

## **Data Quality Objectives for the Trends Component of the PM<sub>2.5</sub> Speciation Network**

The Data Quality Objectives (DQO) process is a strategic planning approach used to prepare for a data collection activity, as described in the document *EPA QA/G-4: Guidance for the Data Quality Objectives Process* (U.S. EPA, 1994). Its purpose is to ensure that the type, quantity, and quality of environmental monitoring data will be sufficient for the data's intended use. The process simultaneously ensures that resources are not wasted collecting unnecessary, duplicative, or overly precise data. The process provides a systematic approach for defining the criteria that a data collection design should satisfy, including when, where, and how many samples to collect and how precise these samples need to be.

The DQO process consists of seven steps. These are: (1) state the problem, (2) identify the decision, (3) identify the inputs to the decision, (4) define the study boundaries, (5) develop a decision rule, (6) specify tolerable limits on decision errors, and (7) optimize the design. This document presents the results of applying the DQO process to the collection of the PM<sub>2.5</sub> speciation data to be used for measuring national trends.

### **1.0 The Problem**

Revised particulate matter National Ambient Air Quality Standards were promulgated July 18, 1997 (U.S. EPA, 1997a and 1997b). These new standards include requirements for monitoring of the chemical species composing particulate matter with aerodynamic diameters less than 2.5 microns (PM<sub>2.5</sub>). The regulations require that approximately 50 sites be established to provide nationally consistent data for the assessment of trends in the chemical constituents of PM<sub>2.5</sub> and that these 50 sites be part of the National Air Monitoring Stations (NAMS) network.

Generally, the DQO process would be used to determine combinations of the sampling frequency, the location of the samples, and the tolerable measurement errors needed to achieve desired levels of errors associated with decisions that will be based on data collected by the PM<sub>2.5</sub> speciation trends sites. However, most of the monitoring characteristics have already been established for the trends network as the result of regulations or recommendations from the PM<sub>2.5</sub> Speciation Expert Panel (Koutrakis, 1998) and PM<sub>2.5</sub> Speciation Workgroup. Thus, the issues to be addressed with this DQO process included (1) estimating the decision errors resulting from the characteristics of the network, (2) recommending changes to the sampling plan if the resultant decision errors were unacceptably large, and (3) proscribing required measurement precision.

The following items summarize the monitoring characteristics that had been

established for the trends network prior to the beginning of the DQO process.

**Number of sites:** The PM<sub>2.5</sub> speciation trends network is to consist of approximately 50 sites, as specified in 40 CFR Part 58, Appendix D. EPA has proposed 53 sites which are documented in *Particulate Matter (PM<sub>2.5</sub>) Speciation Guidance Document* (U.S. EPA, 1998).

**Location of sites:** According to 40 CFR Part 58, Appendix D, approximately 25 of the sites are to be located in Photochemical Air Monitoring Stations (PAMS) areas. The remaining sites are to be selected in coordination among the EPA, the Regional Offices, and the States and locals. Twenty-four of the 53 proposed sites are in PAMS areas. The rationale for the selection and the resulting locations of all the sites are documented in *Particulate Matter (PM<sub>2.5</sub>) Speciation Guidance Document* (U.S. EPA, 1998).

**Sampling frequency:** The PM<sub>2.5</sub> Speciation Expert Panel and EPA have determined that the sampling frequency for the trends sites should be once every 3 days which is documented in *Summary of the Recommendations of the Expert Panel on the EPA PM<sub>2.5</sub> Speciation Guidance Document* (Koutrakis, 1998).

**Sampler type:** The sampler will be a multiple filter device that collects 24-hour integrated samples.

**Analytes to be measured and the method of measurement:** The species to be measured include:

- elements Na through Pb using x-ray fluorescence spectroscopy (XRF),
- major ions (sulfate, nitrate, chloride, ammonium, sodium) using ion chromatography (IC), and
- total, elemental, and organic carbon using thermal optical analysis (TOA).

## 2.0 The Decision

The DQO process incorporates input from a planning team consisting of program staff, technical experts, managers, a quality assurance/quality control advisor, and a statistician. This enables data users and relevant technical experts to specify their particular needs prior to data collection. The members of the PM<sub>2.5</sub> Speciation DQO planning team, referred to as the decision makers, are listed in Table 1.

These decision makers decided that the primary objective of the trends component of the PM<sub>2.5</sub> speciation network is to detect trends in individual component species on a site-by-site basis. Specifically, the decision makers wanted to be able to detect a 3-5% annual trend (increasing or decreasing) with 3-5 years of data.

**Table 1. PM<sub>2.5</sub> Speciation DQO Planning Team Members**

<b>PM<sub>2.5</sub> Speciation DQO Decision Makers</b>			
<b>Name</b>	<b>Address</b>	<b>Phone Number</b>	<b>Electronic Mail</b>
Mike Poore Alice Westerinen	Air Resources Board P.O. Box 2815 Sacramento, CA 95812	(916) 323-4498 (916) 324-6191	mpoore@arb.ca.gov awesteri@arb.ca.gov
Peter Scheff	Environmental and Occupational Health Sciences University of Illinois School of Public Health 2121 W. Taylor Street, Rm. 422 Chicago, IL 60612-7260	(312) 996-0800	pscheff@uic.edu
Alan Leston	DEP Planning & Stds Div 5 <sup>th</sup> Floor 79 Elm Street Hartford, CT 06106	(860) 424-3513	alan.leston@po.state.ct.us
Phil Galvin	NY State DEC 50 Wolf Road Albany, NY 12233-3256	(518) 457-7794	pjgalvin@gw.dec.state.ny.us
Terence Fitz-Simons	USEPA/OAQPS AQTAG (MD-14) Research Triangle Park, NC 27711	(919) 541-0889	fitzsimons.terence@epa.gov
Jim Homolya Joann Rice	USEPA/OAQPS MQAG (MD-14) Research Triangle Park, NC 27711	(919) 541-4039 (919) 541-3372	homolya.james@epa.gov rice.joann@epa.gov
<b>Contractor for Development of PM<sub>2.5</sub> Speciation DQOs</b>			
<b>Name</b>	<b>Address</b>	<b>Phone Number</b>	<b>Electronic Mail</b>
Nancy McMillan	Battelle 505 King Avenue Columbus, OH 43201-2693	(614) 424-5688	mcmillan@battelle.org
<b>Primary EPA Contact for Development of PM<sub>2.5</sub> Speciation DQOs</b>			
<b>Name</b>	<b>Address</b>	<b>Phone Number</b>	<b>Electronic Mail</b>
Shelly Eberly	USEPA/OAQPS MQAG (MD-14) Research Triangle Park, NC 27711	(919) 541-4128	eberly.shelly@epa.gov

Although the data collected by the PM<sub>2.5</sub> speciation network will be invaluable for a multitude of data analyses, the detection of trends is the primary objective of the NAMS portion of the PM<sub>2.5</sub> speciation network, as stated in 40 CFR Part 58 Appendix D. The decision makers and the PM<sub>2.5</sub> Speciation Expert Panel (Koutrakis, 1998) concurred about this being the primary objective and therefore the one on which the DQOs should be based. This means that the tolerable decision errors will be based exclusively on trends analyses, even though other data uses might have larger resultant decision errors. The need for accurate trends at the site level is due to the manner in which the trends will be used. The decision makers decided that trends are needed to evaluate the long-term effectiveness of control strategies. Incorrect estimation of trends may lead to incorrect decisions about the effectiveness of implemented control strategies. Since control strategies likely will be developed, applied, and/or evaluated at the Metropolitan Statistical Area (MSA) level and given that at most one trend site will be within an MSA, the trends need to be accurate on a site by site basis. Additionally, the decision makers thought that regional or national trends would be difficult, if not impossible, to interpret because of the geographical variability in meteorology, species composition, and control strategies.

Variation in meteorology can mask or attenuate trends that are due to changes in emissions. Given the intended use of the trends data, the decision makers decided that meteorological variation needs to be removed before the trend analysis is performed. That is, the trends in which the decision makers wanted to have the specified decision errors are ones for which the impact due to variation in meteorology has been removed. The details for how this adjustment was accomplished are included in the appendix. Basically, a seasonal component based on the number of days into a year was added to the statistical model of the data.

Lastly, the development of the DQOs was done for four analytes, those being sulfate, nitrate, total carbon, and calcium. The target analytes of interest for the speciation trends sites were selected to include those which have been historically measured within the Interagency Monitoring of Protected Visual Environments (IMPROVE) network. To ensure that data from the speciation trends sites could be compared with IMPROVE data sets, the trends DQO development considered an analysis of the ability to sample and measure selected analytes which are thought to be major components of aerosols collected in both networks (sulfate, nitrate, and total carbon) and whose concentrations could be expected to vary with the implementation and effectiveness of emissions controls. Sulfate is a direct indicator of anthropogenic emissions, primarily from fossil-fuel fired combustion sources and can be effectively measured by most fine particulate sampling systems. Sulfate levels are usually the highest in the eastern US. By contrast, nitrate is an indicator of secondary atmospheric aerosol formation resulting from nitrogen oxides emissions and is somewhat difficult to quantitatively sample because of volatilization artifacts which can occur in many sampling systems. Nitrate levels are usually the highest in the western US. Total carbon in fine aerosol particles is associated with wood combustion and mobile source emissions and also represents an analyte which has the potential for either positive or negative sampling artifacts. Calcium was included since it is an element which is

generally associated with nonanthropogenic emissions such as windblown soils, mineral materials, and dusts. Calcium is usually assumed to occur in particles predominantly greater than 2.5 microns. Therefore, fine particulate calcium should be present at low background levels and represents an aerosol constituent which is not expected to vary significantly with source emissions controls implementation.

### 3.0 Inputs to the Decision

Data from the IMPROVE program was used for estimating the variability likely to be observed in national PM<sub>2.5</sub> speciation measurements. This is because each of the analytes to be monitored at the 50 NAMS sites is currently being monitored in the IMPROVE program and the analytical methods used in the IMPROVE program are similar to those to be used in the national program. Table 2 provides a matrix of filter types, target species, and analytical methods to be used in the national program.

**Table 2. PM<sub>2.5</sub> Chemical Speciation Filter Medium, Target Species and Methods**

Filter Medium	Target Species	Analytical Technique
PTFE (Teflon®) filter	Elements: Al through Pb; and mass	EDXRF (IO-3.3) and Gravimetry
Nylon filter with nitric acid denuder	Anions: nitrate and sulfate	IC (IMPROVE Method)
	Cations: ammonium, sodium, and potassium	IC (IMPROVE Method)
Pre-fired quartz fiber filter with gaseous organic denuder	Total carbon (including organic, elemental, carbonate carbon)	TOA (NIOSH 5040)
EDXRF - Energy Dispersive X-ray Fluorescence		
IC - Ion Chromatography		
TOA - Thermal Optical Analysis		

The national chemical analysis methods for calcium and total carbon differ from those used in the IMPROVE network. The IMPROVE network uses proton induced x-ray emission (PIXE) to analyze the PTFE filter for calcium and thermal optical reflectance (TOR) to analyze the quartz filter for total carbon. Due to the lack of long-term data collected using EDXRF and TOA, it is an assumption of this DQO process that EDXRF and PIXE have similar percentages of non-detects and levels of precision and similarly that TOR and TOA have similar percentages of non-detects and levels of precision. This assumption is questionable for calcium based on a recent article that indicates that the detection limit for calcium using XRF may be 5 times that for PIXE, 2.4 ng/m<sup>3</sup> for XRF

versus 0.5 ng/m<sup>3</sup> for PIXE (Nejedly, 1998). The recent literature supports the assumption regarding the comparability of TOR and TOA (Birch, 1998).

An additional difficulty in using the IMPROVE data for the national trend network planning is that the IMPROVE sites, by design, are predominantly located in rural areas. This will not and should not be true of the NAMS sites. Anticipated differences in variability of speciated PM<sub>2.5</sub> data between rural and urban sites was factored into estimates obtained based on the IMPROVE data. This was accomplished by analyzing the data from the urban IMPROVE site located in Washington DC (WASH1), the only long-term urban IMPROVE site.

### 3.1 Summary of Model for All IMPROVE Sites

The model used to describe the IMPROVE data is documented in the appendix. Basically, the model is for log-transformed concentrations and includes a seasonal component, a linear trend component to indicate the increase or decrease in concentrations from one year to the next, and a term for auto-correlation to reflect that data collected close in time are more similar than data collected farther apart in time. The estimates for the parameters of this model are shown in Table 3 for sulfate, nitrate, calcium, and total carbon measurements for the IMPROVE data *on a site by site basis*. Average parameters across all sites are reported as well as the range of values observed.

**Table 3. IMPROVE Data Summary - All Sites**

Parameter	Seasonality (Ratio of Summer Peak Concentration to Winter Trough Concentration)		Time Trend (Percent Reduction in Concentration from One Year to the Next)		Auto-correlation (Estimated Correlation Coefficient Between 3 day Apart Log-transformed Concentrations)		Baseline Concentration (Estimated Concentration During Winter Trough of 1988, µg/m <sup>3</sup> )		Error (Coefficient of Variation)	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
Sulfate	3.485	0.640-12.716	2.90%	-8.0%-15.2%	0.224	0.014-0.404	0.995	0.072-3.982	0.719	0.482-1.529
Nitrate	1.979	0.061-18.370	4.30%	-10.8%-28.7%	0.237	0.079-0.457	0.755	0.024-6.752	1.064	0.502-2.373
Calcium	5.368	0.654-22.528	1.90%	-23.1%-21.6%	0.328	0.037-0.681	0.013	0.002-0.049	0.936	0.558-1.557
Total Carbon	1.934	0.280-6.328	3.70%	-5.1%-18.3%	0.287	0.013-0.540	1.605	0.200-11.940	0.589	0.375-0.915

**Table 3 (continued). IMPROVE Data Summary - All Sites**

Parameter	Concentration ( $\mu\text{g}/\text{m}^3$ )		Measurement Error (Average coefficient of variation)		Percentage Non-Detects (%)	
	Mean	Range	Mean	Range	Mean	Range
Sulfate	1.357	0.246-4.932	0.055	0.036-0.124	0.293	0.000-3.343
Nitrate	0.261	0.029-1.321	0.133	0.046-0.502	4.118	0.000-37.021
Calcium	0.026	0.008-0.083	0.089	0.062-0.129	2.078	0.000-13.566
Total Carbon	1.295	0.342-4.119	0.176	0.064-0.421	0.000	0.000-0.000

The first segment of Table 3 presents results from fitting linear models to log-transformed sulfate, nitrate, calcium, and total carbon concentrations at each of the approximately 60 IMPROVE sites that have recorded multiple years of data. Based on examination of the raw data, seasonal trends were fitted with  $\frac{1}{2}$  a sine wave with a peak around June 30. Time trends were assumed to be linear in the log-transformed concentrations. The correlation between consecutive samples collected at the same site was assumed to be  $\rho$ . IMPROVE data is primarily collected on Wednesday and Saturday, and, thus, this parameter represents the correlation between concentration measurements either 3 or 4 days apart. These correlations were adjusted to be appropriate for one in three day sampling and are presented in Table 3. Baseline concentrations are presented next and can be used in conjunction with the seasonal and time trend parameters to estimate the concentration for an individual day. The final columns of the first segment of Table 3 present the unexplained variation in sulfate, nitrate, calcium, and total carbon after fitting the model that includes the seasonal, time trend, and auto-correlation adjustment. Variation is expressed as a coefficient of variation for the untransformed concentrations.

The second segment of Table 3 presents descriptive statistics on the geometric mean concentrations, measurement error rates, and non-detect percentages. The measurement error rates and percentage of non-detects are quantities provided on the IMPROVE database. Generally, small non-detect percentages were observed for each of the species across the network. This suggests that the ability of laboratories to detect concentrations of each of the species at levels actually present in rural areas is good. Since urban concentrations are anticipated to be higher than rural concentrations, inability to detect species should not be an issue. Note that the measurement error is small compared to the variability remaining after fitting the model. This can be seen by comparing the coefficient of variation in the second part of the table (measurement error) with the coefficient of variation in the first part of the table (variability remaining after fitting the model).

### 3.2 Summary of Model for Washington DC IMPROVE Site

The model used to describe the Washington DC IMPROVE data is identical to that used for the other IMPROVE sites. The estimates for the parameters are shown in Table 4 for sulfate, nitrate, calcium, and total carbon measurements.

**Table 4. IMPROVE Data Summary — Urban Washington Site**

Parameter	Seasonality (Ratio of Summer Peak Concentration to Winter Trough Concentration)	Time Trend (Percent Reduction in Concentration from One Year to the Next)	Auto-correlation (Estimated Correlation Coefficient Between 3-day Apart Log-transformed Concentrations)	Baseline Concentration (Estimated Concentration During Winter Trough of 1988, $\mu\text{g}/\text{m}^3$ )	Error (Coefficient of variation)
Sulfate	2.250	2.90%	0.110	3.405	0.574
Nitrate	0.289	3.80%	0.228	3.389	0.850
Calcium	1.458	1.90%	0.108	0.034	0.572
Total Carbon	0.866	3.30%	0.019	5.387	0.524

**Table 4 (continued). IMPROVE Data Summary — Urban Washington Site**

Parameter	Concentration ( $\mu\text{g}/\text{m}^3$ )	Measurement Error (Average Coefficient of Variation)	Percentage Non-detects (%)
Sulfate	4.932	0.036	0.000
Nitrate	1.240	0.046	0.000
Calcium	0.039	0.083	0.125
Total Carbon	4.119	0.064	0.000

Baseline concentrations are higher for the Washington DC site than for an average IMPROVE site (Table 3) as seen by comparing either the baseline concentrations for the winter trough of 1988 or by comparing the geometric mean concentrations, while the seasonal effects, auto-correlation, and coefficients of variation seem to be lower. Time trends at the Washington site are similar in direction and magnitude to average IMPROVE sites, and the time trends are statistically significant. Due to the higher concentrations observed at the Washington site than at a typical IMPROVE site, measurement error and percentages of non-detects are lower. The reduced auto-correlation and error observed at the Washington site will make it easier to detect time trends in sulfate, nitrate, calcium, and total carbon than at a typical IMPROVE site. However, the lower auto-correlation estimate is of some concern. Recall that the sampling schedule employed at IMPROVE sites is Wednesday and Saturday. Auto-correlation between consecutive samples could be reduced due to a day of the week effect. For instance, decreased vehicular traffic on

Saturdays as compared to a weekday may be the reason for the reduced auto-correlation. A model that incorporates an effect for the day of the week might be more appropriate but was not considered for this DQO process.

### **3.2 Incorporation of Meteorological Variability into Model**

The effect of meteorological information (today and yesterday's temperature and relative humidity) on PM<sub>2.5</sub> species concentrations was considered in separate analyses. The appendix contains details about this analysis. Based on the described analysis, it can be concluded that simple meteorological models only slightly reduce variability and auto-correlation. Seasonal effects are significantly reduced from those indicated based on the non-meteorologically adjusted data, as expected.

### **4.0 Study Boundaries**

The results from this DQO process are to be limited strictly to the NAMS portion of the PM<sub>2.5</sub> speciation network.

### **5.0 Decision Rule**

The decision makers wanted to be able to detect a 3-5% annual trend with 3-5 years of data on a site-by-site basis after adjusting for seasonality. Thus, the parameter of interest is the percent reduction in PM<sub>2.5</sub> sulfate, nitrate, calcium, or total carbon concentration after adjustment for seasonal effects and auto-correlation. This parameter should be estimated by regressing log-transformed sulfate, nitrate, calcium, or total carbon concentration measurement, simultaneously, on a seasonal effect and a linear time trend while accounting for, at a minimum, first-order auto-correlation in the data points. It can be assumed, with a Type I error rate of 0.05, that if the estimated time trend parameter divided by the standard error associated with that estimate is less than -1.645, then a reduction in sulfate, nitrate, calcium, or total carbon concentration has occurred. If the estimated time trend parameter divided by the standard error associated with that estimate is greater than 1.645, then an increase in sulfate, nitrate, calcium, or total carbon concentration has occurred. The technical details are included in the appendix.

### **6.0 Tolerable Limits on Decision Errors**

Two types of decision errors are possible: 1) false positives, claiming that a trend is detected when in fact there is no trend and 2) false negatives, claiming that a trend is not present when in fact there is a trend. By using a statistical hypothesis test, the chances of false positives and false negatives can be estimated in advance. Power curves provide

information about false positives and false negatives. Figure 1 provides an example power curve with interpretation of the power curve annotated on the graph. This particular power curve demonstrates the power to detect a reduction in sulfate based on 1 in three day sampling for five years. Power to detect an *increase* in sulfate is equivalent. (See the appendix for mathematical definitions of increase and reduction in this setting.)

Curves indicating the statistical power with which a trend of a given size can be detected were computed for each of the four species concentration, as described in the appendix. The curves were computed based on the variability and auto-correlation observed when additive seasonal and linear time trend models were fitted to the Washington DC IMPROVE site log-transformed PM<sub>2.5</sub> species concentration data. Figures presented in the appendix present these power curves under a number of assumptions on sampling frequency (daily sampling, 1 in 3 days sampling, and 1 in 6 days sampling), duration of sampling (3 and 5 years), and measurement error (1.0 and 2.0 times IMPROVE's Washington DC site's measurement error rate).

Input from the decision makers indicated that a 3-5% trend per year needs to be detected with 0.8 power within 3-5 years of initiation of sampling. Table 5 summarizes the percent reductions (increases) that can be detected with 0.8 power under a variety of assumptions about sampling frequency and length of time until detection of a trend. Based on this we conclude that with 1 in 3 day sampling for five years, we can detect annual trends that are greater than 5% or less than -5% for sulfate, calcium and total carbon. For nitrate, annual trends must be greater than 6.3% or less than -6.3% in order for us to detect them. Daily sampling provides little improvement in the ability to detect trends.

Analyses documented in the appendix indicate that power is relatively robust to changes in measurement error, up to 2 times IMPROVE's rate. This is because measurement error is small compared to variability left unexplained by the seasonal and time trend component model. Analyses depicting the effect of measurement error if a better data model were developed are presented in the appendix. It is anticipated that measurement error may be more critical for uses of the speciated PM<sub>2.5</sub> data other than trend detection. Therefore, it is advantageous to strive for levels of measurement error comparable to those achieved in the IMPROVE program. Thus, one in three day sampling for five years with measurement error rates similar to IMPROVE's is recommended for trend identification.

**Table 5. Percent Increase or Reduction Providing Power of 0.80 For Two Sampling Periods and Three Sampling Frequencies**

Species	Daily Sampling		One in Three Day Sampling		One in Six Day Sampling	
	Three Years	Five Years	Three Years	Five Years	Three Years	Five Years
Sulfate	7.5	3.6	8.6	4.1	10.9	5.2
Nitrate	12.2	5.9	13.0	6.3	15.3	7.4
Calcium	7.4	3.5	8.5	4.1	10.9	5.2
Total Carbon	5.5	2.6	7.3	3.4	10.0	4.8

## 7.0 Optimized Design

The sampling design to be employed is like that stated in Section 1.0, that is, the sampling frequency will be 1 in 3, the number of sites will be approximately 50 sites with no more than one site per MSA, the sampler will be filter based, the laboratory analyses to be used include TOA, IC, and XRF, and the measurement errors rates and percentage of non-detects should be similar to those seen in the IMPROVE program. The error rates and percentage of non-detects are summarized in Table 4.

## 8.0 Conclusions

Use of the sampling design that includes 53 trends sites, a sampling frequency of 1 in 3, a filter based sampler, laboratory analyses described in Table 2, and measurement error rates and percentage of non-detects like those shown in Table 4 will achieve the decision makers' goal for all species except nitrate. The design will allow for the detection of annual trends greater than 5% (or less than -5%) within 5 years of collection of data, with a power of 0.8. For nitrate, annual trends must exceed 6% (or be less than -6%) to be detected. The decision makers decided that the annual trend required for nitrate was not sufficiently different from their goal to require adjustment to the sampling design.

The decision makers further recommend that the DQOs be reevaluated once data from the trends network becomes available. This is needed due to the assumptions that had to be made in this DQO process because of use of IMPROVE data.

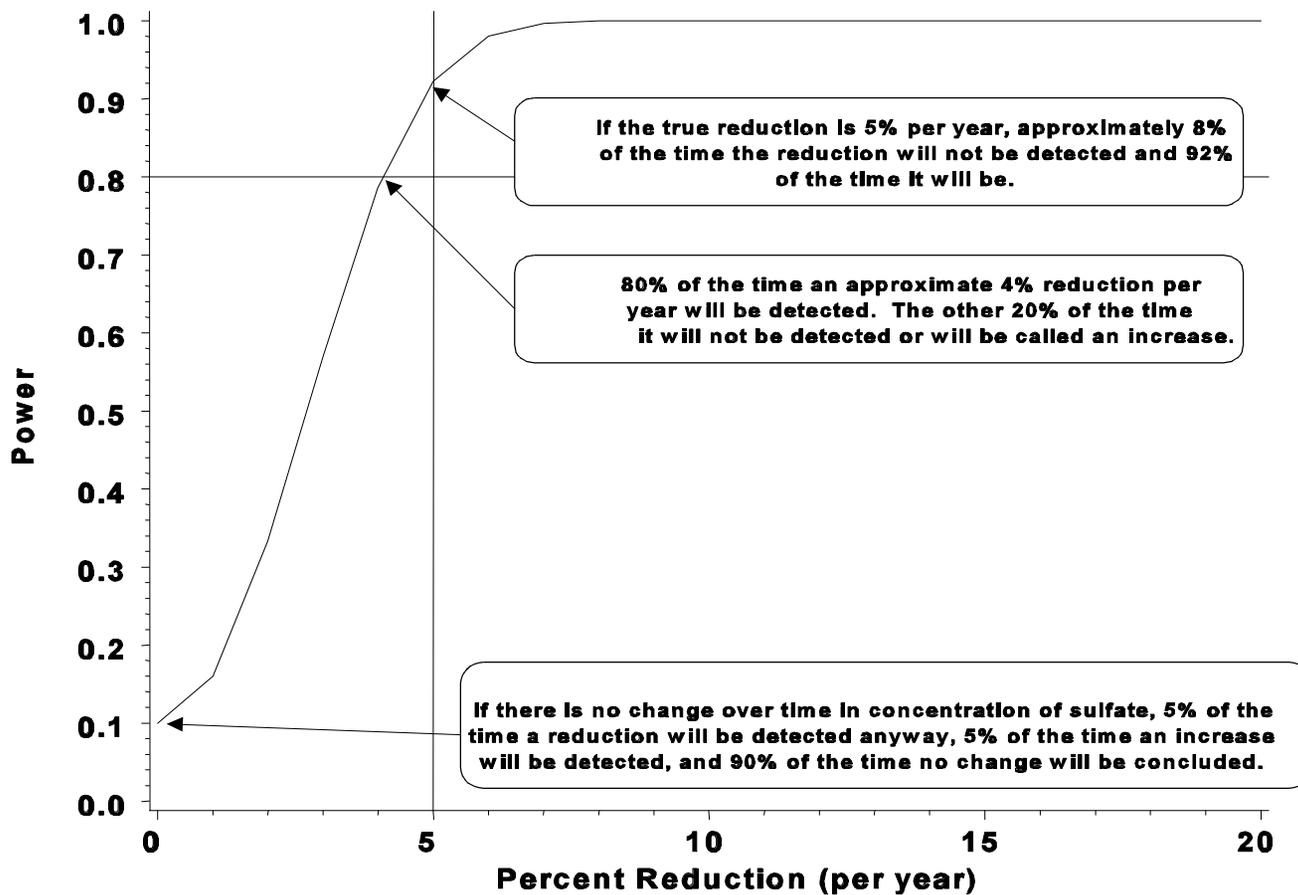


Figure 1. Illustrative Power Curve for 1 in 3 Day Sampling of Sulfate for 5 Years. Interpretation of the Power Curve Annotated

## 9.0 References

- Birch, M.E. (1998). "Analysis of Carbonaceous Aerosols: Interlaboratory Comparison," *The Analyst*, May, Vol 123(851-857)
- Koutrakis, P. (Chair) (1998). "Summary of the Recommendations of the Expert Panel on the EPA PM<sub>2.5</sub> Speciation Guidance Document," May, <http://www.epa.gov/ttn/amtic/files/ambient/pm25/spec/speciate.pdf>
- Nejedly, Z. et. al. (1998). "Inter-Laboratory Comparison of Air Particulate Monitoring Data," *JAWMA*, May, 48:386-397
- U.S. EPA (1994). Guidance for the Data Quality Objectives Process: EPA QA/G-4, Report No. EPA/600/R-96/055, U.S. EPA, Washington, DC
- U.S. EPA (1997a). National Ambient Air Quality Standards for Particulate Matter - Final Rule. 40 CFR Part 50. *Federal Register*, **62**(138):38651-38760. July 18, 1997
- U.S. EPA (1997b). Revised Requirements for Designation of Reference and Equivalent Methods for PM<sub>2.5</sub> and Ambient Air Quality Surveillance for Particulate Matter - Final Rule. 40 CFR Parts 53 and 58. *Federal Register*, **62**(138):38763-38854. July 18, 1997
- U.S. EPA (1998). *Particulate Matter (PM<sub>2.5</sub>) Speciation Guidance (Draft)*, U.S. EPA, Research Triangle Park, NC

## Appendix Technical Details

### 1.0 Notation and Statistical Modeling

Our assumed data model based on preliminary analysis of the IMPROVE data is:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + e_i \quad (\text{A-1})$$

where  $Y_i$  is the  $i^{\text{th}}$  day's log-transformed species concentration measurement,

$$X_{1,i} = \sin\left[\frac{X_{2,i} \bmod 365}{365} \pi\right] \quad (\text{A-2})$$

is the  $\frac{1}{2}$  sine wave seasonal effect,  $X_{2,i}$  is the number of days between January 1, 1988 and the day on which the  $i^{\text{th}}$  observation was collected (the time trend) and  $e_i$  is a first order auto-regressive error term with correlation  $\rho$  and variance  $\sigma^2$ . We assume that this first-order auto-regressive model is appropriate for daily data. This model was fitted separately to data at each IMPROVE site with multiple years of sampling data. A summary of the coefficients estimated is provided in Table 3 of the main text. The results of the model fit to the Washington DC IMPROVE site are provided in Table 4 and Figure A-1. The seasonal effect in the total carbon model is not significant. These plots illustrate that a significant amount of variability in the species concentration data remains unexplained by the model.

Generally, inclusion of a seasonal effect in the species concentration model has been shown by the IMPROVE data to significantly reduce residual variability. This allows more accurate estimation of the other model parameters, specifically the trend parameter. Additional work with meteorological parameters (temperature and relative humidity) at a subset of the IMPROVE sites has shown that use of these parameters, daily and lagged one day, can eliminate the need for a seasonal effect. However, there is only a very slight reduction in residual variation and only very minor changes in auto-correlation when both seasonal and meteorological effects are considered (only simple models were considered). Thus, power to detect a trend, which will be shown subsequently to be a function of how precisely the trend parameter can be estimated, can be increased by considering a model including a seasonal effect. Additional gains from considering meteorology are negligible.

Suppose that 1 in D day sampling is performed rather than daily sampling. This leads to data with the same variance,  $\sigma^2$ , but different correlation between consecutive measurements. In particular the correlation between consecutive measurements given 1 in D day sampling is  $\rho^D$ . This relationship is useful for transforming the correlation coefficient estimated via the

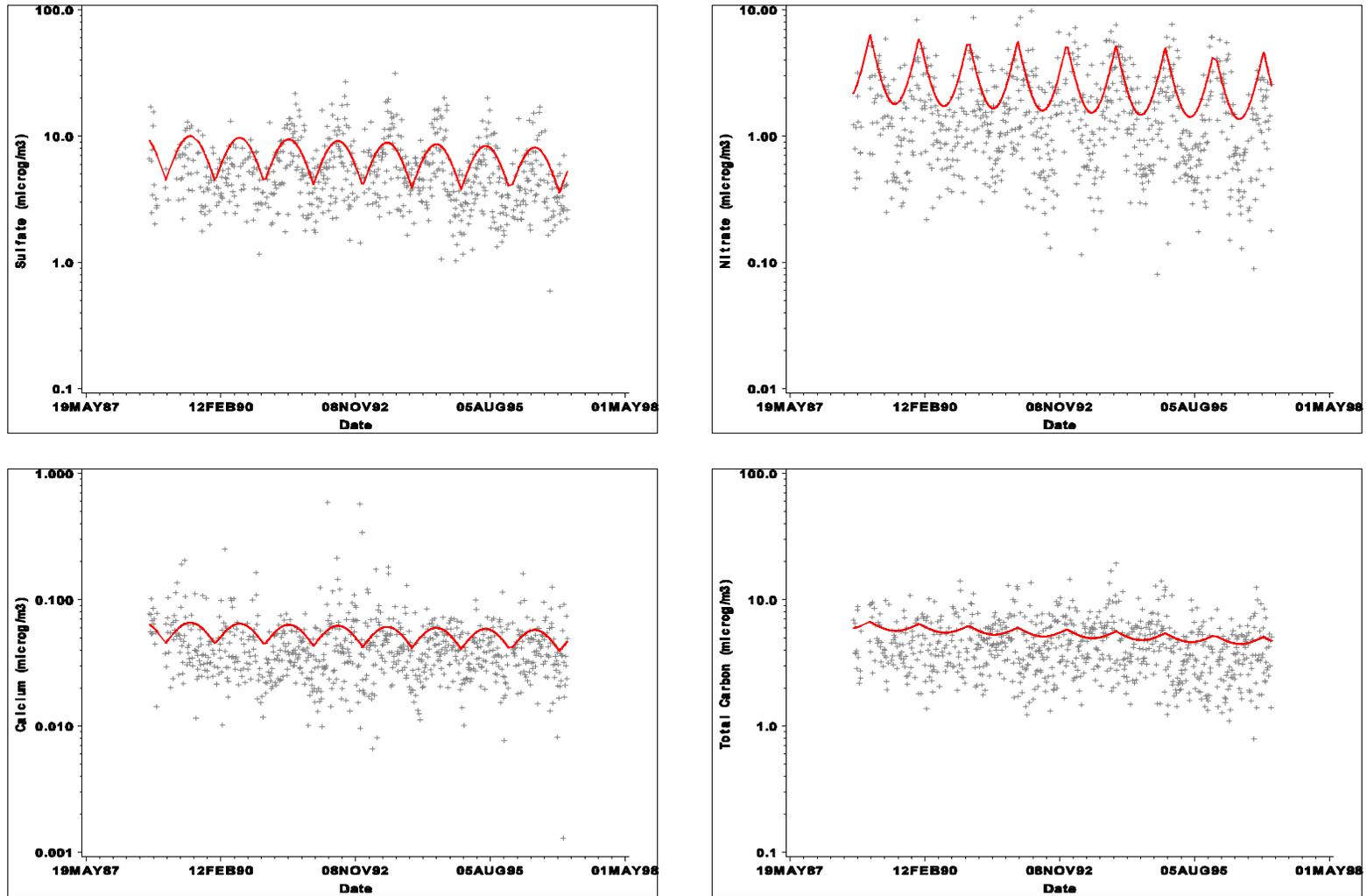


Figure A-1. First Order Auto-regressive Model Fit for Washington IMPROVE Data

IMPROVE data (1 in 3-4 day sampling) to correlation coefficients for alternative sampling frequencies.

The coefficient of variation (square root of variance divided by mean value) of the original concentration measurements based on the above model is a function of  $\sigma^2$ . In particular, This is a

$$CV = \sqrt{\exp(\sigma^2) - 1} \quad (\text{A-3})$$

useful quantity for assessing the residual variability in species concentration data and, in particular, the contribution of measurement error to that variability. Suppose that measurement error,  $ME_{CV}$ , is defined as the sampling and analysis error associated with individual concentration measurements divided by the individual concentration measurements. Based on this definition, the percent of variability due to measurement error, using a variance components approach on the log-transformed concentrations, is

$$\frac{S_{me}^2}{S^2} = \frac{\ln[ME_{cv}^2 + 1]}{\ln[CV^2 + 1]} \quad (\text{A-4})$$

Similarly, the residual variability expected if measurement error is changed by a factor of F, as measured by the coefficient of variation, can be estimated by

$$CV_F = \sqrt{e^{\sigma_E^2 + \ln[(1-F)^2 CV_{ME}^2 + 1]} - 1} \quad (\text{A-5})$$

Using this relationship, measurement errors larger or smaller than IMPROVE's can be considered by transforming  $CV_F$  into  $\sigma_F^2$  using the relationship between these two quantities provided by A-3.

Suppose that 1 in D day sampling has been performed for Y years. The model described can be written multivariately as

$$Y = X\beta + e \quad (\text{A-6})$$

where Y is a vector containing the log-transformed species concentration measurements of length

$$N = \text{Integer}\left[\left(\frac{365}{D}\right) * Y\right], \quad (\text{A-7})$$

X is the (Nx3) design matrix defined by the linear model described,  $\beta$  is a (3x1) vector containing the three parameters described in the linear model, and e is a (Nx1) vector of error terms with covariance structure described by  $\Sigma$  (NxN).  $\Sigma$  is the product of  $\sigma_F^2$  and a standard first-order auto-regressive correlation matrix (NxN) with correlation parameter  $\rho^D$ . This multivariate model

description will be useful later to describe how the monitoring assumptions enter into the power calculations.

## 2.0 Time Trend Power Calculation

Ultimately, to test for the presence of a time trend in the data we are interested in  $\beta_2$ . We interpret  $\beta_2$  by calculating R, the reduction in species concentrations *from one year to the next*, and showing that R is a function of  $\beta_2$  only:

$$R = \frac{\text{last year's mean} - \text{this year's mean}}{\text{last year's mean}} = 1 - \exp(365 * \beta_2) \quad (\text{A-8})$$

For symmetry in our power calculations we also define I, the increase in species concentrations from one year to the next,

$$I = \frac{\text{this year's mean} - \text{last year's mean}}{\text{this year's mean}} = 1 - \exp(-365 * \beta_2) \quad (\text{A-9})$$

We perform our power calculations considering only R, as I is a function of R,

$$I = 1 - \frac{1}{1 - R}. \quad (\text{A-10})$$

Therefore, the null hypothesis that R, the reduction (I, the increase) in species concentration per year, is zero is equivalent to the null hypothesis that  $\beta_2$  is zero. The alternative that R is less than or greater than zero is equivalent to  $\beta_2$  greater than or less than zero. A 0.1 size hypothesis test for the defined null and alternative hypotheses is to reject the null hypothesis if

$$\left| \frac{B_2}{S_b} \right| > 1.645 \quad (\text{A-11})$$

where  $B_2$  is the least squares estimate of  $\beta_2$  and  $s_b$  is the estimated standard error of  $B_2$ . A positive  $B_2$  is equivalent to an increase in concentration over time and a negative  $B_2$  is equivalent to a reduction. This test is based on the assumption that the least squares estimate of  $\beta_2$  is normally distributed with mean equal to  $\beta_2$ . This assumption is appropriate asymptotically, and in this case the sample size (N) is relatively large.

Power is the probability of rejecting the null hypothesis and is a function of the true reduction in species concentration:

$$Power(R) = 1 - P[-1.645 * s_{\beta} < B_2 < 1.645 * s_{\beta} | R = 1 - \exp(365 * \beta_2)] \quad (A-12)$$

Based on this definition and the assumption that  $B_2$  the least squares estimator of  $\beta_2$  is normally distributed with variance estimated by  $s_{\beta}^2$  it can be shown that

$$Power(R) = 1 - \Phi\left[1.645 - \frac{\left(\frac{\ln(1-R)}{365}\right)}{s_{\beta}}\right] + \Phi\left[-1.645 - \frac{\left(\frac{\ln(1-R)}{365}\right)}{s_{\beta}}\right] \quad (A-13)$$

where  $\Phi$  is the normal cumulative distribution function. (Substituting A-10 into one of the cumulative functions demonstrates the symmetry between power to detect reductions and increases.) Thus, all of the concerns relevant to the power to detect a size R reduction in species concentration enter into the power calculations via the variability associated with  $B_2$ , the estimate of  $\beta_2$ . Clearly, smaller  $s_{\beta}^2$  values increase power and larger  $s_{\beta}^2$  values reduce power.

To quantify  $s_{\beta}^2$ , recall that B (3x1), the least squares estimate of  $\beta$  is defined by

$$B = (X^T \Sigma^{-1} X)^{-1} X^T Y \quad (A-14)$$

Thus, the covariance of B (3x3) is

$$Cov(B) = (X^T \Sigma^{-1} X)^{-1} \quad (A-15)$$

where the sampling frequency (1 in D day), the sampling duration (Y years), and the measurement error assumed (F factor) all enter into the calculation of the covariance of B through the parameters governing  $\Sigma$  (D and F) and the size of the matrices (i.e., the number of observations,  $N = (365/D) * Y$ ). The quantity relevant to power calculations,  $s_{\beta}^2$ , is the third diagonal element of the covariance matrix of B. The impact on power of sampling more frequently (decreasing D) is less than the impact of sampling longer (increasing Y) because of the trade-off in the effect of D on sample size (increasing N) and the effect of D on  $\Sigma$  (increasing correlation).

Using the power function described above, which quantifies power as a function of *percent reduction in species concentration per year*, a number of power curves were calculated and graphed. The figures provided (Figures A-2, A-3, and A-4) assume coefficient of variation, auto-correlation, and baseline measurement error equal to the values estimated for the urban Washington DC site. The figures vary sampling frequency (1 in 6, 1 in 3, and daily sampling), duration of sampling (3 and 5 years), and measurement errors (equal to and 2 times the Washington site's). In order, the figures are for 1 in 6 day sampling, 1 in 3 day sampling, and daily sampling. Duration of sampling and measurement error are varied within the figures. The left column assumes three years of sampling, the right 5 years. The first row assumes measurement error equal to IMPROVE's, the second twice IMPROVE's measurement error.

By inverting the power function, an equation providing the percent reduction (or increase) necessary to achieve a prescribed power can be calculated. Table A-1 presents the percent reduction (or increase) per year which yields a power of 0.8.

**Table A-1. Percent Reduction or Increase (per year) Yielding 0.8 Power**

Species	Measurement Error Factor	Daily Sampling		1 in 3 Day Sampling		1 in 6 Day Sampling	
		3 Years	5 Years	3 Years	5 Years	3 Years	5 Years
Sulfate	1	7.5	3.6	8.6	4.1	10.9	5.2
	2	7.5	3.6	8.6	4.1	11.0	5.2
Nitrate	1	12.2	5.9	13.0	6.3	15.3	7.4
	2	12.2	5.9	13.1	6.3	1.53	7.4
Calcium	1	7.4	3.5	8.5	4.1	10.9	5.2
	2	7.7	3.7	8.8	4.2	11.2	5.4
Total Carbon	1	5.5	2.6	7.3	3.4	10.0	4.8
	2	5.6	2.6	7.4	3.5	10.2	4.9

### 3.0 Meteorological Analysis

Meteorological data was obtained from the IMPROVE network. Through this network both nephelometer and transmissometer data were made available. A quick comparison of the two measures did not reveal any obvious benefits for using one over the other. Therefore, the transmissometer data was chosen because of the simplicity of the data structure. The key accompanying the data indicated a relative humidity validity code value equal to 0 identifies a valid measurement. Only these values were kept. The key did not indicate which values were valid for temperature. A short analysis revealed a need for some validity criterion. Therefore, any temperatures exceeding 45°C were set to missing. Daily averages for both the relative humidity and temperature were calculated. Approximately, 25 sites were used in the final analysis.

A two stage approach was used to evaluate the effect of adding the meteorological data into the analysis of the species concentrations. The first stage involved modeling the species concentrations as a function of the meteorological parameters. The following model was used:

$$Y_i = b_1 T_i + b_2 RH_i + b_3 T_{i-1} + b_4 RH_{i-1} + b_5 T_i * RH_i + e_i \quad (A-16)$$

where  $Y_i$  is the  $i^{\text{th}}$  day's log-transformed species concentration measurement,  $T_i$  and  $RH_i$  are the  $i^{\text{th}}$  day's average ambient temperature and average relative humidity, respectively.  $T_{i-1}$  and  $RH_{i-1}$

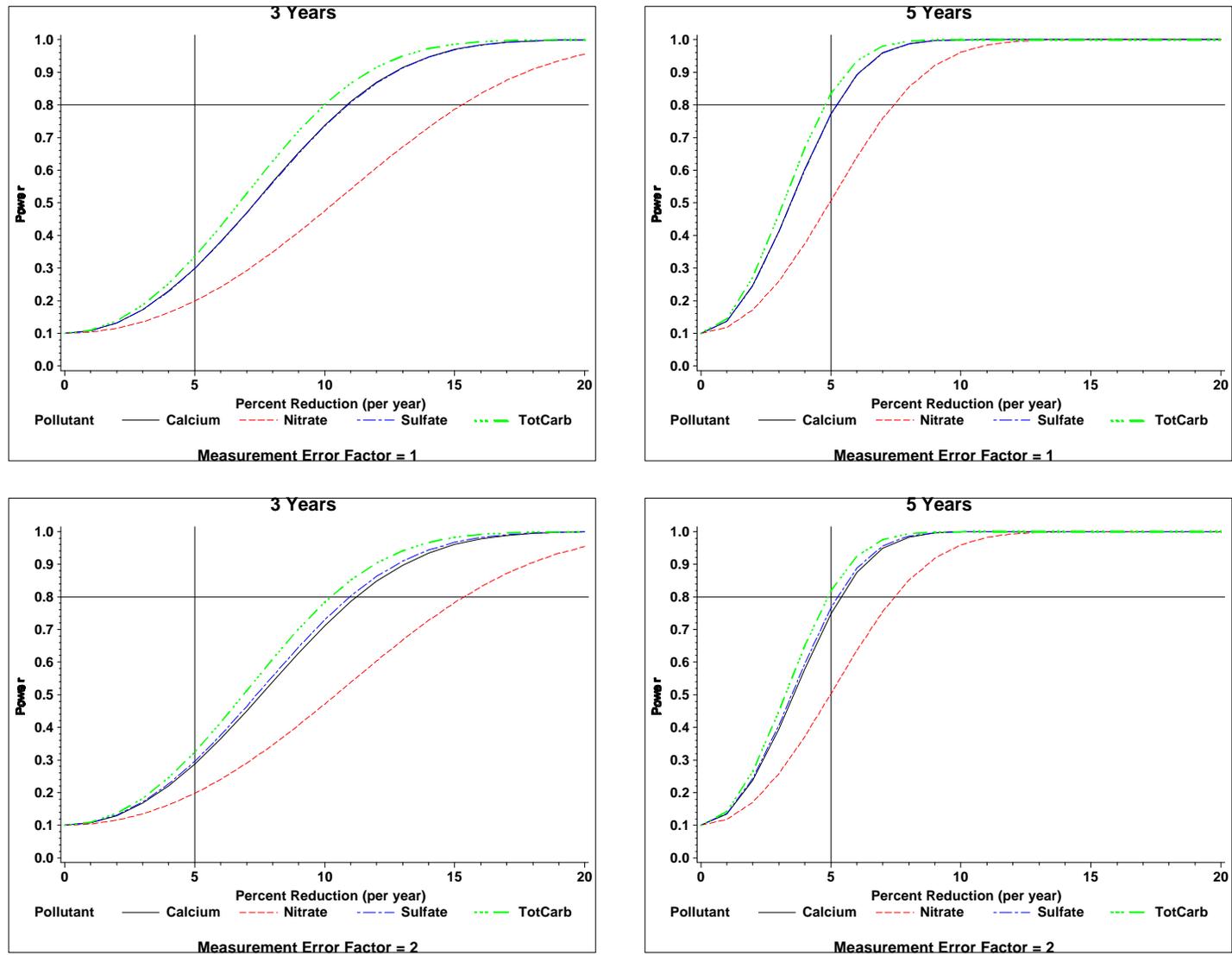


Figure A-2. Power Curve Comparisons for 1 in 6 Day Sampling Between 3 and 5 Year Duration, Assuming Washington IMPROVE Data

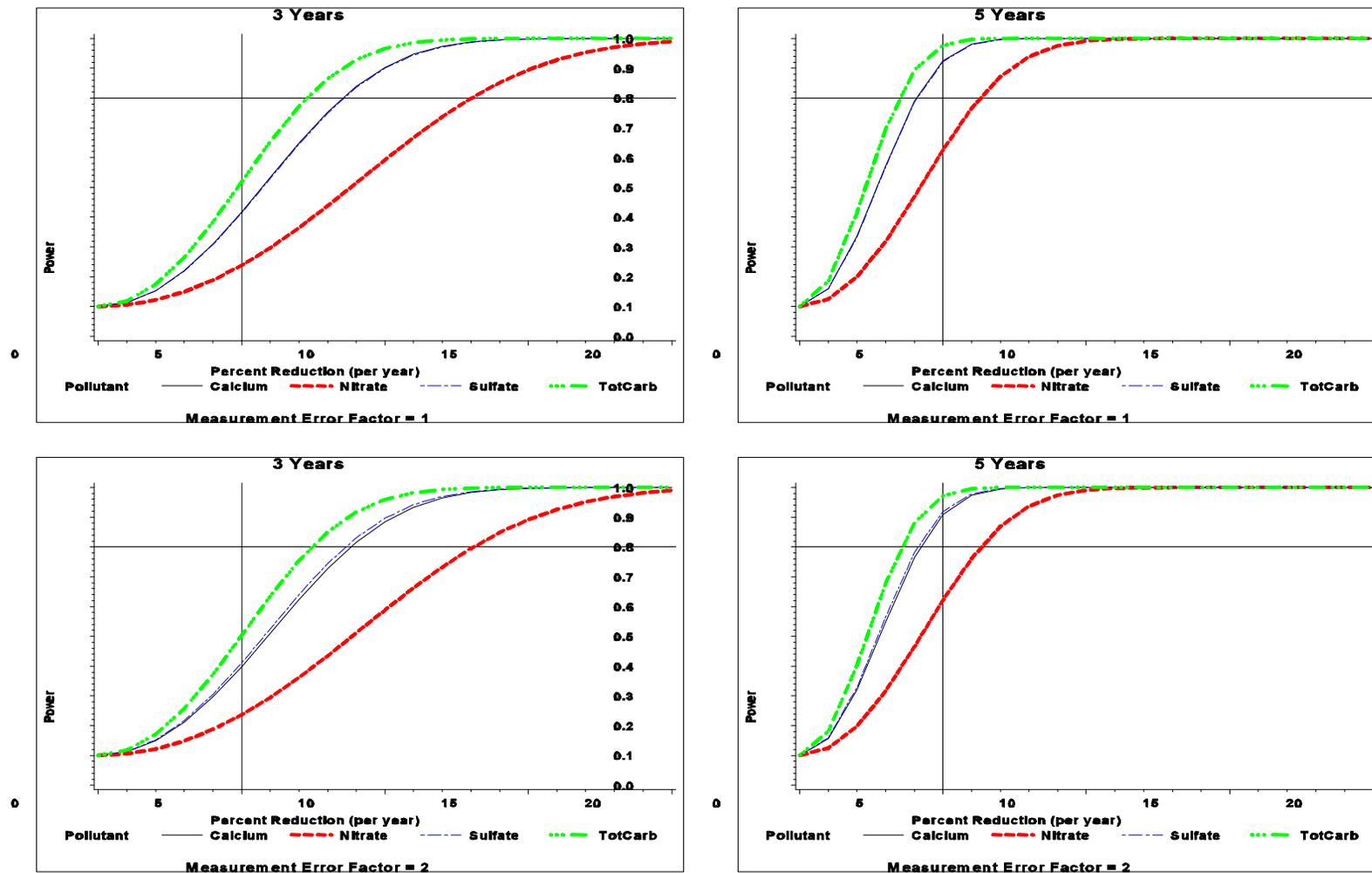


Figure A-3. Power Curve Comparisons for 1 in 3 Day Sampling Between 3 and 5 Year Duration, Assuming Washington IMPROVE Data

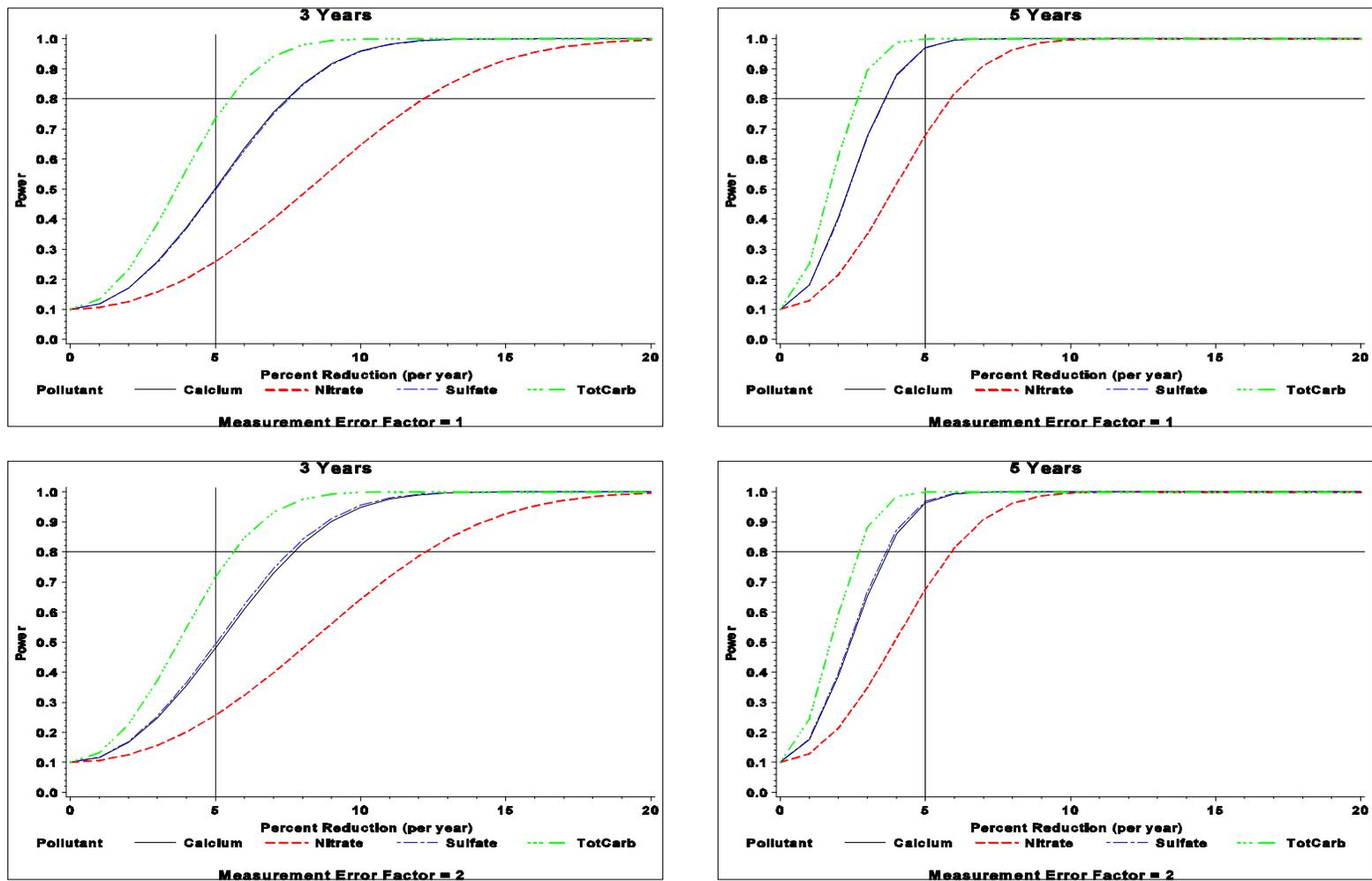


Figure A-4. Power Curve Comparisons for Daily Sampling Between 3 and 5 Year Duration, Assuming Washington IMPROVE Data

**Table A-2. IMPROVE Data Summary – Meterologically Adjusted Data**

Parameter	Seasonality (Ratio of Summer Peak Concentration to Winter Trough Concentration)		Time Trend (Ratio of One Year's Concentration to Previous Year's Concentration)		Auto-correlation (Estimated Correlation Coefficient Between 6 Day Apart Log-transformed Concentrations)		Baseline Concentration (Estimated Concentration During Winter Trough of 1988, mg/m <sup>3</sup> )		Error (Coefficient of Variation)	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
Sulfate	1.327	0.797-2.605	2.70%	-1.4%-6.5%	0.210	0.069-0.340	NA	NA	0.602	0.412-0.790
Nitrate	1.609	0.631-4.508	4.30%	-2.8%-8.4%	0.207	0.082-0.303	NA	NA	1.038	0.673-2.379
Calcium	1.586	0.714-2.927	4.80%	-6.8%-28.6%	0.360	0.121-0.487	NA	NA	0.713	0.527-0.957
Total Carbon	1.009	0.658-1.849	3.30%	2.0%-9.9%	0.254	0.062-0.470	NA	NA	0.513	0.382-0.693

are the temperature and relative humidity from the previous day, respectively.  $T_i * RH_i$  is the interaction between the relative humidity and temperature for the  $i^{th}$  day and  $\epsilon_i$  is the random error associated with the  $i^{th}$  observation. The residual, or unexplained variability from this analysis was then used as the response variable for the time trend analyses mentioned earlier.

The first stage of the modeling yielded ranges of  $R^2$  between .06 and .5 for sulfate, .04 and .3 for nitrate, .24 and .64 for calcium, and .09 and .5 for total carbon. The results from the meteorological adjustment of the  $PM_{2.5}$  species data are presented in Table A-2. (Meteorological data was not available at the Washington IMPROVE site.) Thus, variability and auto-correlation indicated in Table A-2 are present in spite of accounting for meteorology. The seasonal effects indicated in Table A-2 are significantly reduced from those indicated based on the unadjusted data ( Table 2 of the DQO document ) as expected. Variability and auto-correlation values, however, are not changed significantly from the values reported when meteorology was not accounted for (Table 2). Thus, we conclude power calculations, which depend on variability and auto-correlation, can be performed without adjusting for meteorology at least as considered in this simple model (A-16).

#### 4.0 Measurement Error Analysis

The power curves presented in Section 2.0 suggested that the impact of increasing measurement error was small. This is because measurement error explains a relatively small percentage of the residual variability left after fitting the model accounting for seasonal and time trend effects, e.g., for sulfate in Washington

$$\frac{S_{ME}^2}{S^2} = \frac{\ln[ME_{cv}^2 + 1]}{\ln[cv^2 + 1]} = \frac{0.0013}{0.285} = 0.45\% \quad (A-17)$$

To explore the impact of measurement error if residual variability could be further reduced by building a better model to explain species concentrations, perhaps one accounting for meteorology in a sophisticated manner, further measurement error power calculations were performed.

Once again the focus was turned to the Washington data to better emulate the variability associated with urban sites.  $R^2$  is a measure of the amount of variability that can be explained by the model. In terms of partitioned error,  $R^2$  can be written as follows:

$$R^2 = SSR / (SSE + SSR) = 1 - (SSE / (SSE + SSR)) \quad (A-17)$$

where SSR is the (sum of squared) error explained by the model or true error and SSE is the residual or (sum of squared) unexplained error. Inspection of the  $R^2$  values from the time trend and seasonality analysis shows a considerable amount of variability is left unexplained. The  $R^2$  values are .19, .22, .05 and .03 for sulfate, nitrate, calcium and total carbon, (Washington site) respectively. An artificial inflation of  $R^2$  would represent a smaller residual error. Assuming a

variance components relationship between residual error,  $\sigma^2$ , and measurement error,  $S_{me}^2$ ,

$$S^2 = S_{me}^2 + S_e^2 \quad (\text{A-18})$$

measurement error makes up only a tiny portion of this unexplained variability. Thus, artificially increasing  $R^2$  would make measurement error a larger portion of the error. Given that:

$$S^2 = SSE/df \quad (\text{A-19})$$

and the relationship between  $R^2$  and SSE stated above we can back solve to find  $\sigma^2$ , as a function of  $R^2$ . New power curves (Figure A-5) were created assuming hypothetical model  $R^2$  values. An  $R^2$  of .7 was considered reasonable for some Ozone models. We used this  $R^2$  to create the top two plots in Figure A-5. The plot on the left represents while the plot on the right represents a doubling of this error. The bottom two plots IMPROVE's measurement error are for an assumed  $R^2$  of .95. Both sets of plots represent 1 in 3 day sampling over a one year time frame. Table A-3 presents the percent increase or reduction per year which yields a power of 0.8. It is clear that even at an  $R^2$  of .7 measurement error is such a small portion of the residual error that doubling it had little effect.

**Table A-3. Percent Increase or Reduction (per year) Yielding 0.8 Power Assuming 1 in 3 Day Sampling for 1 Year While Varying  $R^2$**

Species	Measurement Error Factor	$R^2 = .70$	$R^2 = .95$
Sulfate	1	24.8	11.0
	2	25.1	12.0
Nitrate	1	36.4	16.9
	2	36.8	18.2
Calcium	1	23.0	10.2
	2	25.2	15.2
Total Carbon	1	19.7	8.6
	2	21.1	11.8