter Temperature Control	7.4.10	The sampler shall provide a means to limit the temperature rise of the sample filter from isolation and other sources to no more than 5C above the temperature of the ambient air surrounding the sampler.	It is desirable to minimize the temperature difference between the ambient air and the location where sample are collected and analyzed in a continuous method to provide for minimal volatilization of PM; however, in some cases heating may be necessary due to moisture interference or other reasons. Each potential continuous method should be designed to optimize this temperature difference with respect to avoiding moisture interference, PM volatilization, and stable measurement readings.

Filter Temperature Sensor	7.4.11	Capable of operating over the range of operating conditions	Generally applicable. However, may not always be required depending on the measurement principle of the continuous method.
Clock/timer system	7.4.12	Capable of maintaining local time and date including year, month, day of month, hour, minute, and second to an accuracy of ± 1.0 minute per month.	Generally applicable.
Outdoor Environmental Enclosure	7.4.14	Suitable to protect the instrument	Generally applicable for those instruments intended to be located outside. Not necessarily applicable to those instruments intended to be located in a station trailer or other environmentally controlled housing
Electrical Power Supply	7.4.15	105 to 125 volts AC (RMS) at a frequency of 59 to 61 Hz.	Generally applicable.
Data Output Port Requirements	7.4.17	Standard RS-232C	The Standard RS-232C data output connection can be utilized. Additionally, it is strongly encouraged to have a provision for an analog output that can be conveniently connected to a typical data logger utilized by ambient air monitoring agencies. For example, 0 - 10mV, 0-100mV, 0-1V, 0-5V, or 0-10V.

Section 10. Linkage to national monitoring strategy

The EPA in partnership with it's principal grantees; States, local agencies and Tribes, are formulating a national air monitoring strategy that strives to enhance the overall effectiveness of major regulatory based monitoring efforts throughout the nation. The continuous PM monitoring plan addressed here is a major sub-component of this more comprehensive air monitoring strategy. A brief overview of the air monitoring strategy with selected attachments is provided to understand the larger context of the of role of continuous PM monitoring in the nation's reshaping of air monitoring

The monitoring strategy includes establishing a future direction for the shape and scope of air networks throughout the United States. This direction must incorporate knowledge acquired in air quality research and management practices over the last two decades, and take advantage of the much of the existing infrastructure of operating networks and monitoring agencies. The experience over the last 20 years suggests three basic enhancements in national network design:

1) ***multiple and collocated pollutant measurements*** to better diagnose cause effect phenomena in health association and atmospheric process characterization efforts,

2) ***regional scale air quality characterization*** to understand the linkage between background and transport concentrations (regional, continental, global scales) as they impact both rural and urban environments, an increasingly important need as the separation between rural and urban air pollution levels continues to decrease.

3) ***accommodating new technologies*** to provide timely reporting of air quality information to the public and to improve basic characterization of physical, chemical, temporal and spatial composition of air quality.

Consistent with these enhancements, the strategy has identified needed improvements to the monitoring program:

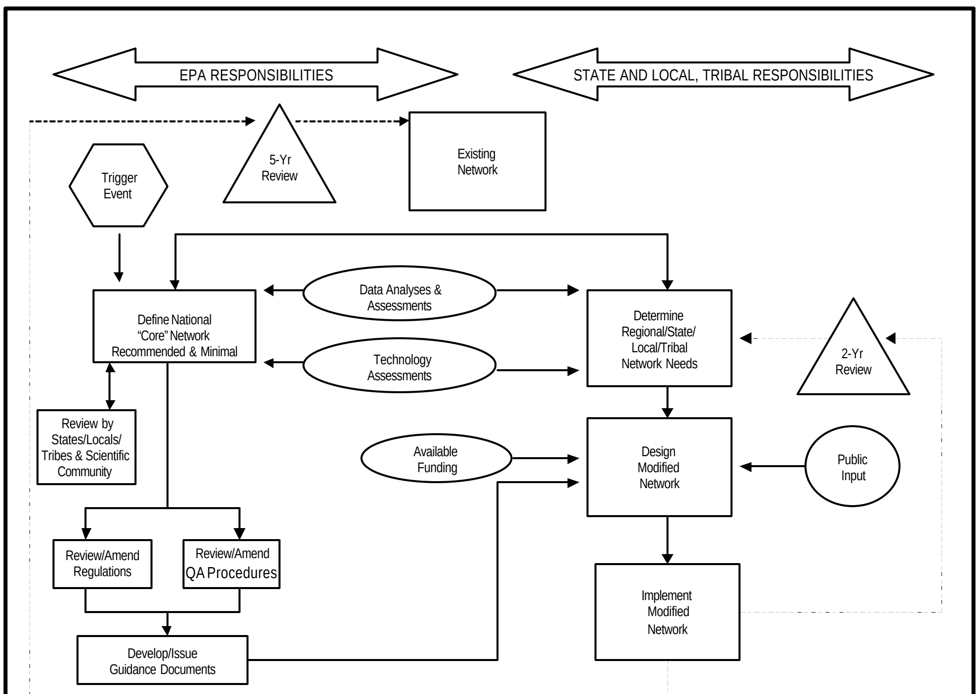
- characterization of hazardous air pollutants (HAPs)
- continuous particulate matter monitoring
- information transfer and delivery
- integration across pollutant programs; and

divestment in much of the existing criteria pollutant monitoring networks.

The strategy is challenged to create adequate flexibility for States, Tribes and local agencies to address area specific problems and simultaneously yield a core of consistent measurements nationally within an anticipated flat resource allocation. This strategy is being guided by the National Monitoring Steering Committee (NMSC), which combines a combination of monitoring and air program management leadership from States, local agencies, Tribes and EPA. The NMSC will be delivering the strategy, which is largely a set of directional and specific recommendations for change in monitoring nationwide for broader public and scientific review in early 2002. Several efforts are underway to evaluate the effectiveness of existing networks and provide a vision for future operations, including:

- Development of network objectives and priorities to guide future investments and divestment;
- a network design proposal for nationally consistent multi pollutant measurement stations;
- National and regional based assessments of existing criteria pollutant networks that attempt to identify existing opportunities for criteria monitoring divestment;
- modifications of existing regulations and quality assurance practices to implement the recommendations emerging from the assessment; and
- accommodation of advanced monitoring and information transfer technologies to enhance scientific value of data collected and dissemination of public information.

Figure 1 illustrates the information flow across these various components.



Information flow and integration across strategy elements indicating influence on networks

This PM continuous monitoring implementation plan provides a test case for this important fifth element of the air monitoring strategy, and the ability to implement continuous monitors is impacted by all of the strategy elements. The broader vision for a PM network includes an integrated hybrid network of filter based and continuously operating samplers. The current $PM_{2.5}$ network of approximately 1100 integrated samplers (FRMs) and nearly 200 “uncoordinated” continuous samplers should evolve into a system of perhaps 700- 1000 $PM_{2.5}$ samplers with a more even distribution (e.g., 50-50) of integrated and continuous methods. The continuous methods must be integrated to ensure data compatibility with the current FRM network . Currently, EPA provides only limited specification on operational guidance or performance expectations for continuous samplers, which limits the ability to utilize many of the existing continuous monitors to support an array of spatially oriented data uses such as model evaluation and PMF applications. Currently, only a very small fraction of continuous monitors enhance the spatial depth of the existing FRM network The challenge in this strategy is to maximize the benefit of continuous samplers so that data analysts are not constantly confronted with screening out instrument types for non regulatory use. This goal is challenged further by an existing inventory of diverse methods using various measurement principles, and the recognition that the measurement from an integrated sampler in many instances has several inherently different (and meaningful) physical and chemical features with respect to a filter measurement.

Assuming no new resource initiative for PM monitoring, resources for the enhancement and integration of continuous monitors will largely come from the existing resource base. This assumption implies a substantial reduction of FRM operations to free resources for operation of continuous samplers. The assessment work (element 3) to date has identified several areas where there is redundancy of samplers and therefore the potential for a reduction of FRM monitors. EPA needs to develop specific guidance for selecting candidate sites for removal based on the assessment and related spinoff products. Such guidance would incorporate design objectives that seek to eliminate sampling redundancy through correlation or related analysis, and enhance spatial coverage through mapping and kriging approaches. EPA will deliver this guidance in mid-2002. Meanwhile, a set of specific recommendations to modify existing PM monitoring regulations will be delivered as part of the strategy in early 2002. These modifications will reduce the number of required $PM_{2.5}$. FRM sites to free operational resources and enable agencies to invest in continuous methods. The recommended revisions must address the performance expectations and test requirements for Regionally Equivalent Monitors (Section 4) and lay out the basic network design framework (Section 5) expected for an integrated $PM_{2.5}$ system.

The NMSC has identified a national need to move toward a multi pollutant network that emphasizes hazardous air pollutants, continuous PM and advanced information transfer technology. In addition to addressing methods, the technology component (element 5) of the strategy provides the rationale and approach for enhancing information transfer and data analysis to increase data usage. The network design effort (element 2) is recommending a set of National Core (NCore) multi pollutant monitoring sites located in major metropolitan areas and selected rural environments. The goals of these sites include assistance for health and exposure studies, air quality management and monitoring m methods. The research community should realize long term benefits from these goals, which are similar to the objectives being addressed in the existing Supersites program. The use of these NCore sites to serve as multi pollutant methods platforms that collocate continuos and integrated PM measurements is critical to the long term integration of continuous and filter based methods. The relationship between a continuous sampler and an FRM is impacted by composition and meteorology which vary in time and space. The NCore platforms could maintain system integration by supporting iterative review of the statistical relationships between collocated integrated and continuous methods as aerosol composition changes arise from future demographic shifts and implementation of emission reduction strategies.

Section 11. Regulatory Changes and Schedule

There are a number of federal regulations that are used to provide the framework for ambient air quality monitoring. These regulations cover the sampling and analytical methods used, how new methods are approved, quality assurance and control procedures, and basic monitoring objectives for certain air pollutants. A great deal more technical information is provided in guidance documents and through the Internet.

Guidance documents are relatively easy to modify as new procedures and technologies appear within the monitoring community. Federal regulations are not particularly easy to modify; however, periodic reviews and revisions are necessary in order to create an air monitoring system that is responsive to current environmental data needs. Along with a variety of topics, we intend to review and modify our regulations to incorporate more continuous particle techniques as part of a larger overall national monitoring strategy. We also intend to establish mechanisms for incorporating continuous techniques by using guidance documents whenever possible.

Specific Regulations to be Reviewed:

There are three main regulatory "Parts" of the Code of Federal Regulations (CFR) that we will investigate in our work to modify the monitoring regulations. These regulations are all part of CFR Title 40 which deals with the environment. Specifically:

40 CFR 50[11] Appendices: National Primary and Secondary Ambient Air Quality Standards (NAAQS), Appendix L. This regulation provides us with the NAAQS and the federal reference methods for measuring each air pollutant with an established standard. We are NOT going to modify the national ambient air quality standards with this regulatory review. Reviews and, if needed, revisions of the NAAQS occur in separate formal processes. We do want to review a portion of the minor requirements in the Appendix L portion of this regulation which describes the reference method for measuring $PM_{2.5}$. The overall reference method will not be modified; however, we do want to examine some of the requirements for reporting supplementary data on the samplers' performance. We have successfully completed two annual quality assurance reports on the $PM_{2.5}$ FRM network operation, and we believe that we can reduce the amount of supplementary data being reported to EPA, specifically in Table L-1. This is a small change; however, it may provide some relief to State, local, tribal, and other monitoring agencies' data managers.

40 CFR 53 Ambient Air Monitoring Reference and Equivalent Methods. This regulation provides air quality monitoring instrument manufacturers with the application and testing requirements for reference and equivalent methods that must be followed in order to have their sampler/analyzer approved for regulatory use. The EPA's Office of Research and Development (ORD) is currently responsible for these approvals. This regulation describes the

[11]Regulations are cited in documents using the format "Title# CFR Part#".

complexities of how new criteria pollutant methods can be formally introduced into the ambient air monitoring network. EPA is a strong proponent of this formal process given the policy and financial impact that decisions using data from federal reference and equivalent methods can carry. We will review this regulation; however, changes to it may or may not be taken in this package. The particulate matter National Ambient Air Quality Standard is being reviewed separately by the EPA. This separate process will also be used to promote continuous particulate matter monitoring technologies within our regulations.
The EPA's ORD has established a Reference and Equivalent Method Board that includes members from OAQPS and ORD. This Board's function has been to review new and modified proposals for fine particulate matter monitoring candidate methods, and to provide broader program input into the approval process. OAQPS proposes to expand the role for this Board to include identifying how to incorporate regional equivalency into the existing reference and equivalent method testing program prior to any actual regulatory change. This approach may need to take the form of a pilot project initially. We will also need to examine our regulatory authority for making such a change. There is a precedent for approving regionally based equivalency within the particulate matter program, specifically with the approval of the Oregon DEQ Med-Vol sampler. It will be necessary to follow-up with any regional equivalency process with formal regulatory changes to Part 53.

40 CFR 58 Ambient Air Quality Surveillance. This regulation is a primary focus of our efforts to both incorporate new technologies and to provide data as outlined in the national monitoring strategy. Nearly all data collection and reporting requirements, all the quality assurance requirements, the NAAQS pollutant network design criteria, the air quality index reporting, and annual data certification requirements are included within this regulation. This regulation describes how the Clean Air Act air monitoring authority has been interpreted and implemented by the EPA and our State and local agency partners for air pollutants with established NAAQS. Tribal agencies are not regulated under this provision; however, the technical requirements within should be familiar to any tribal agency that plans to conduct monitoring.
We expect to change the 40 CFR 58 regulations to allow more flexibility in designing the particulate matter monitoring network. One of these changes would include modifying the existing correlated acceptable continuous (CAC) particulate matter monitoring approach to allow for a more network-based approach rather than only the site-by-site approach as defined currently. The original CAC provisions were developed prior to the full deployment of sequential federal reference methods (FRMs) for fine particles as a way to provide sampling frequency relief from daily sampling. Since the sequential FRMs have been available and are working, the CAC provision has largely been ignored by air monitoring agencies. EPA will modify this provision so that it will provide a better mechanism for incorporating continuous particle monitors into the network.

Participants in the Regulatory Review

We have solicited input from a variety of parties for this regulatory review process. Through the larger air monitoring strategy, we have created a National Monitoring Strategy Committee that is providing advice and recommendations for the national air monitoring program. Some of these recommendations will be realized only after regulatory change has taken place. The NMSC has been discussed in section 10 of this document.

We have also created three separate work groups, one each for the subjects of regulatory review, quality assurance, and technology. These work groups were established to make some concrete progress on the program changes needed to realize the national monitoring strategy goals. The quality assurance group will provide recommendations for changes to the quality assurance provisions of the monitoring regulations as well as all existing quality assurance practices; and the technology work group will make recommendations for use in the methods sections of the regulations and in technical guidance used by monitoring agencies. The regulatory review work group must take information from all of these parties, in addition to the NMSC and the work group's own recommendations, and develop an appropriate regulatory package.

The NMSC and the three work groups include representatives from the EPA OAQPS, the ten EPA Regional Offices, State agencies, local agencies, and tribal governments. All regulatory changes will undergo public review and comment inherent within the regulatory modification process. EPA will also work through existing mechanisms such as the STAPPA/ALAPCO Monitoring Committee and the Standing Air Monitoring Work Group (SAMWG) to communicate with stakeholders on these regulatory changes.

Schedule

Regulatory changes typically take a minimum of 18 to 24 months to complete, including the original proposal package preparation, publication and comment periods, reviewing and responding to comments, and finalizing a package for publication.

Cost estimates are generally prepared for a rule-making action such as the one. A complete funding review of air monitoring grant funds will also be needed, but this should be part of the overall monitoring strategy implementation, and not tied as directly to this package.

Key Milestones (later milestones are subject to change):
October - NMSC recommendations on the national network.
October 23-25 - Monitoring Strategy Workshop
December 1 - Draft rule-making language prepared for work group review.
January - External scientific review of monitoring strategy
June 2002 - Proposal in the Federal Register
July-September 2002 - Public comment period
October - December 2002 - Review public comments, prepare responses
January 2003 - Final regulatory package published in Federal Register

Section 12. Summary of Issues and Action Items

This document serves as a bridge between initial concepts for integrating continuous PM monitors presented at the meeting with the Clean Air Science Advisory Committee's Subcommittee on Particle Monitoring in January, 2001 and comprehensive guidance for monitoring agencies. There remain numerous details not addressed at this time that should be addressed to ensure a satisfactory outcome. These issues and other areas of concern include:

- ***Complex program***. The concepts and elements incorporated in this plan are singularly and collectively complex therefore creating a communications challenge. Other approaches were considered, but the potential drawbacks of a simplistic approach were not acceptable. That is, it would have been easy to develop a rigorous non-flexible program easily communicable but conveying little motivation for deployment. Similarly, a program without constraints would likely compromise data quality and interpretability. Thus, a decision was made to accommodate both flexibility and data comparability at the expense of developing and communicating a complex program.

- ***Annual standard versus daily***. The DQO analyses performed to date have assumed that the annual standard is the driving standard. Since the annual standard involves the average of three numbers, each of which is based on at least 44 numbers but usually more than 60, it is clear why decision errors are not very sensitive to measurement imprecision and why it is proposed that the measurement precision performance criterion be 20% CV. DQOs based on the daily standard, which involves the average of 3, annual 98^{th} percentiles, may show that decision errors are sensitive to measurement imprecision. Additional analyses will be performed to assess the importance of measurement imprecision for decision errors associated with the daily standard. Similarly, analyses will be performed to assess the importance of measurement imprecision for decisions made with non-aggregated data, such as AQI reporting.

- ***Rescinding REM certification based on future poor performance***. The REM program is based on demonstrating an acceptable level of comparison between FRM and continuous samplers. This relationship may change as a result of atmospheric changes due to deployment of emission mitigation strategies. Guidance, albeit complex, will allow for a non static relationship. Nonetheless, this potential for aerosol change will require iterative evaluation of instrument performance that is likely, in some instances, to show that a previously approved REM fails performance goals.

- ***Guidance for developing and approving regional equivalent domains***. The information in this document can be applied in a somewhat straightforward manner for approving an instrument for CAC or REM purposes at an individual site. The larger goal is to broaden this acceptance to a "region" where the meteorological and aerosol composition characteristics exhibit consistent behavior and hence throughout which the continuous and FRM methods exhibit similar relationships. Regionality is further complicated by administrative and demographic issues (e.g., multiple monitoring agencies and State boundaries intersecting within a given "region"). This topic has not been adequately addressed in this document and requires additional effort. The overall complexity of regionality and the use of transformation models might suggest development of a review board to handle REM requests on a case by case basis.

- ***Reliance on FRM measurements as an indicator***. The underlying approaches require comparability of continuous and FRM measurements. The reason for this is that so many objectives relate to the FRM measurement (e.g., NAAQS comparisons, AQI, air quality model application). In many instances, there is no technical reason to expect comparability between disparate measurement approaches. Such comparability is desired given the utility of relating continuous measurements to a wealth of existing FRM data and to incorporate a reference marker. The downside of this approach is that the value of an FRM measurement is assumed or inferred to be greater than that of a candidate method, when in some cases the candidate method may better reflect "true" characteristics of an aerosol.

- ***Specific Guidance on Performance Specifications***. Sections 5 - 8 introduce performance specifications for bias and precision, but several specific details are not addressed. For example, how is bias measured? What is the statistic as well as what is the source of the data to be used in the statistic? Are bias estimates based only on existing collocated instruments or is an independent audit required? How are bias and precision treated on a regional basis, does the failure of one site constitute failure for a region, or are all estimates averaged across a region? What is the appropriate frequency for checking bias and precision? These unique considerations warrant development of a dedicated Quality Assurance program for CAC and REM applications.

- ***Data interpretation and management***. Transformed data are to be submitted to AIRS. How do analysts gain access to raw non-transformed data? Transformation models are based on 24-hr comparisons, yet transformed data will be reported continuously, which may create odd results in discrete hourly reporting. Coding specifications for CAC and REM need to be developed.

- ***Demonstration of performance***. The bias and precision estimates are based on existing network performance. This implies that the testing to meet such specifications should be conducted under conditions consistent with routine operations. This approach should not be interpreted as excluding desired vendor participation. Responsibilities for conducting testing, developing transformations and communicating performance results requires further effort.

- ***Consistency with FEM***. The current Class III equivalency requirements appear to be more strict than what a FRM can meet. That is, the imprecision in the FRM is such that the R^2 requirement can not be met, not because of the challenging instrument, but because of the instrument being used as the standard. This inconsistency needs to be addressed. In doing so, it may make it possible for an instrument to acquire a Class III equivalency.

ATTACHMENT A – $PM_{2.5}$ Continuous /FRM Analysis

TO BE PROVIDED AT A LATER DATE

Attachment B. DRAFT TECHNICAL REPORT

Data Quality Objectives (DQOs) for $PM_{2.5}$

for

U.S. ENVIRONMENTAL PROTECTION AGENCY
Office of Air Quality Planning and Standards
Emissions, Monitoring, and Analysis Division (MD-14)
Research Triangle Park, North Carolina 27711

Michael Papp, EPA Work Assignment Manager
Vickie Presnell, EPA Project Officer

Contract No. 68-D-98-030
Work Assignment 4-08

July 25, 2001

Prepared by

Basil Coutant, Tim Pivetz, and Jesse Frey
BATTELLE
505 King Avenue
Columbus, Ohio 43201-2693

and

Michael Papp
OAQPS/EMAD (MD-14)
U.S. ENVIRONMENTAL PROTECTION AGENCY
Research Triangle Park, North Carolina 27711

TABLE OF CONTENTS

List of Tables

List of Figures

DRAFT TECHNICAL REPORT

DATA QUALITY OBJECTIVES (DQOS) FOR $PM_{2.5}$

1.0 INTRODUCTION

An important concern in any organization that is collecting and evaluating environmental data must be the quality of the results. A quality system [1] must be developed and documented to ensure that the $PM_{2.5}$ monitoring results:

- meet a well-defined need, use, or purpose;
- satisfy customers expectations;
- comply with applicable standards and specifications;
- comply with statutory (and other) requirements of society; and
- reflect consideration of cost and economics.

The development of a quality system for $PM_{2.5}$ requires a coordinated effort between EPA and the State and local monitoring community and tribal organizations. Elements of the quality system include planning, implementation, and assessment. As part of the planning effort, EPA is responsible for developing National Ambient Air Quality Standards (NAAQS), defining the quality of the data necessary to make comparisons to the NAAQS, and identifying a minimum set of QC samples from which to judge data quality. The State and local organizations are responsible for using this information to develop and implement a quality system that will meet the data quality requirements. Then, it is the responsibility of both EPA and the State and local organizations to assess the quality of the data and take corrective action when appropriate. This document describes the approach used in developing a quality system for the $PM_{2.5}$ monitoring program. It is based on both the initial DQO development done in 1997, prior to the network establishment, and an assessment of the major assumptions that went into that development using 1999 and 2000 data from the network. Following the planning, implementation, and assessment theme, the discussion includes the:

1. development of data quality objectives (DQOs);
2. identification of the types and frequencies of QC samples, based upon the DQOs, to evaluate and control measurement uncertainty;
3. data quality assessment (DQA) process used to compare measurement uncertainty to the DQO; and
4. consequences of failing to meet the DQOs.

1.1 Data Quality Objectives

DQOs are qualitative and quantitative statements derived from the DQO Process that clarify the monitoring objectives, define the appropriate type of data, and specify the tolerable levels of measurement errors for the monitoring program [2]. By applying the DQO Process to the development of a quality system for $PM_{2.5}$, the EPA guards against committing resources to data collection efforts that do not support a defensible air quality management program. The DQO Process that follows illustrates the steps taken to assess the quality of data needed for making comparisons to the $PM_{2.5}$ NAAQS. The focus of this document is the annual NAAQS based on the 3-year annual arithmetic mean concentration. Throughout this document, the term *decision maker* will be used. This term represents individuals that are the ultimate users of ambient air data and, therefore, may be responsible for: setting the NAAQS, developing a quality system, evaluating the data, or making comparisons to the NAAQS to determine if a standard is or is not violated. The DQOs will be based on the data requirements of the decision maker(s).

In order to understand the DQO Process, a discussion on data uncertainty will follow, which will lead into the discussion of the $PM_{2.5}$ DQO.

1.2 Data Uncertainty

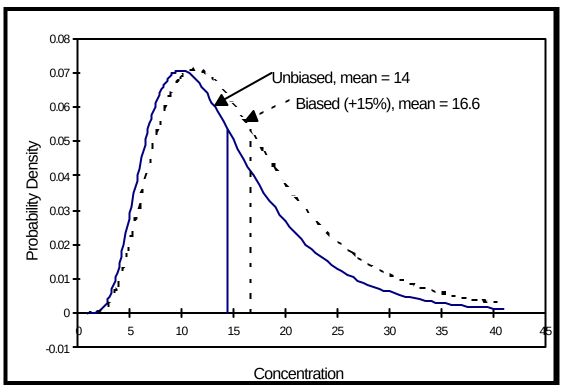


Figure 1. Effect of positive bias on the annual average estimate.

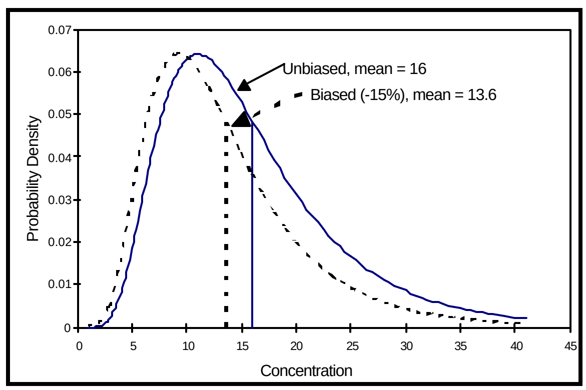


Figure 2. Effect of negative bias on the annual average estimate

Decision makers need to feel confident that the data used to make environmental decisions are of adequate quality. The data used in these decisions are never error free and always contain some level of uncertainty. Because of these uncertainties or errors, there is a possibility that measurements may yield annual averages above 15.0 :g/m^3 when the average is actually below 15.0 :g/m^3 (false positive error as illustrated in Figure 1) or below 15.0 :g/m^3 when actually the mean is above 15.0 :g/m^3 (false negative error as illustrated in Figure 2). Therefore, decision makers need to understand and set limits on the probabilities of these types of uncertainties in these data.

The DQO defines the acceptable level of data uncertainty. The term "uncertainty" is used as a generic term to describe the sum of all sources of error associated with a given portion of the measurement system. The estimate of the overall uncertainty that the decision makers are willing to accept leads to the DQO. Overall data uncertainty is the sum of **total population uncertainty** and **total measurement uncertainty**.

Total Population Uncertainty is defined as the natural spatial and temporal variability in the population of the data being evaluated. Population uncertainty can be controlled through the use of statistical sampling design techniques, the proper placement of ambient air quality monitors, spatial averaging (as allowed by the $PM_{2.5}$ NAAQS), and maintaining sampling frequency and completeness standards. Since the population of concern for the $PM_{2.5}$ NAAQS violation decision is a single instrument (each instrument can effect the attainment/nonattainment decision), the population uncertainty would be the uncertainty over the 3-year averaging period. During the development of the NAAQS, population uncertainty, due to temporal variability, was incorporated into the standard by stating that 3 complete years of data determines compliance with the NAAQS, even though the expected value may be different. Therefore, temporal variability would be considered completely accounted for, as long as every day sampling was implemented. However, 1-in-6-day sampling and 1-in-3-day sampling, or any deviation from every day sampling, have an impact on uncertainly that must be understood, and, if possible, quantified.

Total Measurement Uncertainty is the total error associated with the environmental data operation. The environmental data operation for $PM_{2.5}$ represents various data collection activities or phases including: the initial weighing of the filters (and the conditions in which they are weighed), the transportation of the filters, the calibration of the instrument and its maintenance, the handling and placement of the filters, the proper operation of the instrument (sample collection), the removal, handling and transportation of the filter, the storage and weighing of the sampled filter, and, finally, the data reduction and reporting of the value. At each phase of this process, errors can occur that, in most cases, are additive. The goal of a QA program is to control total measurement uncertainty to an acceptable level through the use of various quality control and evaluation techniques. In a resource constrained environment, it is most important to be able to calculate/evaluate the total measurement uncertainty and compare this to the DQO. Various phases (field, laboratory) of the measurement system can be evaluated, subject to the availability of resources.

Two data quality indicators are most important in determining total measurement uncertainty:

- **Precision** - a measure of mutual agreement among individual measurements of the same property usually under prescribed similar conditions. This is the random component of error. Precision is estimated by various statistical techniques using some derivation of the standard deviation. For the $PM_{2.5}$ DQO, the coefficient of variation (CV) is used, which is the standard deviation divided by the mean, multiplied by 100.

- **Bias** - the systematic or persistent distortion of a measurement process that causes error in one direction. Bias will be determined by estimating the positive or negative deviation from the true value as a percentage of the true value.

Accuracy has been a term frequently used to represent closeness to "truth" and includes a combination of precision and bias error components. For $PM_{2.5}$, the term accuracy will be used when measurement uncertainty cannot be separately associated with precision or bias.

2.0 THE $PM_{2.5}$ DQOS

The $PM_{2.5}$ DQOs were developed to reduce the probability of decision errors by controlling precision, bias, and sampling representativeness. The development was based on a series of assumptions and input criteria. These key assumptions are discussed below. The main assumptions that are data driven were compared with 1999 and 2000 data from the $PM_{2.5}$ mass network. The key inputs from decision makers have been reviewed by decision maker representatives. The power curves in Figure 9 and the error rates in Table 4 incorporate any modifications to the items below as indicated by this review. See Appendix A for additional assumptions and input criteria.

2.1 The DQO is Based on the Annual Arithmetic Mean NAAQS

The $PM_{2.5}$ standards are a 15 :g/m^3 annual average and a 65 :g/m^3 24-hour average. The annual standard is met when the 3-year average of annual arithmetic means is less than or equal to 15 :g/m^3. Due to rounding, the 3-year average does not meet the NAAQS if it equals or exceeds 15.05 prior to rounding. The 24-hour average standard is met when the 3-year average 98th percentile of daily $PM_{2.5}$ concentrations is less than or equal to 65 :g/m^3.

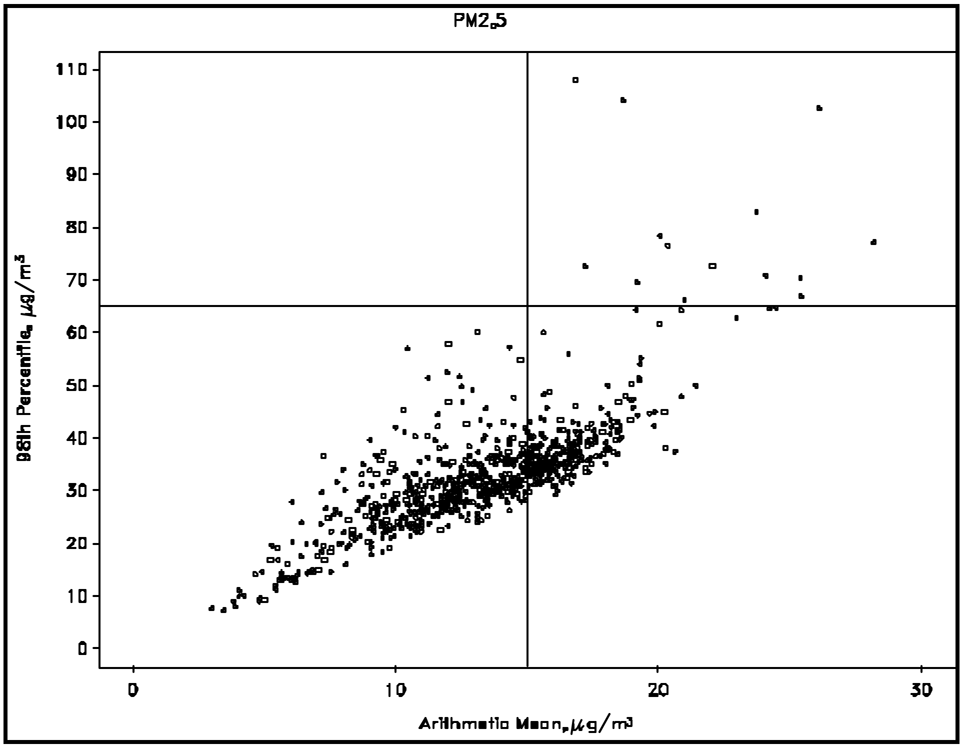


Figure 3. Annual arithmetic mean and 24-hour 98th percentiles from AIRS data (extracted on April 4, 2001)

The original $PM_{2.5}$ DQOs were developed using some $PM_{2.5}$ information (a total of 47 single-year estimates of the annual average and the 24-hour 98th percentile) as well as available PM_{10} information. In order to review and revise these standards, two years of AIRS $PM_{2.5}$ data (extracted on April 4, 2001) were investigated. These data represent the first two years of the mass network. Identifying sites with 90 or more observations in the year 2000 (which represented the first full year of data collection) yielded 757 measurements of annual averages and 24-hour 98th percentiles. These points are plotted in Figure 3. Figure 3 does not display estimates derived according to the standard, as the averages represent one-year averages as opposed to three-year averages, but it does indicate the relative importance of the two standards. Points to the right of the vertical line may be viewed as exceeding the annual average standard. Approximately 34 percent of the annual average measurements exceeded the standard. Only 14 measurements or about 2 percent exceeded the 65 :g/m^3 24-hour standard.

2.2 The Distribution of the Measurement Error

Error in environmental measurements is often assumed to be normal or lognormal. Figures 4 and 5 attempt to illustrate what happens to the normal and lognormal distribution functions for the same median concentration at two values for measurement error (CV's of 10 percent and 50 percent). In the case of $PM_{2.5}$, the measurement error is expected to be in the range of 5 to 10 percent of the mean, as shown in Figure 4, where normal or lognormal errors produce close to identical results. Therefore, due to these comparable results and its simplicity in modeling, the normal distribution of error was selected.

Additionally, measurement error is assumed to be independent from day to day. It is also assumed that the standard deviation of the measurement error is assumed to be proportional to the true concentration being measured . The first of these assumptions is quite reasonable to expect. The second may not be entirely true. However, as long as the measurement error is less than the amount implied by a 10 percent CV, the decision errors will be controlled at the desired levels.

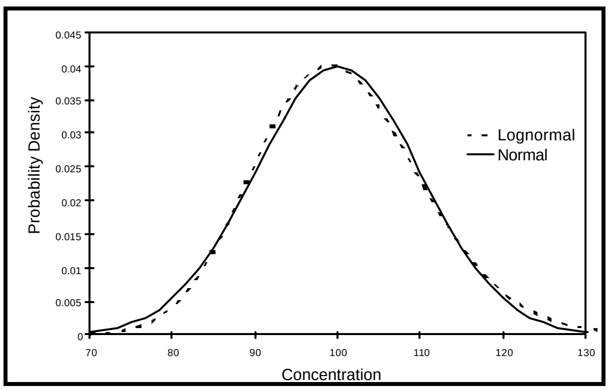


Figure 4. Comparison of normal and lognormal density functions at low measurement error (10 percent CV)

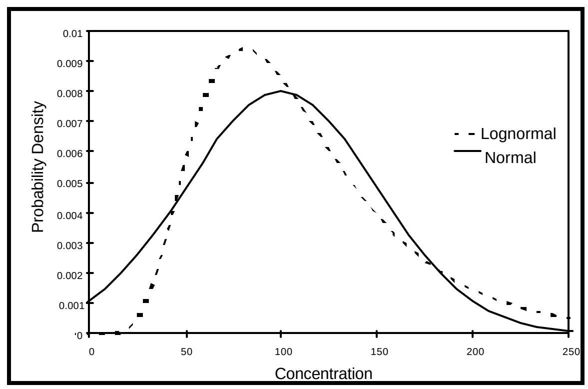


Figure 5. Comparison of normal and lognormal density functions at higher measurement errors (50 percent CV)

2.3 Errors Can Occur When the Estimated 3-Year Average Differs from the Actual, or True, 3-Year Average

Errors in the estimate are due to population uncertainty (sampling less frequently than every day) and measurement uncertainty (bias and imprecision). The false positive error occurs whenever the estimated 3-year average exceeds 15.0:g/m^3 and the actual 3-year average is less than 15.0:g/m^3 (Figure 1). The false negative error occurs whenever the estimated 3-year average is less than 15.0:g/m^3 and the actual 3-year average is greater than 15.0:g/m^3 (Figure 2).

2.4 The Limits on Precision and Bias Are Based on the Smallest Number of Sample Values in a 3-Year Period

Since the requirements allow 1-in-6-day sampling and 75 percent data completeness each quarter, the minimum number of values in a 3-year period is 144. It can be demonstrated that obtaining more data, either through more frequent sampling or the use of spatial averaging, will lower the rate of errors at the same precision and bias acceptance levels. It is assumed that any missing values are random and, thus, unlikely to have significant impact on precision and bias levels.

2.5 The Error Limits Were Set at 5 Percent

For the two cases in Section 3, the decision maker will make the correct decision at least 95 percent of the time if precision and bias are maintained at the acceptable levels and the completeness criteria are satisfied. For cases that are less “challenging” (i.e., have annual average values that are farther from 15.0 $\mu g/m^3$ or are made from less variable data), the decision maker will make the errors less often. Sampling more frequently will also reduce the probability of making an error. Finally, if precision and bias prove to be lower than the values used in the DQO development, the decision maker can expect errors less than 5 percent of the time.

2.6 Measurement Imprecision Was Established at 10 Percent Coefficient of Variation (CV)

The original DQO analysis reviewed available AIRS data and other $PM_{2.5}$ studies to determine that it was reasonable to allow measurement imprecision at 10 percent CV. While measurement imprecision has relatively little impact on the ability to avoid false positive and false negative errors, it is an important factor in estimating bias. CV’s greater than 10 percent make it difficult to detect and correct bias problems. The DQOs are developed assuming a worst case scenario with respect to the bias, in that the bias is always assumed to be in the direction that would result in a decision error.

Other assumptions made concerning precision and bias include the assumption that they will be constant throughout the 3-year period. Similarly, it is assumed that precision and bias are checked sufficiently often to detect significant deviations from the DQOs.

2.7 Assumptions About the Underlying Variability of the Day-to-Day PM Concentrations

For the original DQOs, PM_{10} data from AIRS were reviewed to find a reasonable statistical model for $PM_{2.5}$. These analyses led to choosing a sinusoidal model for the long-term seasonal pattern. The review of the 1999 and 2000 $PM_{2.5}$ network data indicates that this is a reasonable choice for the cases with the largest amount of natural variation.

Specifically, the original DQOs were based on assuming a mean sinusoidal seasonal pattern for the $PM_{2.5}$ concentrations such that the:

- ratio between the high and the low points of the curve was 5.63;
- random population variation about the mean seasonal curve has a normal distribution with a standard deviation that is proportional to the seasonal mean;
- random population variation about the mean seasonal curve has a CV of at most 50 percent; and
- natural day-to-day variation about the sine curve is statistically independent.

Each of these assumptions were meant to reflect a worst case scenario with respect to the assumption's influence on the decision error rates.

A subset of the 1999 and 2000 network data was extracted from AIRS on April 4, 2001 to investigate the original DQO assumptions. The data were limited to sites with an annual mean between 10 and 20 micrograms per cubic meter. This was done mainly to represent the range that is most important to the DQOs. Also, the relative variability (CV) that can be measured could easily be biased by sites with the more extreme means. Next, completeness criteria were applied to ensure representativeness of the results. (See below.) Tables 1 and 2 show the distribution of two of the key characteristics in the DQO development, the ratio of highest to lowest mean values and the CV about those means.

Table 1: Distribution of ratios of highest to lowest monthly or bimonthly mean at a site

		Monthly	Bimonthly
# of sites		289	292
Mean		2.07	1.76
Minimum		1.24	1.11
Percentile	90.0	2.60	2.12
	91.0	2.65	2.36
	92.0	2.79	2.38
	93.0	2.87	2.49
	94.0	3.01	2.57
	95.0	3.70	3.17
	96.0	4.41	3.36
	97.0	4.61	3.90
	98.0	5.25	4.03
	99.0	6.05	4.69
Maximum		6.54	4.89

Table 2: Distribution of CVs about monthly and bimonthly means

		Monthly	Bimonthly
# of Estimates		3,398	1,752
Mean		49.6	50.7
Minimum		16.1	22.9
Percentile	10.0	34.6	37.6
	25.0	40.4	42.8
	50.0	48.1	49.4
	75.0	56.3	56.9
	90.0	66.6	64.7
	95.0	73.7	70.5
	96.0	75.4	72.3
	97.0	78.2	75.9
	98.0	83.8	79.1
	99.0	93.5	89.8

To estimate the ratio of the high and the low means throughout the year and to estimate the variability about those means, monthly and bimonthly averages were considered. Technically, the DQO parameters of interest are daily means and the CV's about those means. These cannot be directly estimated without multiple measurements for each day of the year taken over many years. However, since the assumed sinusoidal behavior does not change much over the period of 1 to 2 months, the data were aggregated to monthly and bimonthly levels. Hence, the ratio of the highest to lowest points on the sine curve is approximated by the ratio of the highest monthly (bimonthly) mean to the lowest. Similarly, for each site and month (or consecutive 2-month period) a CV could be calculated for the period.

For the monthly averages, a site's data were used when there were at least 11 months, each with at least 10 valid measurements. For the bimonthly averages, a site was considered if there were at least 19 valid measurements in each of the six two-month intervals. A site's ratio was then the ratio of the maximum monthly mean (or bimonthly mean) to its minimum monthly mean (or bimonthly mean). The distribution of these ratios is included in Table 1. For each site and month or bimonthly period, a CV was also estimated. Table 2 contains the distribution of CVs about the means. The DQOs need to guard against the most variable cases, so the highest portion of the distribution is most important to this work. However, since the original DQOs were based on a CV of 50 percent, Table 2 indicates that some of the middle portion of the distribution of CVs is important as well.

These analyses show that the ratio of maximum to minimum used in the original DQOs was slightly higher than would be necessary as an estimate of the worst case scenario for the seasonal variability. From the distribution shown in Table 2, it was concluded that the estimate for the upper bound on the CV used in the original analyses was too low. A ratio of 5.3 and CV of 80 percent were chosen to represent the worst case for use in the DQOs and the case studies below (compared to 5.63 and 50 percent, respectively, in the original DQOs).

Figures 6 and 7 show an example of the $PM_{2.5}$ data extracted from AIRS for a fixed site. The mean of the data from this site for the time period shown (January 1999 through November 2000) is 16.3 $\mu g/m^3$. A sine curve with a mean of 16.3 and a ratio of 5.3 between the highest and lowest points on the curve would be given by

$$16.3 + 11.125 \sin(2 \pi D / 365 + \text{phase shift}) \quad .$$

The curve shown in Figures 6 and 7 is

$$16.3 + 11.125 \sin(2 \pi D / 365 + 1.9) \quad ,$$

where D is the number of days into the 3 year cycle. The phase shift 1.9 was chosen to minimize the square error between the sine curve and the data values. Figure 7 has the same data and curve plotted against the number of days into the year rather than the number since January 1, 1999.

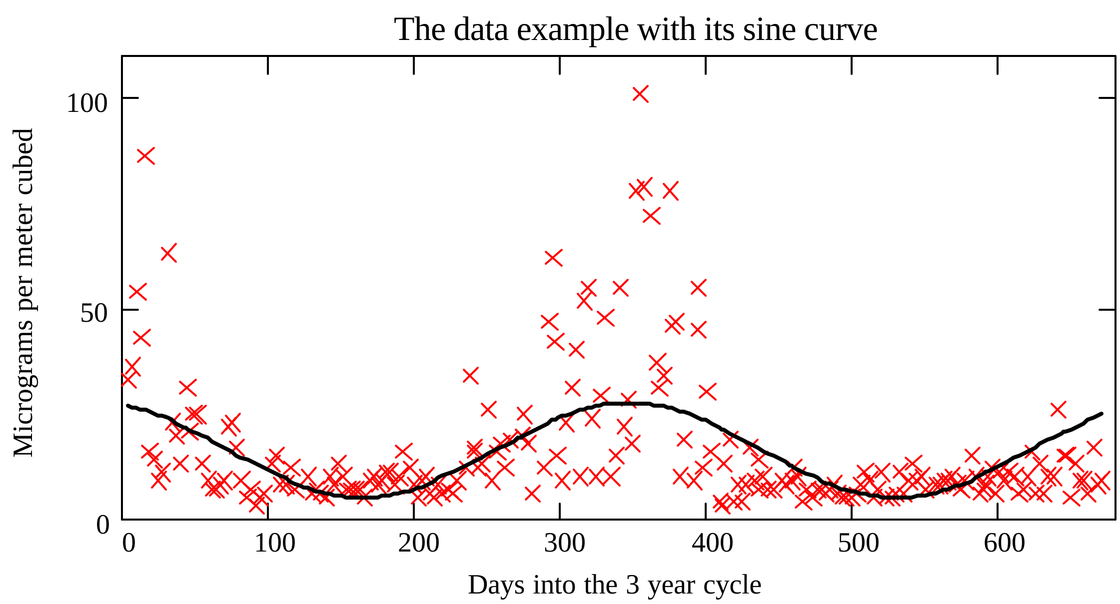


Figure 6. $PM_{2.5}$ concentrations showing a sinusoidal seasonal pattern along with the DQO sine curve that is associated with these data

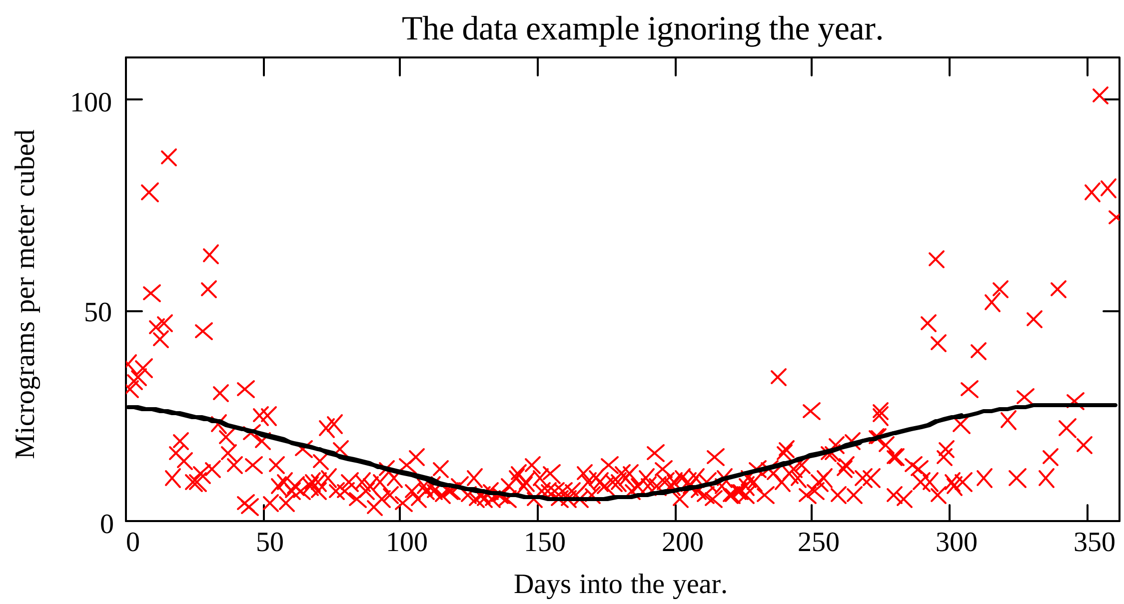


Figure 7. The data example and curve plotted against the number of days into the year

The figures show several key points that the DQO model is designed to simulate. First, the data exhibits seasonal variation. For this particular site, the ratio of the highest monthly mean to the lowest is 5.2. The periods with the highest means have much more variability. The variability “about” the sine curve is not symmetric; the deviations above the curve extend further away from the curve than the deviations below the curve. In fact, if the sign of the deviations from the curve were reversed there would be many negative values. There is also a considerable amount of day-to-day variation. After accounting for the sinusoidal seasonal variation and subtracting out a measurement error CV of 10 percent, the remaining variation is just over 80 percent.

Sinusoidal simulation models were also considered as starting points for developing DQOs. The following model mimics the situation in Figure 7 except for the phase shift[1].

C_D = concentration on Day D = [16.3 + 11.125 sin(2 B D / 365)]* $*_D$,

D = 1,2,... is the number of days into the 3-year cycle ,

where $*_D$ is a random factor that is log-normally distributed with mean one and standard deviation equal to 80 percent. Figure 8 illustrates this function together with simulated $PM_{2.5}$ levels for three years. (Compare with the real data in Figures 6 and 7.) The long-term average concentration is 16.3 :g/m^3. A station having $PM_{2.5}$ levels following this model would virtually always be in a true state of non-attainment, based on the average of three years’ data with no measurement system error.

In revising the DQOs, instead of assuming a normal distribution about the sinusoidal curve, a lognormal distribution was utilized. The lognormal distribution does not produce negative values and is skewed. (See Figure 2.) The original DQOs were based on a normal distribution about the sine curve with a CV of 50 percent. This produced very few negative values that were ignored. However, using a normal distribution with a CV set at 80 percent would have resulted in negative numbers more than ten percent of the time. Hence, the lognormal distribution was chosen to more realistically simulate the natural variation about the sine curve.

1 For modeling purposes, the phase shift is not important since complete years will be used.

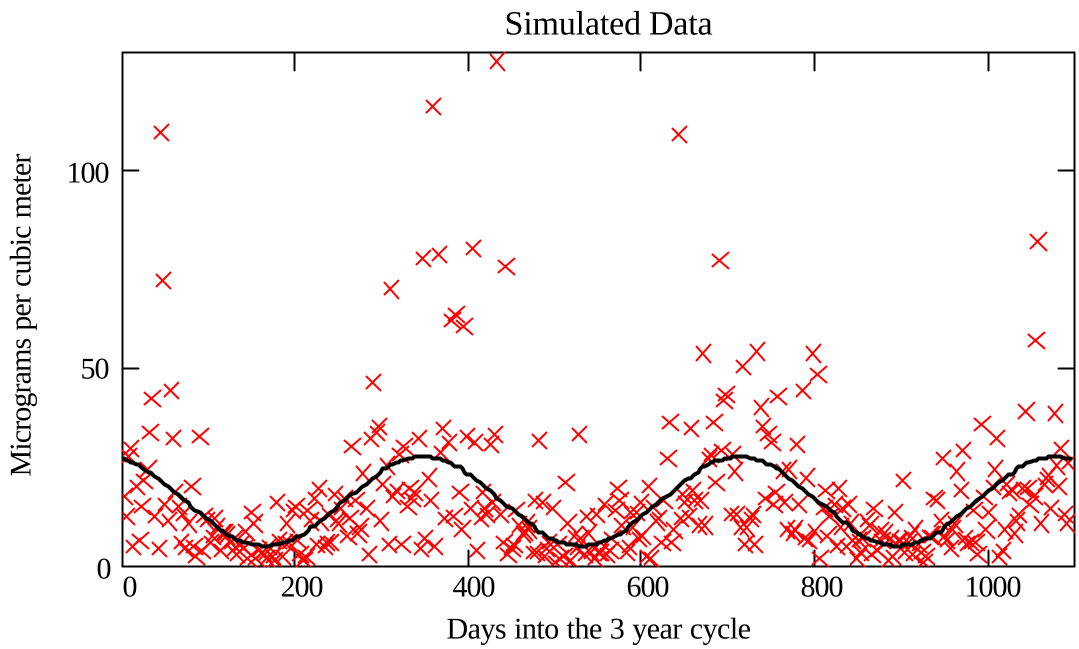


Figure 8. Simulated $PM_{2.5}$ data for every third day with a long-term mean of 16.3 :g/m³ and a population CV of 80 percent

3.0 THE MODELING PROCESS

The relationship of the DQO assumptions and input criteria to the decision error rates was established by Monte Carlo simulation of both the true 3-year annual mean concentrations and the 3-year mean sample mean concentration. A true 3-year mean concentration establishes the correct attainment/nonattainment decision. The sample mean determines the decision.

Each of the items listed in Section 2 impacts the output of this process. The two key outputs are power curves and the associated gray zone. The power curves relate the true 3-year means to the probability of a measured annual mean being above 15.0 :g/m^3. The gray zone is the range of 3-year annual means about the standard where the decision error rate is unavoidably higher than the 5 percent limit set in Section 2.5.

3.1 Gray Zone Boundary Cases

The following two examples illustrate the process used to investigate the effects on ~~decision~~ error rates of various values of precision, bias, and sampling frequency. They show the extremes where the 5 percent error rate is met. As mentioned above, less "challenging" cases, such as cases with means that are further from the NAAQS standard, will have lower error rates.

<u>Case 1: Suppose a site has a true three-year mean of 12.2 :g/m^3.</u>

The correct measurement is an annual mean below 15.0:g/m^3. The probability of the false positive error for sampling every sixth day depends on the measurement system bias and precision. (See Table 4.) As stated in Section 2.6, the data in Table 4 show that precision alone has little impact on error, but is an important factor for determining the bias, which is an important factor in error rates. Figure 9 illustrates the power curves associated with 75 percent complete 1-in-6-day sampling, a population CV of 80 percent (about the sine curves), a measurement CV of 10 percent, and biases of +/- 10 percent. The actual mean obtained from sampling would likely differ from 12.2 because of sampling error (not sampling every day), measurement error, and measurement bias. However, the probability that these factors would combine to yield a mean of at least 15.05 :g/m^3 is only 5 percent.

<u>Case 2: Suppose a site has a true three-year mean of 18.8 :g/m^3</u>.

The correct measurement is an annual mean above 15.0:g/m^3. The probability of the false negative error for sampling every sixth day depends on the measurement system bias and precision. (See Table 4.) As stated in Section 2.6, the data in Table 4 show that precision alone has little impact on decision error, but is an important factor for determining the bias, which is an important factor in decision error. Figure 9 illustrates the power curves associated for 75 percent complete 1-in-6-day sampling, a population CV of 80 percent (about the sine curves), a measurement CV of 10 percent, and biases of +/- 10 percent. The actual mean obtained from sampling would likely differ from 18.8 because of sampling error (not sampling every day), measurement error, and measurement bias. However, the probability that these factors would combine to yield a mean of less than 15.05 :g/m^3 is only 5 percent.

Combinations of precision and bias that yield error probabilities around 5 percent were considered acceptable. After reviewing Cases 1 and 2, and based upon the acceptable decision error of 5 percent, the DQOs for acceptable precision (10 percent CV) and bias ($\pm$ 10 percent) were chosen.

Table 3. Summary of Case 1 and 2 parameters

	3-Year Mean	Correct Decision	Incorrect Decision	Tolerable Error Rate
Case 1	12.2 ug/m^3	Attainment	F(+) = nonattainment	5 %
Case 2	18.8 ug/m^3	Nonattainment	F(-) = attainment	5 %

3.2 Modeling

The probability estimates in Table 4 and the power curves in Figure 9 are developed by modeling 3-year sets of data similar to that shown in Figure 8, except all 3*365 days are generated. For a given long-term expected mean between 10 and 20, many sets of data representing a true 3-year set of data are generated. Each set of random data determines a true 3-year mean that is rounded to the nearest tenth and is either in attainment (at most 15.0) or out of attainment (over 15.0). Then associated with each these data sets, a random set of 144 days is selected such that 12 days are selected from each quarter from a 1-in-6-day sampling scheme. To these values, normally distributed random measurement error and both a positive and negative bias are added. The random measurement error has a mean of 0 and a standard deviation that depends on the magnitude of the particulars day's true value (10 percent for the power curves). This generates a set of sampled data values and a sample mean. Finally, the power curve is generated by repeating this process for a range of long-term expected means. Power is calculated as the percent of the time that sample means from a fixed true 3-year mean are over 15.05.

The power curves show the probability or percent of the time that a measurement of an annual mean is above 15.0 : g/m^3 under different conditions. Power curves are the standard statistical tool for comparing the effects of various input parameters on decision errors. The curves shown in Figure 9 and the error rates in Table 4 are based on the assumptions and input criteria discussed in Section 2. In particular, the key assumptions are a long-term sinusoidal daily mean with ratio of 5.3 between the high and low points of the curve, a lognormal variation about the sine curve with a CV of 80 percent, 75 percent complete 1-in-6-day sampling, normal measurement error with a 10 percent standard deviation, and bias of +/- 10 percent.

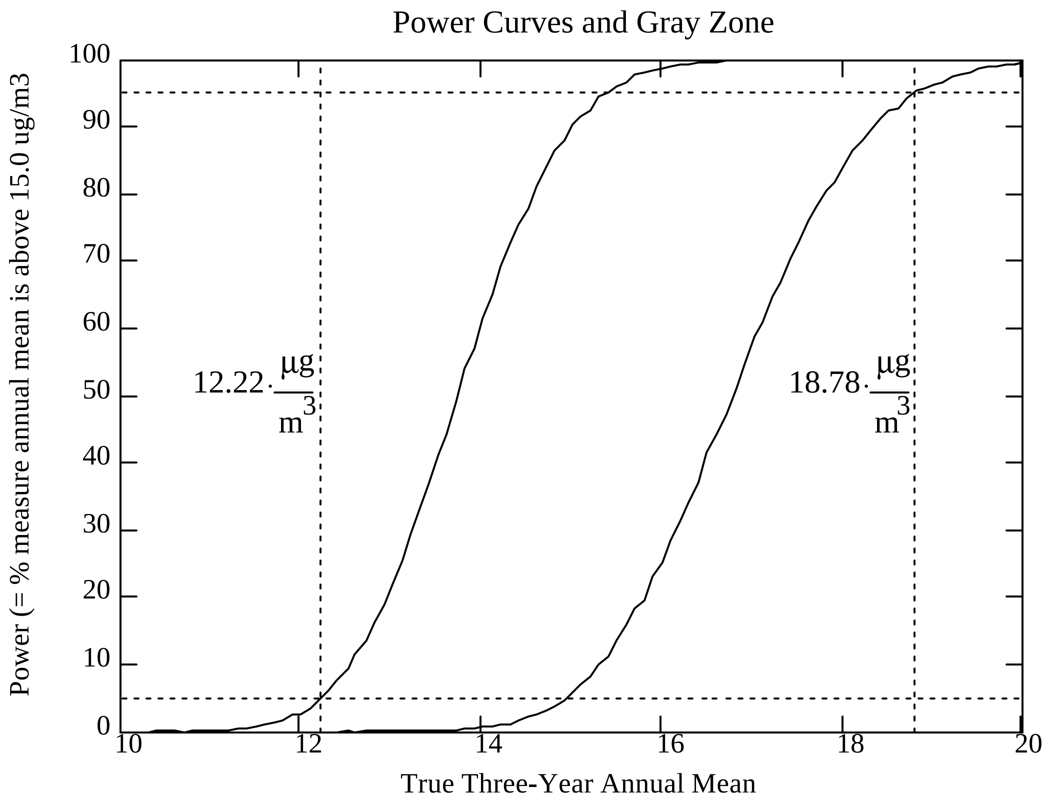


Figure 9. **Power curves for 75 percent complete 1-in-6-day sampling with 10 percent measurment CV and ± 10 percent bias.**

Table 4. Maximum error rates

Absolute Bias *	Measurement CV	Error rate at 12.2 mg/m^3	Error rate at 18.8 mg/m^3
5	0	1%	1%
5	10	1%	1%
5	80	8%	8%
5	100	10%	12%
10	0	5%	5%
10	10	5%	5%
10	80	15%	17%
10	100	18%	21%
15	0	15%	21%
15	10	15%	21%
15	80	26%	31%
15	100	28%	34%

* The bias is taken in the direction that causes the most error.

The Data Quality Objectives Process sets concrete goals to produce sufficient, high quality data for decision makers and data user needs. The process needs to continue throughout the data collection cycle. Assumptions made in the process need to be checked and, if necessary, updated as the data are collected. For $PM_{2.5}$ monitoring program, the combination of at most a 10 percent measurement CV and at most an absolute bias of 10 percent have been chosen to ensure that at most 5 percent decision errors will occur outside the range of 12.2 to 18.8 :g/m^3 for a 1-in-6-day sampling scheme. The key assumptions that went into the original choice of 10 percent measurement CV and 10 percent absolute bias have been checked against the 1999 and 2000 data from the network The basic structure of the assumptions has been left intact, but some of the particular parameters have been modified to more realistically model the network data.

REFERENCES

[1] American National Standard, Specifications and Guidelines for Quality Systems for Environmental Data Collection and Environmental Technology Programs, ANSI/ASQC E4-1994, American Society for Quality Control, 1994

[2] Guidance for the Data Quality Objectives Process, U.S. Environmental Protection Agency, Quality Assurance Management Staff, EPA QA/G-4, March 14, 1994.

[3] Guidance for the Data Quality Assessment Process EPA QA/G-9, U.S. Environmental Protection Agency, QAD EPA/600/R-96/084, July 1996.

[4] Rhodes, R.C., "Guideline on the Meaning and Use of Precision and Accuracy Data Required by 40 CFR Part 58, Appendices A and B." EPA60014-83-023, U.S. Environmental Protection Agency, Research Triangle Park, N.C. 27711, June 1983.

[5] "Model Quality Assurance Project Plan for the $PM_{2.5}$ Ambient Air Monitoring Program at State and Local Air Monitoring Stations (SLAMS)," Quality Assurance Guidance Document, U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, EPA 454/R-98-005, April 1998.

APPENDIX A:

ASSUMPTIONS AND INPUT CRITERIA FOR THE DQOs FOR AMBIENT AIR MONITORING OF $PM_{2.5}$

APPENDIX A: ASSUMPTIONS AND INPUT CRITERIA FOR THE DQOS FOR AMBIENT AIR MONITORING OF $PM_{2.5}$

The DQO process clarifies the monitoring objectives of a network to ensure that the data collected is of the appropriate type, quantity, and quality to meet program goals. The process necessarily makes various assumptions about the nature of the data to be collected in order to quantitatively bridge the decision goals with the required quantity and quality of monitoring data. A key part of the DQA process is then to assess these assumptions, once monitoring data are available. AIRS $PM_{2.5}$data from 1999 and 2000 were used to review the assumptions made in developing DQOs for the $PM_{2.5}$ NAAQS compliance. The NAAQS $PM_{2.5}$ standards are met if:

1. The 3-year average of the annual arithmetic means of the daily $PM_{2.5}$ concentrations is less than or equal to 15 micrograms per cubic meter; and
2. The 3-year average 98th percentile of the daily $PM_{2.5}$ concentrations is less than or equal to 65 micrograms per cubic meter.

Both the original DQO process and an outline for the DQA are documented. EPA developed a model Quality Assurance Project Plan QAPP for states to follow. This included a general DQO development that was intended to be used nationwide to provide for uniform data quality. Section 7 of the Model QAPP documents the DQOs and Section 24 outlines an assessment plan for checking some of the assumptions made in the DQO development. Included in this outline is a list of seven assumptions and input criteria statements used in the DQO development. They are enumerated in the Model QAPP as:

1. The DQO is based on the annual arithmetic mean NAAQS.
2. Normal distribution of measurement error.
3. Decision error can occur when the estimated 3-year average differs from the actual, or true, 3-year average.
4. The limits on the precision and bias are based on the smallest number of required sample values in a 3-year period.
5. The decision error limits were set at 5 percent.
6. Measurement imprecision was established at 10 percent coefficient of variation (CV).
7. Achievement of bias and precision limits.

Further examination of the DQOs shows that there are some additional assumptions that should be verified. (Or verify that there is negligible impact from making an incorrect assumption.) These are enumerated with some comments below.

1. The DQO is based on the annual arithmetic mean NAAQS.

   The review of the 1999 and 2000 data shows that annual arithmetic mean is the more stringent requirement for the network. However, since the main calculations in the DQO development are done via Monte Carlo simulation, it is not necessary to make this assumption. Using Monte Carlo simulation, DQOs for the 98th percentile could be included as well. This was not done because, compared to the annual mean, the quality of the estimate of the 98th percentile will be much more dependent on the distributional assumptions made in the simulations.

2. Normal distribution of measurement error.

   The statement and the suggested check in the Model QAPP is concerned only with the distributional nature of the measurement error. There is an additional implied assumption that is directly tied to this assumption, namely:

   > The measurement error standard deviation is assumed to be proportional to the true concentration..

3. Error can occur when the estimated 3-year average differs from the actual, or true, 3-year average.

   This is a fundamental assumption of the process. The assumption is that the methods described for calculating the error rate are realistic.

4. The limits on the precision and bias are based on the smallest number of required sample values in a 3-year period. In particular, it is assumed that 75 percent of 1-in-6-day sampling is both sufficient and attainable.

   It is also assumed that the missing values are completely at random. It is, of course, assumed that the missing values are independent of the daily value. However, the missing data could be random, but clustered because the completeness requirement is applied quarterly, any clustering is unlikely to have a significant impact.

5. The decision error limits were set at 5 percent.

   This is an input criteria needed to carry out the calculations being made. It is not necessary that both the false positive error rate and the false negative error rate be the same, but this was the case that was chosen by decision makers.

6. Measurement imprecision was established at 10 percent coefficient of variation (CV).

   This input criteria statement has the implied assumption is that a 10 percent CV is attainable. (And, as noted above, it makes sense to measure the precision in terms of a percent of the mean.)

7. Achievement of bias and precision limits.

   As with Assumption 6, the DQO was developed assuming that an absolute bias of 10 percent was achievable. There is also the underlying assumption that the bias should be measured as a percent of the mean rather than as an absolute bias. The DQOs are developed assuming a worst case scenario with respect to the bias. (The bias is always assumed to be in the direction that would result in a decision error.)

8. There is a mean seasonal pattern to the PM concentrations that can be adequately described by a sinusoidal curve such that the ratio of the high to the low in this pattern is 5.3 (this is assumed as a worst case).

   The original value of 5.63 was reduced to 5.3 based on the review of the 1999 and 2000 data. (See Section 2.7 and Table 1 in the main text.)

9. The random population variation about the mean seasonal curve is log-normal with a standard deviation that is proportional to the seasonal mean.

   The original DQO development assumed normally distributed variation. This was changed to lognormal to more realistically model the data in the Monte Carlo simulations.

10. The random population variation about the mean seasonal curve has a CV of at most 80 percent.

    This was increased from 50 percent based on the review of the 1999 and 2000 data. (See Section 2.7 and Table 2 in the main text.)

11. The precision and bias will be constant throughout the 3-year period (or at least consistently within their respective limits).

12. The precision and bias are checked sufficiently often to detect significant deviations from the DQOs.

13. The day-to-day values (of the truth) are assumed to be independent.

The review of the network data found that this is NOT true across the network. However, it is nearly true; the minimum correlation between daily values is about 0.3. To model the worst case for the decision maker, values close to 0 (=uncorrelated) should be used. Incorporating a small amount of correlation into the DQO modeling has very little effect on 1-in-6-day sampling. The effect would be more pronounced on a DQO based on 1-in-3-day sampling or daily sampling.

14. The measurement error is assumed to be independent from day-to-day.

This is much the same as saying that the bias is consistent. In particular, it is assumed that the sampler will not be high for a couple of months, then low for a while, and average out to some acceptable bias.

Attachment C. DQO Guidance for AQI Applications

See: http://www.epa.gov/ttn/amtic/files/ambient/monitorstrat/aqidqorept.pdf

ATTACHMENT D - Draft National Monitoring Strategy

See: http://www.epa.gov/ttn/amtic/stratdoc.html