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%.

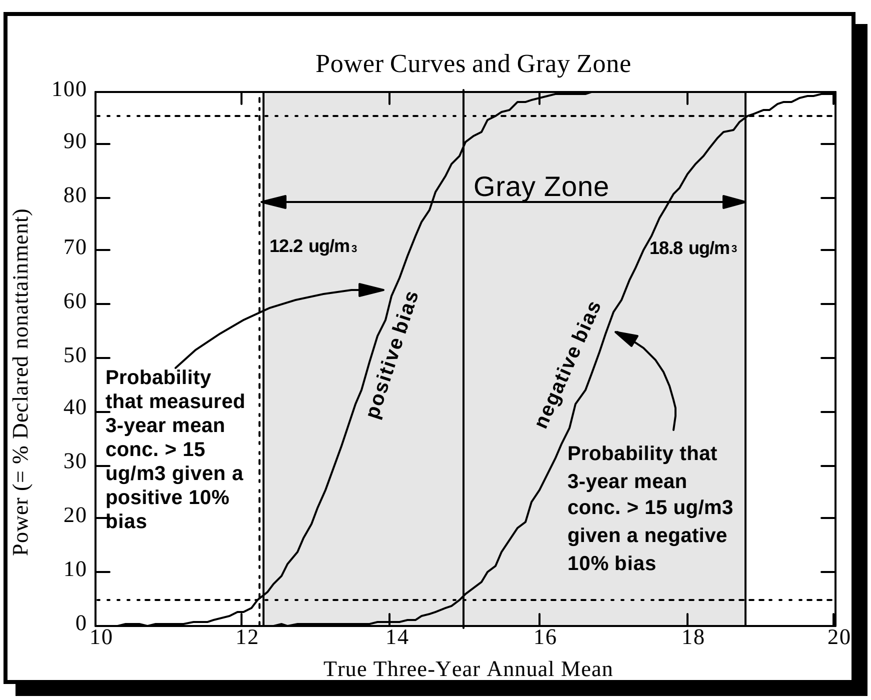

Figure 6.1 Power curve based on 2001 assumptions

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.

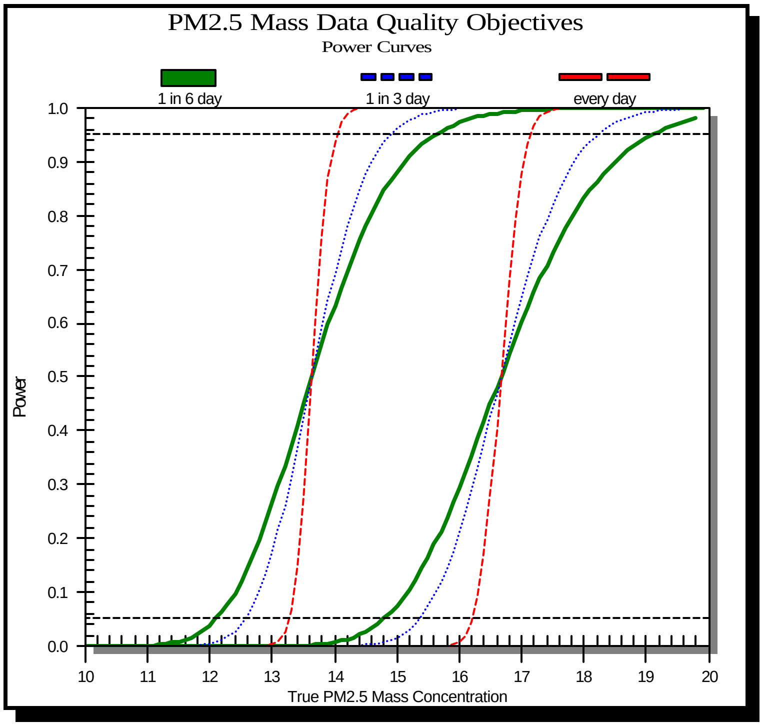

Figure 6.2. Power curve changes due to changes in sampling frequency

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 $\mu g/m^3$ (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 $\mu g/m^3$ (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

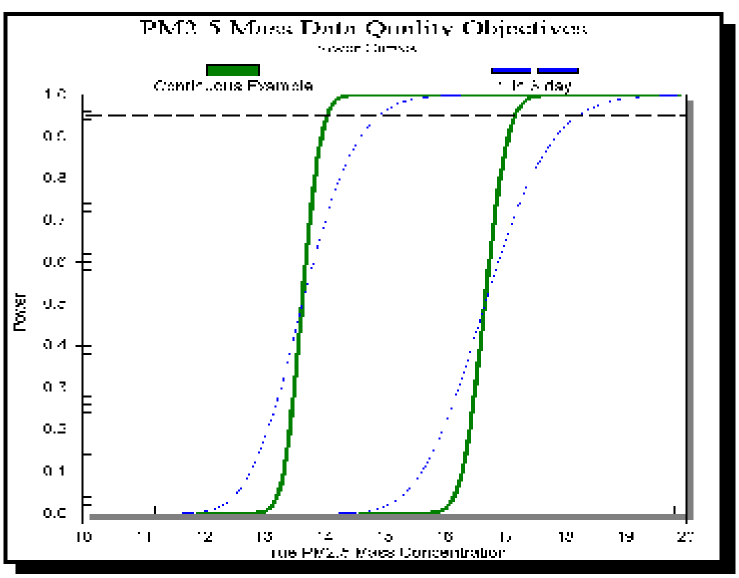

Figure 6.3. Example continuous monitoring network power curve relative to 1 in 3 day power curve.

- 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 $\mu g/m^3$ (lower left line green solid) and 17.1 $\mu g/m^3$ (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 µg/m^3 when concentration is <40 µg/m^3 (24 hour sample) or <30 µg/m^3 (48 hour sample); Rpj <5% for concentration > 40 µg/m^3 (24 hour sample) or >30 µg/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.)