Attachment 5

Technical Report

on

Sensitivity of the $PM_{10\text{-}2.5}$ Data Quality Objectives to Spatially Related Uncertainties

Attachment 5

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

Sensitivity of the $PM_{10-2.5}$ Data Quality Objectives to Spatially Related Uncertainties

EXECUTIVE SUMMARY

At the July 22, 2004 Clean Air Scientific Committee (CASAC) meeting, the use of a data quality objectives (DQO) approach was presented as it related to developing the appropriate measurement quality objectives for the $PM_{10-2.5}$. DQOs are qualitative and quantitative statements that help define the appropriate type of data, and specify the tolerable levels of data uncertainty. A detailed report of that work titled: *Use of a Performance Based Approach to Determine Data Quality Needs of the PMcoarse (PMc) Standard* can be found at the following website: http://www.epa.gov/ttn/amtic/casacinf.html

Using some of the same techniques that were used to develop DQOs for fine particulate NAAQS ($PM_{2.5}$), the U.S. Environmental Protection Agency (EPA) had a DQO software tool developed that provided decision makers with an understanding of the consequences of various uncertainty input parameters such as sampling frequency, data completeness, precision, and bias, and how these uncertainties affect our confidence in concentration estimates. Since both manual and continuous (automated) methods may be available for use in estimating the coarse particulate fraction, the DQO software tool was developed to address both manual and continuous methods and is useful in weighing the benefits and disadvantages of these methods.

The DQOs were developed using data collected from sites providing coarse particulate estimates from around the country as well as data from multi-site performance evaluations conducted by the EPA National Exposure Research Laboratory (NERL). These data provided estimates of reasonable sample population and measurement uncertainty input parameters that were used to generate the DQO performance curves (gray zones).

In general, EPA received positive feedback on the DQO approach. Some specific comments were accepted and implemented; others required a more detailed assessment. Two comments that were brought up at the meeting, in addition to the submissions from a number of committee members, were to look at the effects of spatial variability and multi-modal distributions. This report presents the techniques that were used to address these two issues, how they were incorporated into the DQO tool, and how these components of variability might affect the performance curves.

Preliminary performance curves were assessed for their sensitivity to the input parameters. The assessment found that for the daily standard, the performance curves were most sensitive to sampling frequency, followed by the completeness, the population CV of the coarse fraction of the particulate matter, and the ratio of the mean concentrations between the coarse and fine fractions of the particulate matter. The effect of multi-modal distributions was very small. The effect of the spatial variability is small compared to the parameters mentioned above, but we suggest including this parameter in the DQO evaluation.

1.0 INTRODUCTION

At the July 22, 2004 Clean Air Scientific Committee (CASAC) meeting, the use of a data quality objectives (DQO) approach was presented as it related to developing the appropriate measurement quality objectives for the $PM_{10-2.5}$. DQOs are qualitative and quantitative statements that help define the appropriate type of data, and specify the tolerable levels of data uncertainty. A detailed report of that work titled: *Use of a Performance Based Approach to Determine Data Quality Needs of the PMcoarse (PMc) Standard* can be found at the following website: http://www.epa.gov/ttn/amtic/casacinf.html

Using some of the same techniques that were used to develop DQOs for fine particulate NAAQS ($PM_{2.5}$), the U.S. Environmental Protection Agency (EPA) had a DQO software tool developed [1] that provided decision makers with an understanding of the consequences of various uncertainty input parameters such as sampling frequency, data completeness, precision, and bias, and how these uncertainties affect our confidence in concentration estimates. Since both manual and continuous (automated) methods may be available for use in estimating the coarse particulate fraction, the DQO software tool was developed to address both manual and continuous methods and is useful in weighing the benefits and disadvantages of these methods.

The DQO development used preliminary data collected from sites providing coarse particulate estimates from around the country as well as data from multi-site performance evaluations conducted by the EPA National Exposure Research Laboratory (NERL). These data provided estimates of reasonable input parameters that were used to generate the DQO performance curves (gray zones). [2]

In general, EPA received positive feedback on the DQO approach. Some specific comments were accepted and implemented; others required a more detailed assessment. Two comments that were brought up at the meeting, in addition to the submissions from a number of committee members, were to look at the effects of spatial variability and multi-modal distributions. This report presents the techniques that were used to address these two issues, how they were incorporated into the DQO tool, and how these and other components of variability might affect the performance curves.

2.0 THE PREVIOUS DQO TOOL

The previous DQO tool for $PM_{10-2.5}$ created NAAQS decision performance curves based on a simulation model. The model simulated the decisions based on three-year periods assuming NAAQS standards that would be similar in form to the $PM_{2.5}$ standards (i.e., an annual and a daily standard based on a particular percentile).

Since a difference method was likely to be used (at least as a reference), the basic model simulates true and observed $PM_{2.5}$ and PM_{10} concentrations. The true PM_{10} concentrations are modeled as a sum of the fine fraction and the coarse fraction. These two fractions are simulated with sinusoidal means (to represent seasonality) with distinct long-term means and with random deviations. The seasonal patterns can be phase shifted. The random deviations have distinct (but constant) coefficients of variation (CV) which can be correlated and/or autocorrelated.

The $PM_{2.5}$ and PM_{10} measurements have separate biases and measurement CVs. Negative differences are set equal to zero. (The simulated true coarse PM concentrations are all strictly greater than zero.) A common sampling frequency is assumed (which would be the least frequent if the two were not on the same frequency). Finally, fixed quarterly completeness criteria are assumed and implemented independently of each other (a 75 percent completeness criterion for both measurements corresponds to about 56 percent completeness).

In the tool, all of the characteristics mentioned above correspond to user-supplied parameters. The preliminary data was assessed to establish realistic ranges for each of the parameters (see Table 2.).

3.0 ISSUES ADDRESSED

This section describes the process used to address the issues of multi-modal distributions and spatial variability. The multi-modal distributions issue was fairly explicitly defined by the CASAC review and could be readily addressed with changes to the basic DQO simulation model. The spatial variability issue required some exploratory data analyses to first establish how much spatial variability should be incorporated into the model. The need for additional data analyses of the historical data prompted a review and revision of the database before these analyses were conducted.

3.1 Updating the Historical Database for Unified Analyses

One of the issues noted with the previous analyses was that the historical data used included measurements from a variety of methods (i.e., both high and low volume methods). A key point that should be noted is that the analyses were restricted to estimating the possible ranges of population characteristics; they were not used for estimating measurement error characteristics. Regardless, a new database was created with particular regard to the methods that had been used to take the measurements. This single database is now being used for DQO parameter estimates.

The database still contains the mix of analytic methods, but the data have been flagged for relative data quality. The lower quality measurements are still needed for analyses that require extensive amounts of data.

3.2 Estimating the Spatial Variability from Historical Data

The effect of spatial variability on the DQO process was a priority for investigation this year. The first step in this was to develop a spatial variability model for coarse PM and to estimate the associated parameters to be used in the new simulation model. An exponential spatial model for the variability was assumed based on the performance of that model with ozone and $PM_{2.5}$ data. This model has two parameters: the sill, σ, and the range, θ, which describe the covariance between points based on the distance between them, $\sigma^2 \exp(-\text{distance}/\theta)$.

While there are ample data for understanding temporal variability, there are very limited data available for investigating spatial variability. Estimates of spatial variability are often limited to cases were there are 30 or more data points, in part because the estimation methods may not converge for small data sets. To achieve this, monitors were grouped by state and date, with the hope that the range parameters could be extrapolated reliably for most states. Consistency across dates was also looked for in evaluating the results. The range parameters were found to be between 20 and 125 km and the sill between 60 and 100 percent of the overall variability.

3.3 Redesigning the DQO Simulation Model

The next step was to redesign the DQO simulation model to incorporate the spatial variability and multiple high periods in the year for testing the effects of these changes on the gray zone. Program details are given in the document titled: *The $PM_{10\text{-}2.5}$ DQO Model* that can be found at the following website: http://www.epa.gov/ttn/amtic/casacinf.html

The previous model only simulated temporal variability. The model (before adding measurement uncertainty) for both the fine and coarse fractions is simulated as random deviations with a constant coefficient of variation from a sinusoidal mean, with a period of one year. These were allowed to be phase shifted from each other. The (log-normal) random deviations could be independently autocorrelated and/or correlated with each other.

In the new model, the mean for the coarse fraction can have either one or two periods. Depending on the other parameters, this will cause the simulated PM_{10} concentrations to have more than one high period in the year. Phase shifting the fine and coarse fractions can also achieve that effect.

The new model also simulates spatial variability. The spatial domain is simulated on a grid (Figure 1) that is supposed to represent a neighborhood scale monitor with grid points located every 2 km in a square grid that is 8 km on a side. The temporal variation also includes a seasonal pattern throughout the three-year simulation periods. For each grid point, separate three-year design values are computed. Then the grid's design values (one for the annual standard and one for the daily standard) are computed by averaging over the grid-point specific design values. Finally, this design value is compared to the design value for the simulated observations corresponding to the center point. This whole process is repeated 1,000 times to determine the probability that the observed-design value correctly predicts whether or not the grid's true-design value is above the standard.

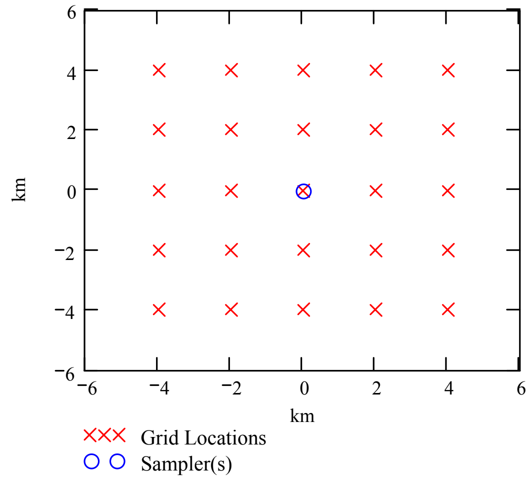

Figure 1. Simulation grid points.

Hence, the comparison is indicating how well a single monitor does in predicting the true mean design value across the grid area. Since the day-to-day shape of the surface is not fixed, on average, throughout the three-year period, the center should be an unbiased indicator of the mean. Consequently, there is no inherent bias at any site being simulated, unless a strong autocorrelation is used to “fix” the shape of the surface.

4.0 SENSITIVITY TESTING

In addition to producing a software tool for use in the DQO development, several scenarios were tested to investigate the relative sensitivity to changes in the parameters. These tests consisted of a base case described below and cases that agreed with the base case except for one parameter. Each of these cases is described in this section.

4.1 The Base Case

Similar to the $PM_{2.5}$ DQO, the base case shown in Table 1 was generally developed from what was assumed to be a near worst case estimate for each parameter. These conservative estimates lengthen the gray zone (causing more uncertainty in the estimate) but help ensure that decision makers understand the ramifications for these realistic input parameters on the potential for decision errors. There are some exceptions that are noted. Table 2 provides the original parameter range estimates derived from the preliminary data and how they compare to the base value listed in Table 1. As a note, the daily standard, annual standard and the daily standard percentile used as a base values will be dictated by the setting of the NAAQS . Changing these values does not have an effect on the conclusions drawn from the sensitivity testing.

Table 1. Base case parameter settings in the $PM_{10-2.5}$ DQO model.

Parameter	Base Value	Comments
Type 1 error	0.05	The same as used for $PM_{2.5}$
Type 2 error	0.05	The same as used for $PM_{2.5}$
Daily standard (µg/m^3)	35	These values were assumed in order to make comparisons. The relative effects reported should be representative for a wide range of standards.
Annual Standard (µg/m^3)	20	
Daily standard percentile	0.98	
Sampling frequency	1	Chosen because it was assumed that most sites would be using continuous samplers.
$PM_{2.5}$ completeness	0.75	The same as used for $PM_{2.5}$
PM_{10} completeness	0.75	Assumed for a difference method. This criterion is applied independently of the $PM_{2.5}$ completeness.
$PM_{2.5}$ measurement CV	0.1	The same as used for $PM_{2.5}$
PM_{10} measurement CV	0.1	Assumed for comparison; not based on known achievable results.
Maximum $PM_{2.5}$ bias	0.1	The same as used for $PM_{2.5}$
Maximum PM_{10} bias	0.1	Assumed for comparison; not based on known achievable results.
$PM_{2.5}$ autocorrelation	0	The same as used for $PM_{2.5}$
$PM_{10-2.5}$ autocorrelation	0	See Table 2.
$PM_{2.5}$ to $PM_{10-2.5}$ correlation	0	See Table 2. (Assumed to be a worse case.)
$PM_{2.5}$ population CV	0.8	The same as used for $PM_{2.5}$
$PM_{10-2.5}$ population CV	1	See Table 2.
Mean $PM_{2.5}$/ mean $PM_{10-2.5}$	0.45	See Table 2. (The tool uses the inverse of the usual ratio so that it can be set to 0 for use with direct methods. 1/0.45 = 2.22)
$PM_{2.5}$ seasonal ratio	5.3	The same as used for $PM_{2.5}$
$PM_{10-2.5}$ seasonal ratio	14	See Table 2.
$PM_{10-2.5}$ phase shift	0	Not extensively investigated.
$PM_{10-2.5}$ periods / year	1	Not extensively investigated.
Spatial sill	0	See the discussion in Section 3.2. A spatial sill of zero causes the $PM_{10-2.5}$ surface to be flat.
Spatial range	NA	

Table 2. $PM_{10-2.5}$ population parameter estimates.

Quantile	2.5	10	20	30	40	50	60	70	80	90	97.5
$PM_{2.5}$ seasonal ratio	1.46	1.63	1.77	1.88	2.02	2.14	2.28	2.58	3.03	4.01	5.72
$PM_{10-2.5}$ seasonal ratio	1.68	2.05	2.32	2.73	3.24	3.82	4.42	5.54	8.01	14.34	52.52
$PM_{2.5}$ population CV	0.35	0.41	0.45	0.48	0.51	0.53	0.56	0.6	0.64	0.69	0.8
$PM_{10-2.5}$ population CV	0.4	0.49	0.56	0.61	0.66	0.71	0.76	0.84	0.93	1.08	1.39
$PM_{2.5}$ autocorrelation	0	0.06	0.25	0.35	0.42	0.45	0.48	0.51	0.59	0.68	0.94
$PM_{10-2.5}$ autocorrelation	0	0.13	0.19	0.22	0.28	0.31	0.44	0.48	0.51	0.64	0.81
Mean $PM_{2.5}$ /$PM_{10-2.5}$	0.30	0.45	0.62	0.77	0.96	1.14	1.39	1.79	2.17	2.72	3.57
$PM_{10-2.5}$ to $PM_{2.5}$ corr.	-0.23	-0.05	0.06	0.12	0.19	0.25	0.31	0.39	0.46	0.56	0.69

4.2 Sensitivity Testing Results

The sensitivity testing showed that the length of the gray zone was most sensitive to the sampling frequency, followed by the completeness, the population CV, and the ratio of the mean concentrations of the two fractions. The results are shown in Table 3 and Figure 2. The two items of interest, the effects of changing the number of modes and the spatial variability (highlighted), have much smaller effects. See the next section for discussions.

Table 3. Sensitivity test results.

Case	Parameter settings		Gray Zone		Test Gray Zone Length / Base Case Length
	Base Case	Test	Lower bound	Upper bound	
Base Case	See Table 1	NA	27.4	48.0	1.00
Sampling Frequency	1	1-in-3	22.2	56.2	1.66
Overall completeness	PM2.5 = 75% PM10 = 75%	PM2.5 =100% PM 10 = 75%	27.5	44.7	0.84
$PM_{2.5}$ autocorrelation	0	0.2	27.7	47.8	0.98
$PM_{10-2.5}$ autocorrelation	0	0.2	27.4	47.7	0.99
$PM_{10-2.5}$ to $PM_{2.5}$ correlation	0	0.25	27.2	48.4	1.03
$PM_{10-2.5}$ pop. CV	100%	60%	28.0	45.7	0.86
Mean $PM_{2.5}$ /Mean $PM_{10-2.5}$	0.45	1	26.4	49.4	1.12
phase shift	0	3 mo.	27.6	47.7	0.98
phase shift		6 mo.	28.0	47.6	0.95
phase shift		9 mo.	27.8	47.6	0.97
periods/year	1	2	27.5	47.5	0.97
Spatial variability	Sill = 0, range = NA	Sill = 1, range = 20 km	26.9	48.5	1.05

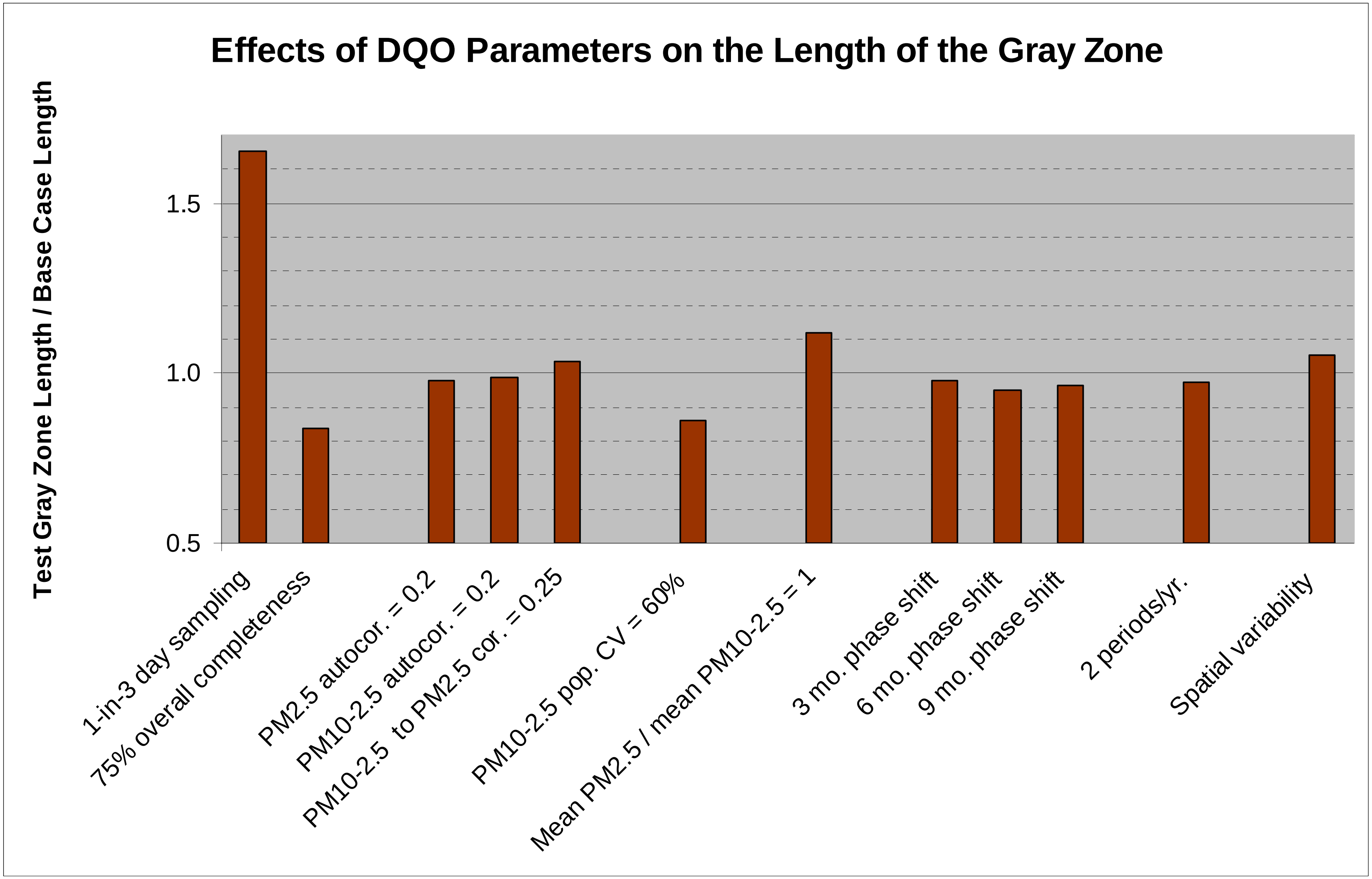

Figure 2. **Effects of the DQO parameters on the length of the gray zone.**

4.3 Sensitivity Testing Results Discussion

This section is subdivided into tests of similar parameters as they are grouped in Figure 2. No testing was done on how the changes in the parameters interact with each other.

4.3.1 Sampling Parameters

The two sampling parameters tested were (1) the sampling frequency that was changed to every third day sampling from the base case of daily sampling and (2) the $PM_{2.5}$ completeness. These two parameters have the largest impacts on the overall length of the gray zone of all of the parameters tested. These both argue for implementing continuous methods, even if the primary DQOs are based on difference methods with less than daily sampling. More detailed analysis is needed to make sure that different measurement biases and precision do not offset the gains from daily sampling. This is further explored in the report regarding the equivalency requirements.

The completeness requirements are modeled independently in the current model. In the test case, the $PM_{2.5}$ completeness was set to 100 percent, so that the overall completeness would be 75 percent each quarter.

4.3.2 Correlation Parameters

The two autocorrelation parameters are essentially measures of how quickly the concentrations change in time. Increasing these parameters aids the ability to make correct decisions, in part because the missing data are not allowed to be as different as the values that do get measured.

Increasing the $PM_{2.5}$ to $PM_{10-2.5}$ correlation has the opposite effect, and the default value should be increased (from zero) for the final DQOs. The overall effect is small, but zero is outside the range found in the historical data. The parameter needs to be revisited when the network has been established. Values greater than 0.5 would seem to indicate common sources of $PM_{10-2.5}$ and $PM_{2.5}$; this is not expected to be the case in regions with high values of $PM_{10-2.5}$.

4.3.3 $PM_{10-2.5}$ Population CV

This parameter was expected to be a key driver to the DQO selection. The base case is approximately equal to the 90th percentile of the estimates from the historical data. The test case is approximately the 30th percentile.

4.3.4 Ratio of the Mean $PM_{2.5}$ Concentration to the Mean $PM_{10-2.5}$ Concentration

The test results show that, for difference methods, the $PM_{2.5}$ is a significant interferant. This should be expected since the $PM_{2.5}$ bias is allowed to be significant. However, in regions with high $PM_{10-2.5}$ concentrations, this ratio will be in a range where it is not a sensitive parameter. This parameter will not be set equal to a near worst case for the nation but within a conservative range where we expect $PM_{10-2.5}$ issues.

4.3.5 Phase Shifts

Phase shifting the high periods for the $PM_{10-2.5}$ and $PM_{2.5}$ decreases the interference that the $PM_{2.5}$ can cause during the time when the $PM_{10-2.5}$ is high.

4.3.6 Multiple High Periods per Year

Setting this parameter equal to 2 decreases the overall length of the gray zone. The results are likely due to the fact that for half of the time, when the $PM_{10-2.5}$ is high, the $PM_{2.5}$ interference is low.

4.3.7 Spatial Variability

The effect of the spatial variability is generally small in the range that is felt reasonable for $PM_{10-2.5}$, (an exponential sill between 60 and 100 percent of the temporal variability and a range parameter equal between 20 to 125 km). The test case shows the settings within this range that correspond to the most spatial variability with the sill equal to 100 percent of the temporal variability and a range equal to 20 km. The effect might be a smaller response than expected due to the way that "truth" was defined for the simulation grid; namely that the 98th percentile design value was calculated for each grid point and then these were averaged over the grid. If the truth were defined to be the maximum rather than the average (or, even more extreme, finding the 98th percentile of the daily maximums across the grid), then there would be a bias associated with the measurements for the center. Since biases are directly reflected in the performance curve, these alternate definitions would both shift and enlarge the gray zones.

4.4 Conclusions

The assessment found that for the daily standard, the performance curves were most sensitive to sampling frequency, followed by the completeness, the population CV of the coarse fraction of the particulate matter, and the ratio of the mean concentrations between the coarse and fine fractions of the particulate matter. The effect of multi-modal distributions was very small. The effect of the spatial variability is small compared to the parameters mentioned above, but we suggest including this parameter in the DQO evaluation tool. Similar to our work with $PM_{2.5}$, as the $PM_{10-2.5}$ network is established and data are collected, EPA will review all the input parameters used to establish the base case and determine from the actual network whether refinements are needed in the DQO model.

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., and Holloman, C.H. (2003). "Estimating Parameters for the PMcoarse DQO Tool." Technical Report to the U.S. EPA, Office of Air Quality Planning and Standards, under EPA Contract No. 68-D-02-061, May.