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

Page

1.0 INTRODUCTION 1
1.1 Data Quality Objectives 1
1.2 Data Uncertainty 2

2.0 THE $PM_{2.5}$ DQOS 5
2.1 The DQO is Based on the Annual Arithmetic Mean NAAQS 5
2.2 The Distribution of the Measurement Error 6
2.3 Errors Can Occur When the Estimated 3-Year Average Differs from the Actual, or True, 3-Year Average 7
2.4 The Limits on Precision and Bias Are Based on the Smallest Number of Sample Values in a 3-Year Period 7
2.5 The Error Limits Were Set at 5 Percent 7
2.6 Measurement Imprecision Was Established at 10 Percent Coefficient of Variation (CV) 8
2.7 Assumptions About the Underlying Variability of the Day-to-Day PM Concentrations 8

3.0 THE MODELING PROCESS 13
3.1 Gray Zone Boundary Cases 13
3.2 Modeling 14

REFERENCES 17

APPENDIX A: ASSUMPTIONS AND INPUT CRITERIA FOR THE DQOS FOR AMBIENT AIR MONITORING OF PM2.5 . . . . A-1

List of Tables

Table 1: Distribution of ratios of highest to lowest monthly or bimonthly mean at a site 9
Table 2: Distribution of CVs about monthly and bimonthly means 9
Table 3. Summary of Case 1 and 2 parameters 14
Table 4. Maximum decision error rates 16

List of Figures

Figure 1. Effect of positive bias on the annual average estimate, resulting in a false positive error 2
Figure 2. Effect of negative bias on the annual average estimate, resulting in a false negative error 2
Figure 3. Annual arithmetic mean and 24-hour 98th percentiles from AIRS data (extracted on April 4, 2001) 5
Figure 4. Comparison of normal and lognormal density functions at low measurement error (10 percent CV) 7
Figure 5. Comparison of normal and lognormal density functions at higher measurement errors (50 percent CV) 7
Figure 6. $PM_{2.5}$ concentrations showing a sinusoidal seasonal pattern along with the DQO sine curve that is associated with these data 10
Figure 7. The data example and curve plotted against the number of days into the year 11
Figure 8. Simulated $PM_{2.5}$ data for every third day with a long-term mean of 16.3 :g/m^3 and a population CV of 80 percent 12
Figure 9. Power curves for 75 percent complete 1-in-6-day sampling with 10 percent measurement CV and +/- 10 percent bias 15

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

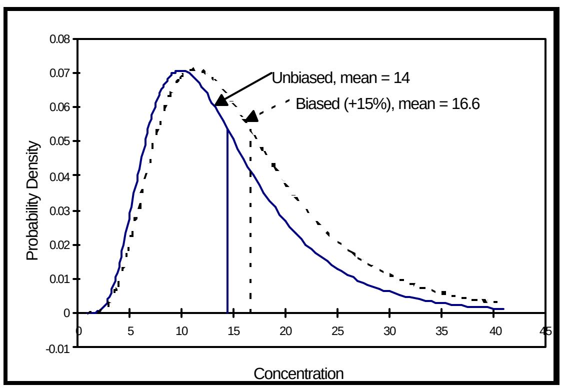

Figure 1. Effect of positive bias on the annual average estimate.

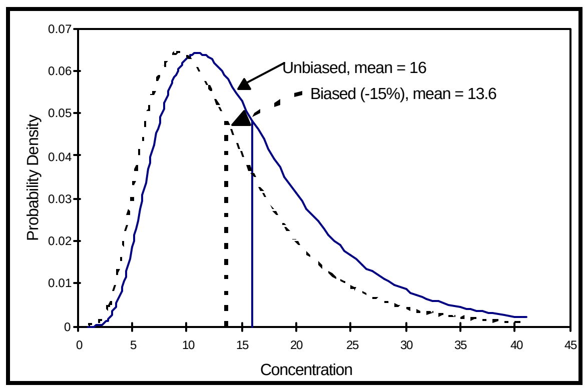

Figure 2. Effect of negative bias on the annual average estimate

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**.

<u>Total Population Uncertainty</u> 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.

<u>Total Measurement Uncertainty</u> 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.

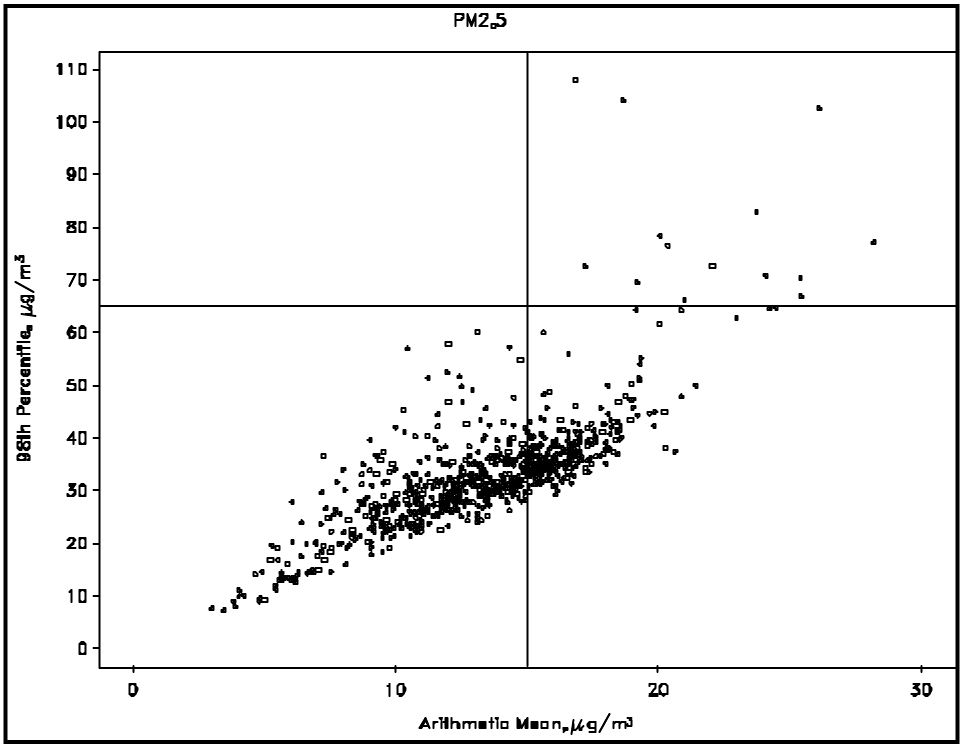

Figure 3. Annual arithmetic mean and 24-hour 98th percentiles from AIRS data (extracted on April 4, 2001)

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.

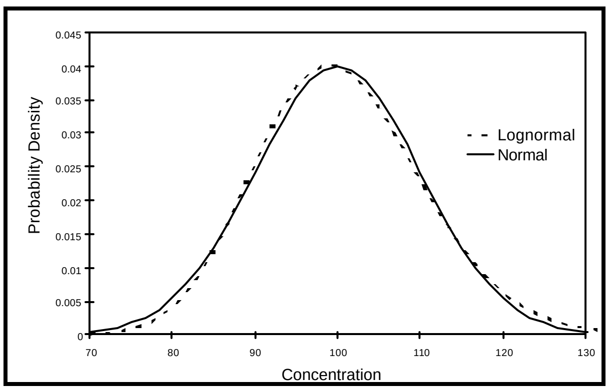

Figure 4. Comparison of normal and lognormal density functions at low measurement error (10 percent CV)

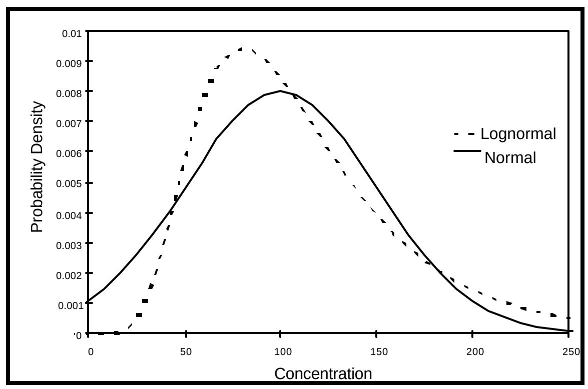

Figure 5. Comparison of normal and lognormal density functions at higher measurement errors (50 percent CV)

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.

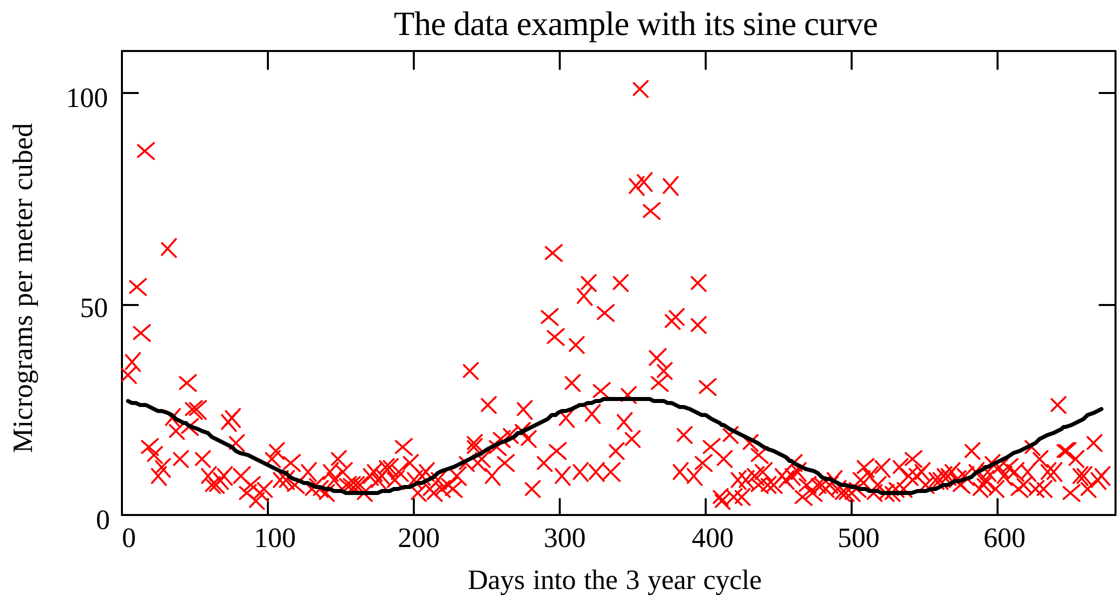

Figure 6. $PM_{2.5}$ concentrations showing a sinusoidal seasonal pattern along with the DQO sine curve that is associated with these data

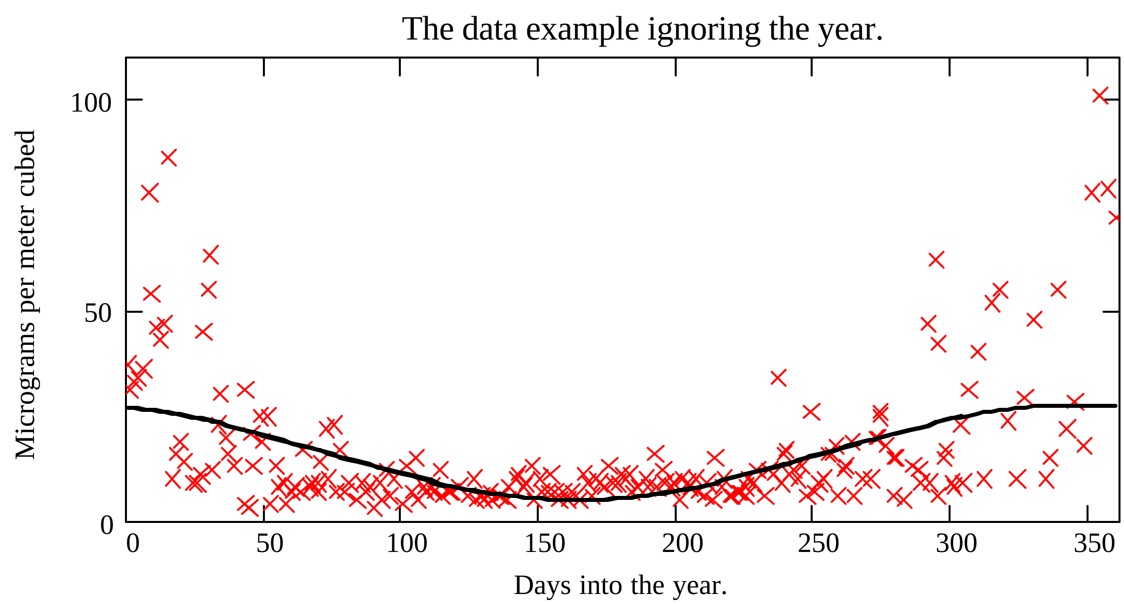

Figure 7. The data example and curve plotted against the number of days into the year

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 = $[16.3 + 11.125 \sin(2 \pi D / 365)] * \delta_D$,

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

where δ_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 $\mu g/m^3$. 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.

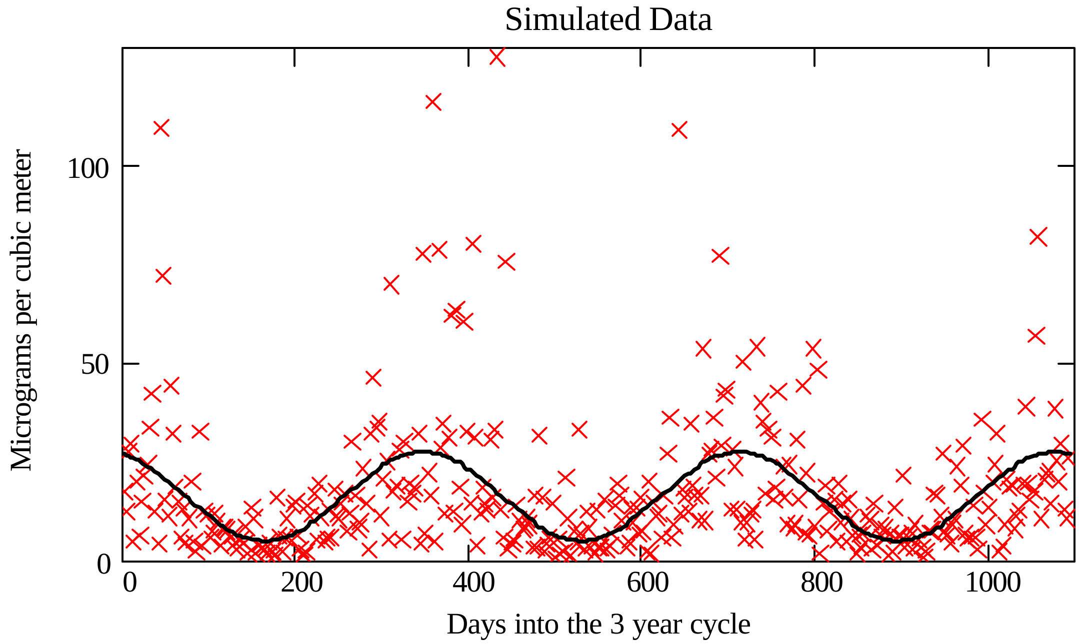

Figure 8. Simulated $PM_{2.5}$ data for every third day with a long-term mean of 16.3 :g/m³ and a population CV of 80 percent

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.

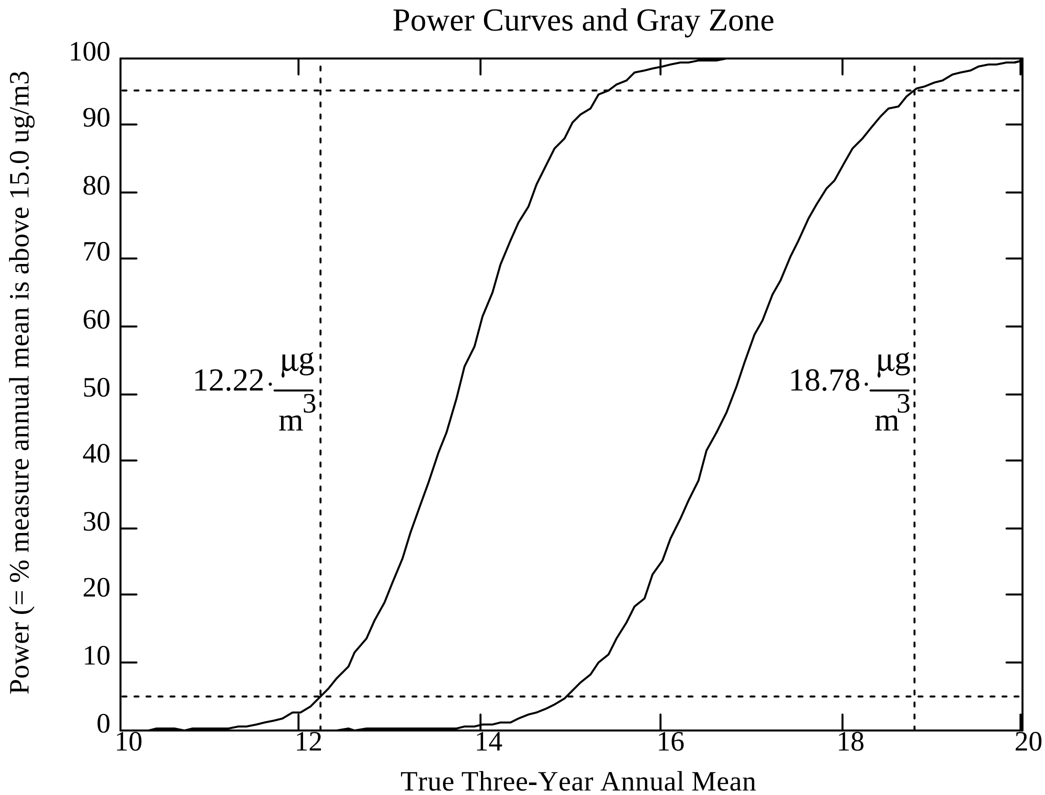

Figure 9. **Power curves for 75 percent complete 1-in-6-day sampling with 10 percent measurment CV and ± 10 percent bias.**

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 $\mu g/m^3$ 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.