Attachment 6

Technical Report

on

$PM_{10-2.5}$ METHOD EQUIVALENCY DEVELOPMENT

by

Office of Air Quality Planning and Standards
Office of Research and Development
U.S. ENVIRONMENTAL PROTECTION AGENCY
Research Triangle Park, North Carolina 27711

BATTELLE
505 King Avenue
Columbus, Ohio 43201-2693

Attachment 6

DISCLAIMER

This report is a work prepared for the U.S. Environmental Protection Agency by Battelle. In no event shall either the U.S. Environmental Protection Agency or Battelle have any responsibility or liability for any consequences of any use, misuse, inability to use, or reliance upon the information contained herein, nor does either warrant or otherwise represent in any way the accuracy, adequacy, efficacy, or applicability of the contents hereof.

TABLE OF CONTENTS

List of Tables

List of Figures

EXECUTIVE SUMMARY

The U.S. Environmental Protection Agency (EPA) staff are currently developing Data Quality Objectives (DQOs) in anticipation of a potential National Ambient Air Quality Standard (NAAQS) coarse particulate matter ($PM_{10-2.5}$). By definition, $PM_{10-2.5}$ is the difference between two concentrations, PM_{10} minus $PM_{2.5}$. Hence, the EPA is considering a $PM_{10-2.5}$ reference method based on separate low-volume Federal Reference Method (FRM) measurements of PM_{10} and $PM_{2.5}$ and calculating $PM_{10-2.5}$ by difference. However, these methods are labor intensive and, hence, costly to operate. Consequently, EPA is developing method equivalency criteria for various continuous measurement methods concurrently with its DQO development.

The objective of the equivalency criteria development is to develop standards for comparing candidate measurement methods ("continuous" and/or "direct" methods) to the reference measurement methods. The results of the candidate-to-reference-method comparison need to reasonably insure that, when local agencies use these methods for making comparisons to the NAAQS, the decision quality is as good as if they had used filter-based reference methods.

This does not mean that the candidate methods need to meet the same DQO criteria developed for the reference methods. The differences in the technologies involved yield different strengths and weaknesses. There are two key differences: (1) the candidate methods can operate at a much higher sampling frequency and (2) they can operate with an effective completeness that is likely to be higher than reference methods. These two factors strongly influence the width of the gray zone (a means of measuring decision quality) [1,2]. Consequently, the continuous methods can be allowed relaxed standards for the precision and bias.

It is anticipated that the equivalency requirements will be based on collocated sampling of several candidate method samplers and several reference method samplers for month-long periods at several sites in different seasons. The criteria, however, will be applied for each site-month combination. In particular, the candidate samplers will have to meet criteria for their precision, correlation with the reference method, and additive and multiplicative bias as found through regression.

Since the DQOs have not been established, the results of the current effort cannot be explicitly stated. Instead, the DQO software tool [1,2] for $PM_{10-2.5}$ has been modified to aid in establishing the criteria for making the appropriate tradeoffs. Examples generated with this tool show that reasonable criteria can be set following the methods used for establishing equivalency requirements for $PM_{2.5}$ [3]. The example shown works well for candidate samplers with a measurement coefficient of variation of 15 percent, a multiplicative bias of no more than 15 percent, a correlation of 0.95 to 0.97 (depending characteristics of the concentrations sampled), and additive biases of about 5 $\mu g/m^3$.

Revised Technical Report
on
$PM_{10-2.5}$ Method Equivalency Development

WA 3-11, Tasks 2 and 3

1.0 INTRODUCTION

The U.S. Environmental Protection Agency (EPA) staff are currently developing Data Quality Objectives (DQOs) in anticipation of a potential National Ambient Air Quality Standard (NAAQS) coarse particulate matter ($PM_{10-2.5}$). By definition, $PM_{10-2.5}$ is the difference between two concentrations, PM_{10} minus $PM_{2.5}$. Hence, the EPA is considering a $PM_{10-2.5}$ reference method based on separate low-volume Federal Reference Method (FRM) measurements of PM_{10} and $PM_{2.5}$ and calculating $PM_{10-2.5}$ by difference. However, these methods are labor intensive and, hence, costly to operate. Consequently, EPA is developing method equivalency criteria concurrently with its DQO development.

The objective of the equivalency criteria development is to develop standards for comparing candidate measurement methods ("continuous" and/or "direct" methods) to the reference measurement methods. The results of the candidate to reference method comparison need to reasonably insure that, when local agencies use these methods for making comparisons to the NAAQS, the decision quality is as good as if they had used filter-based reference methods.

This does not mean that the candidate methods need to meet the same DQO criteria developed for the reference methods. The differences in the technologies involved yield different strengths and weaknesses. There are two key differences: (1) the candidate methods can operate at a much higher sampling frequency and (2) they can operate with an effective completeness that is likely to be higher than reference methods. These two factors strongly influence the width of the gray zone (a means of measuring decision quality) [1,2]. Consequently, the continuous methods can be allowed less stringent standards for the precision and bias.

The DQO software tool [1,2] for $PM_{10-2.5}$ has been modified to aid in establishing the criteria for making the appropriate tradeoffs. A software tool is needed because there are still "decision maker" level choices yet to be made that will finalize the DQO. This document describes how and what the tool does.

2.0 THE ASSSUMED FORM OF THE EQUIVALENCY CRITERIA

A fixed form for the equivalency requirements was assumed in the development of the software tool. The assumptions come from the development of equivalency requirements for $PM_{2.5}$ measurements by continuous methods [3]. In fact, the entire process completely parallels that development.

2.1 The Data Assumptions

To determine the equivalency of a candidate continuous sampler (uniquely specified by manufacturer, brand, and model number) relative to the reference method, daily concentration data need to be obtained from samples collected from collocated candidate and reference method samplers at multiple sites. To support an equivalence evaluation, the data collection process involving collocated samplers needs to adhere to the following requirements:

- Three (3) to five (5) candidate samplers will be collocated with three (3) reference method samplers. Each "sampler" in this context needs to be a complete unit capable of generating a $PM_{10-2.5}$ measurement. For measurements based on a difference of two measurements, the PM_{10} and $PM_{2.5}$ samplers need to be paired. (The statistics below are not applicable unless the number of $PM_{2.5}$ samplers is the same as the number of PM_{10} samplers.) This number of samplers is consistent with existing requirements and improves on the ability to identify statistical outliers in daily concentrations.

- Within a given season of the year, each sampler will be run daily for a target of 30 days (with at least one site having samples collected in multiple seasons).

- On a given day, the required sample collection period for each sampler will be sufficient to be representative of 24-hour sampling. (A minimum of 22 hours is used for the $PM_{2.5}$, and a similar requirement is assumed here.)

- On a given day, valid data must be available for at least two (2) reference method samplers and at least two (2) candidate samplers in order for any data associated with the day's sample collection to be used in the equivalency evaluation.

- Each sampler at a given site will produce valid measurements on at least 75 percent of the sampling days in a given season. For a 30-day sampling period, this corresponds to a minimum of 23 days per season.

- The acceptable concentration range of sample data would be set to be both representative of coarse sampling and limit outliers. (For example, an upper bound should be set that is sufficiently small to avoid any data irregularities from filter clogging.)

Data collection will be replicated at multiple sites to ensure that the sampling is representative of different aerosol types. Furthermore, for at least one site, sampling will occur in at least two distinct seasons of the year. The above sampling requirements will hold across seasons for each site. The above sampling requirements, however, apply to each site and season. The recommendation for the total number of sites has not been established at the time of this report.

2.2 Assumed Calculations in the Equivalency Criteria

From the daily sample concentration data to be collected from the collocated samplers at a given site, the following four measures will be calculated:

- Precision
- Correlation
- Multiplicative bias
- Additive bias

A candidate sampler needs to achieve specified criteria placed on each of these four measures in order to be classified as equivalent to the reference method. Values for these four measures are calculated separately for each site, and the candidate sampler needs to achieve the specified criteria at each site.

The explicit formulas presented below specify how each of these four measures is calculated. The next section deals with how to use the tool to develop acceptance criteria for these measures. In calculating the four equivalency measures, true daily $PM_{10\text{-}2.5}$ concentrations at a given site are estimated from the daily means associated with the reference method samplers.

Precision: The precision associated with the candidate sampler data is calculated as:

$$Cand\_prec = \sqrt{\frac{\sum_{i=1}^{D}\left(Cand\_daily\_CV_i\right)^2}{D}} \quad . \qquad \text{(Eq. 1)}$$

where the summand is the daily coefficient of variation among the candidate samplers and D is the number of days with valid data.

Correlation: Correlation in the daily means between the FRM and candidate samplers is calculated as follows:

$$r = \frac{D\sum_{i=1}^{D}\left(REF\_daily\_mean_i \cdot Cand\_daily\_mean_i\right) - \left(\sum_{i=1}^{D} REF\_daily\_mean_i\right)\cdot\left(\sum_{i=1}^{D} Cand\_daily\_mean_i\right)}{D^2 \cdot Cand\_RMS \cdot REF\_RMS} \qquad \text{(Eq. 2)}$$

Multiplicative bias: The multiplicative bias is the slope of the ordinary least-squares line between the daily means of the candidate and reference method samplers. It is calculated as the correlation (Equation 2) multiplied by the ratio of the root-mean-square deviations from the overall means of the daily means for the candidate and reference method samplers:

$$b = r \cdot \frac{Cand\_RMS}{REF\_RMS} \quad . \qquad \text{(Eq. 3)}$$

Additive bias: The additive bias is the intercept of the ordinary least-squares line between the daily means of the candidate and reference method samplers. It is dependent on the daily means associated with the candidate sampler, the overall mean (associated with the reference method samplers), and the calculated value for multiplicative bias (Equation 3). The formula for additive bias is as follows:

$$a = \left(\frac{1}{D} \cdot \sum_{i=1}^{D} Cand\_daily\_mean_i \right) - b \cdot Overall\_mean \qquad \text{(Eq. 4)}$$

3.0 SETTING THE CRITERIA

There are several steps to developing the criteria for the four measures. The DQO software tool is critical to two of the four. To aid in the discussion, an explicit example is used throughout Section 3.

3.1 Precision and Multiplicative Bias

The requirements for the precision and multiplicative bias need to be based, in part, on what is generally achievable for the class of methods being considered as candidates. Setting the requirements to be slightly less than the requirements in the DQOs may be acceptable. In the example shown below, the DQO requirements are set at 10 percent for each and the requirements for the candidate samplers are set at 15 percent. As these are made less restrictive, the requirements for the other sections will become more restrictive. Some sensitivity testing is recommended.

3.2 The DQO Example

To help with descriptions in the remaining sections, an explicit DQO scenario is used. The specific parameter settings shown in Table 1 are generally reasonable choices for producing a worst-case scenario. Except as noted in Table 1, the parameters have been estimated from historical data for sites across the nation. Then for each parameter, a near “worst case” value was chosen. In this way, the decision error rate is controlled across the nation. Individual sites will generally have better performance than is indicated by these curves. Figure 1 shows the performance curve associated with these parameters.

Table 1. Parameter settings for the DQO example.

Standard		
	Level	**Percentile**
Daily Standard	60 µg/m^3	98th
PM Fraction Characteristics		
	$PM_{10-2.5}$	**$PM_{2.5}$**
Seasonality ratio	14	5.3
Population CV	1	0.8
Autocorrelation	0	0
Global Characteristics		
	Setting	
Phase shift	0	
$PM_{2.5}$ to $PM_{10-2.5}$ correlation	0	
Mean $PM_{2.5}$ / mean $PM_{10-2.5}$	0.45	
$PM_{10-2.5}$ Periods per year	1	
$PM_{10-2.5}$ Spatial sill	1	
$PM_{10-2.5}$ Spatial Range	20	
Sampling frequency	3	
Measurement Error Characteristics		
	PM_{10}	**$PM_{2.5}$**
Bias	0.1	0.1
Measurement CV	0.1	0.1
Completeness	0.75	0.75
Output		
	Lower Bound	**Upper Bound**
Daily Gray Zone	37.7 µg/m^3	95.6 µg/m^3

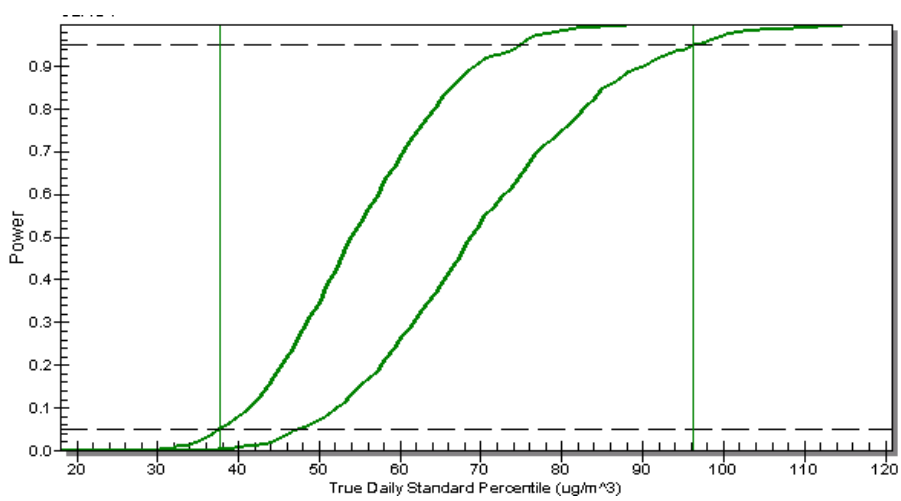

Figure 1. The decision performance curve for the DQO example.

3.3 The Alternate Scenario

The next step in establishing the equivalency requirements is to establish an alternate scenario that is more representative of the candidate methods. Again, an explicit example is chosen for illustrative purposes. The parameter settings are shown in Table 2 and the corresponding performance curve is shown in Figure 2.

Table 2. The alternate parameter settings.

Standard		
	Level	**Percentile**
Daily Standard	60 µg/m^3	98th
PM Fraction Characteristics		
	$PM_{10-2.5}$	**$PM_{2.5}$**
Seasonality ratio	14	5.3
Population CV	1	.8
Autocorrelation	0	0
Global Characteristics		
	Settings	
Phase shift	0	
$PM_{2.5}$ to $PM_{10-2.5}$ correlation	0	
Mean $PM_{2.5}$ / mean $PM_{10-2.5}$	0	
Periods per year	1	
Spatial sill	1	
Spatial Range	20	
Sampling frequency	1	
Measurement Error Characteristics		
	PM_{10}	**$PM_{2.5}$**
Bias	0.15	0
Measurement CV	0.15	0
Completeness	0.75	1
Output		
	Lower Bound	**Upper Bound**
Daily Gray Zone	44.8 µg/m^3	79.8 µg/m^3

Notes:
1. The ratio of the mean $PM_{2.5}$ / mean $PM_{10-2.5}$ has been set to zero so that the simulated PM_{10} concentrations contained no $PM_{2.5}$ fraction.
2. The measurement characteristics for the $PM_{2.5}$ have been set to 0 errors and 100 percent completeness so that the only simulated measurement errors are that of the $PM_{10-2.5}$.

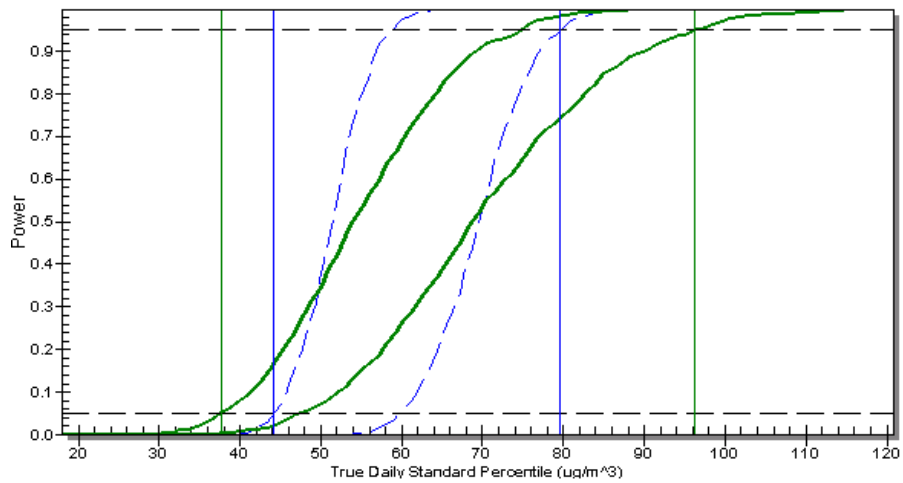

Figure 2. Decision performance curves for both the DQO case (dashed) and the alternate scenario (solid).

It is important to note that the gray zone for the alternate scenario is strictly contained in the gray zone for the DQO case. This is needed because the additive bias is computed as the amount of additive bias needed for the alternate case to yield a gray zone equivalent to the DQO case. If the gray zone for the alternate case is not strictly within the DQO case, then the case with zero additive bias will not be in the acceptable range. This is clearly undesirable.

3.4 Setting the Correlation Requirement

The correlation measures how well the candidate sampler's concentrations vary linearly with the reference method's concentrations. The formula for the expected correlation between measurements with constant measurement CVs is derived by Mosquin, et al., in References [4,5]. The expected correlation is dependent on the coefficient of variation of the concentrations measured. Experience with $PM_{2.5}$ has shown that this is a fairly constant characteristic of a site and it is assumed to be true for $PM_{10\text{-}2.5}$ as well in the DQO model. Historical data should be considered in the selection of sites for testing. As with the $PM_{2.5}$ equivalency criteria, the acceptance criteria should vary with the measured coefficient of variation of the daily means from the reference method.

The software tool shows the expected correlation along with an approximate lower bound that would be expected [6]; both the expected correlation and the lower bound are dependent on the number of samplers used. The tool assumes three reference samplers with a measurement CV of 7.5 percent. The number of candidate samplers, their measurement CV, and the number of sample-days used in the comparison are inputs. Figure 3 shows an example based on five candidate samplers, 23 sample days, and the 15 percent measurement CV used in the alternate scenario.

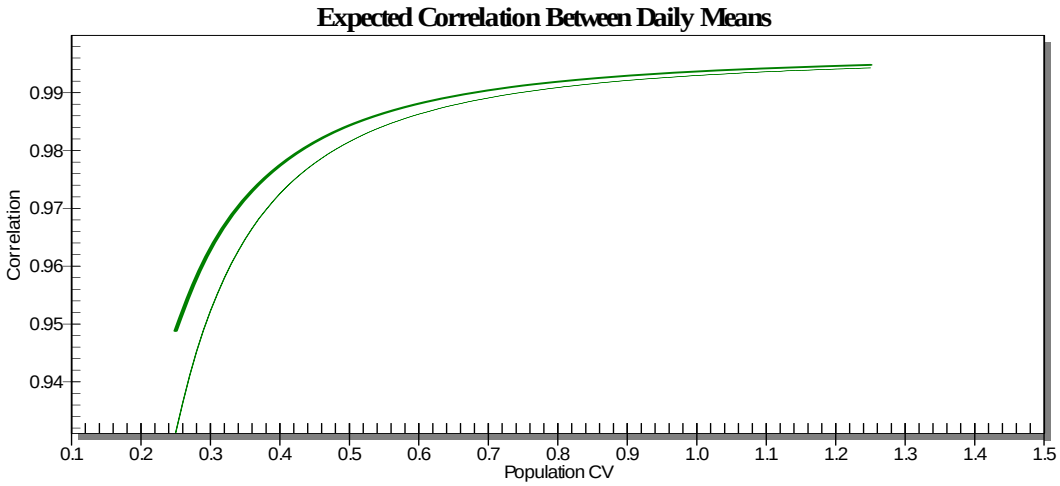

Figure 3. The expected correlation (solid) and approximate lower bound (dashed) for the alternate scenario, 23 sample days, and five candidate samplers.

3.5 Setting the Additive Bias Requirement

As in the case for the $PM_{2.5}$ equivalency requirements, the acceptable amount of additive bias can be made dependent on the multiplicative bias. The tool computes the amount of additive bias that could be allowed for the alternate scenario to obtain a user-defined target gray zone. The additive bias criteria should be at least as restrictive as requiring the point corresponding to the additive and multiplicative biases to be in the parallelogram generated by the tool.

In the example shown in Figure 4, the target gray zone is the gray zone for the DQO. This does not need to be the case. The target gray zone could be chosen to be slightly more restrictive to allow for the case where local agencies may end up operating the candidate samplers with slightly more bias than is found in the equivalency testing.

Finally, as noted at the end of Section 3.3, the target gray zone should strictly contain the gray zone for the alternate case. The tool gives a warning if the values entered do not meet this criterion, but displays the parallelogram. In cases when this happens either the left-hand side of the parallelogram is completely below zero or the right-hand side is completely above zero (or both). This “problem” can be avoided by restricting the multiplicative bias more than has already been done.

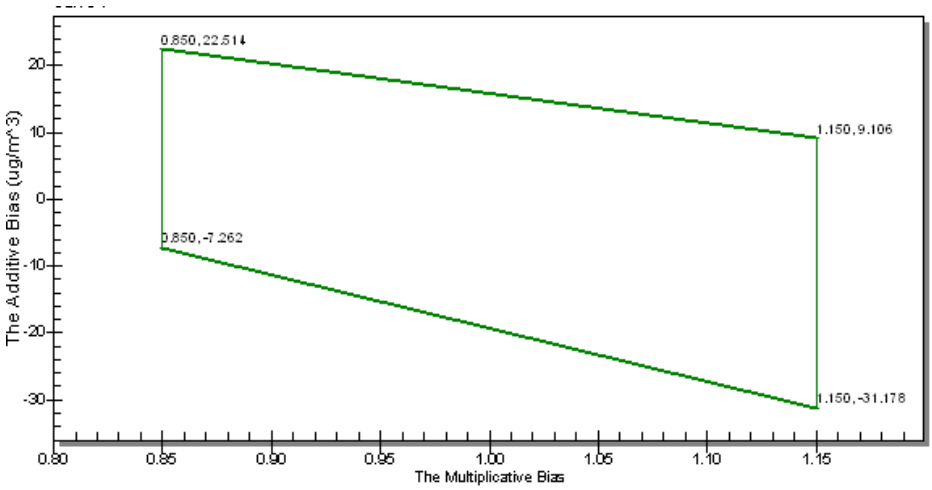

Figure 4. The acceptance range for the additive and multiplicative bias corresponding to the DQO example and the alternate scenario.

4.0 SUMMARY

It is anticipated that the equivalency requirements will be based on collocated sampling of several candidate method samplers and several reference method samplers for month-long periods at several sites in different seasons. The criteria, however, will be applied for each site-month combination. In particular, the candidate samplers will have to meet criterion for their precision, correlation with the reference method, and additive and multiplicative bias as found through regression.

Since the DQOs have not been established, the results of the current effort cannot be explicitly stated. Instead, the DQO software tool for $PM_{10\text{-}2.5}$ has been modified to aid in establishing the criteria for making the appropriate tradeoffs. Examples generated with this tool show that reasonable criteria can be set following the methods used for establishing equivalency requirements for $PM_{2.5}$. The example shown works well for a continuous candidate sampler with a measurement coefficient of variation of 15 percent, a multiplicative bias of no more than 15 percent, a correlation of 0.95 to 0.97 (depending characteristics of the concentrations sampled), and additive biases of about 5 $\mu g/m^3$.

5.0 REFERENCES

[1] Coutant, B.W., Morara, M., and Boehm, R.C. (2003). "DQO Companion for PMcoarse." Software developed for the U.S. EPA, Office of Air Quality Planning and Standards, under EPA Contract No. 68-D-02-061, May.

[2] Coutant, B.W. (2005). "Coarse PM Data Quality Objective Development." Technical Report to the U.S. EPA, Office of Air Quality Planning and Standards, under EPA Contract No. 68-D-02-061, August.

[3] Coutant, B.W., and Sanford, J.D. (2004). "Criteria for Designation of Equivalence Methods for Continuous Surveillance of $PM_{2.5}$ Ambient Air Quality." Technical Report to the U.S. EPA, Office of Air Quality Planning and Standards, under EPA Contract No. 68-D-02-061, September.

[4] Mosquin, P., and McElroy, F. (2004). "Data Quality Objectives for PM Continuous Methods II." ManTech technical report to the U.S. EPA, report number TR-CAN-04-02, June.

[5] Mosquin, P., McElroy, F., Clayton, A., and Vanderpool, B. (2003). "Data Quality Objectives for PM Continuous Methods." ManTech technical report to the U.S. EPA, report number TR-4423-03-08, June.

[6] Casella, G., and Berger, R.L. (1990). Statistical Inference. Belmont, California: Duxbury Press.