September 29, 2008

MEMORANDUM

SUBJECT: Documentation of DQO Process Outputs and Proposals for Quality Control of Pb data from SLAMS Sites

FROM: Michael Papp (EPA, OAQPS, AQAD, AQAG) Michael Papp

TO: Pb Monitoring Rule Docket (EPA-HQ-OAR-2006-0735)

This memo presents summary findings from an analysis that assessed the significance of data completeness, sampling frequency, precision and bias on the Pb design value averaging times and identifies the measurement quality objectives for the data quality indicators of precision and bias.

Executive Summary

The EPA is in the process of revising the lead (Pb) National Ambient Air Quality Standards (NAAQS). As part of the process, the monitoring requirements for Pb are being reviewed and include monitoring options like a:

- Change in *averaging time* of the indicator from the fixed quarterly average to either a rolling quarterly or monthly average; and a
- Change in *monitoring device* from the current Federal Reference Method (FRM), the high-volume TSP sampler, to the low-volume PM_{10} sampler.

Using the DQO Process, EPA explored how changes in design value averaging times, sampling frequency, data completeness, precision and bias affect ones ability to compare Pb estimates to a NAAQS value. RTI, in coordination with Neptune and Company, worked with EPA to create a Pb data set which could then be used to evaluate various data quality scenarios based on the two design value averaging times (monthly and rolling quarterly). The scenarios included:

- Two completeness scenarios (75% and 90%)
- Three sampling frequencies (every day, every three days, every six days)
- Three precision scenarios (10%, 20% and 30%)
- Six bias scenarios ($\pm$ 5%, $\pm$ 10%, $\pm$ 15%)

The evaluation used over 130 Pb ambient air routine monitoring locations. We initially reviewed the data to determine whether we needed to perform DQO assessments of source and non-source oriented sites separately. The data assessment revealed that although concentrations were different, temporal variability of the two site types were similar and one model representing the temporal variability at a monitoring site could be developed. The next step was to select a model to represent a hypothetical monitoring site with a certain amount of variability. We selected six routine monitoring sites that had greater than average temporal variability (in the 80-90 percentile range of the data set) to construct this model. Because temporal variability and mean Pb level are related, these sites also had mean Pb levels higher than the average of all the data. The selection of a model with more than average variability even compared to sites with similar mean values, but within the universe of real data currently being reported, is consistent with the approach used several years earlier to generate ozone and $PM_{2.5}$ DQO's. A detailed report of this

Pb data selection/modeling activity is provided in Attachment A. Tables 1 and 2 provide the descriptive statistics of the population model and Figures 1 and 2 provide graphic representations of the data.

Table 1. Average Monthly Design Value Statistics for the Hypothetical Monitoring Site (µg/m^3). (Daily Sampling, 100% Completeness, No Decrease in Precision or Increase in Bias)

N	Min.	Q(.025)	Q(.25)	Med.	Mean	Q(.75)	Q(.975)	Max.	SE
2500	0.0563	0.0563	0.0890	0.1119	0.1221	0.1385	0.2329	0.2810	0.0487

Table 2. Average Rolling Quarterly Design Value Statistics for the Hypothetical Monitoring Site (µg/m^3). (Daily Sampling, 100% Completeness, No Decrease in Precision or Increase in Bias)

N	Min.	Q(.025)	Q(.25)	Med.	Mean	Q(.75)	Q(.975)	Max.	SE
2500	0.0776	0.0776	0.0997	0.1163	0.1221	0.1375	0.1843	0.1937	0.0297

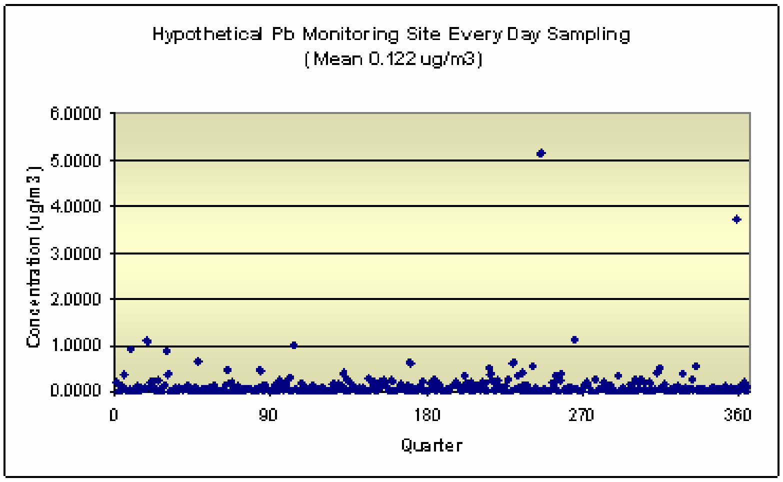


Figure 1 Hypothetical Pb Site

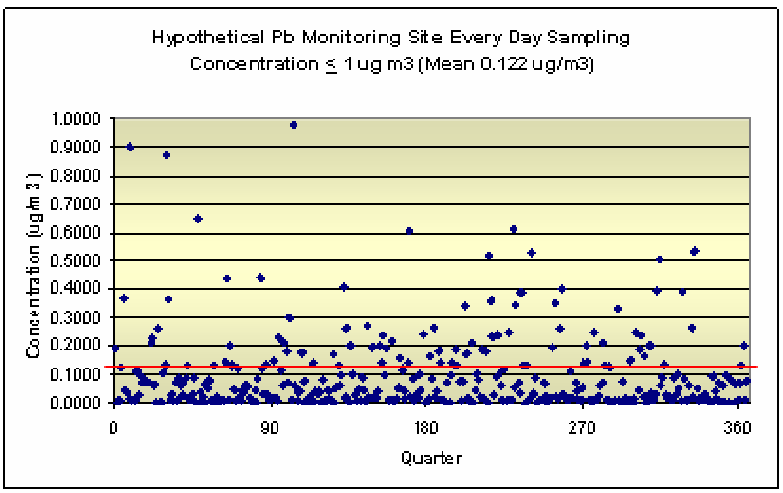


Figure 2 Hypothetical Pb Site (mean value in red)

Once the model is constructed, we use it as a tool to evaluate the effect each scenario has on the uncertainty around either the monthly or rolling quarterly estimate of average Pb concentration. The metric used to measure this significance was the difference between the upper and lower 95% confidence limits (CL) for the variability about the mean (0.122 µg/m^3) of the hypothetical monitoring site.

In order to provide a way to illustrate all the scenarios evaluated, Figure 3 was developed in a manner that two lines (top two lines) would represent data for the monthly average at data completeness of 75% and 90% , and two lines (bottom two lines) would represent data for the quarterly average at the same data completeness of 75% and 90%. Each point along the lines represents a scenario of sampling frequency, imprecision and bias and represents the difference (width) of the upper and lower 95% CLs of the data where the true (hypothetical) mean is 0.122 µg/m^3. Each scenario was simulated 2,500 times.

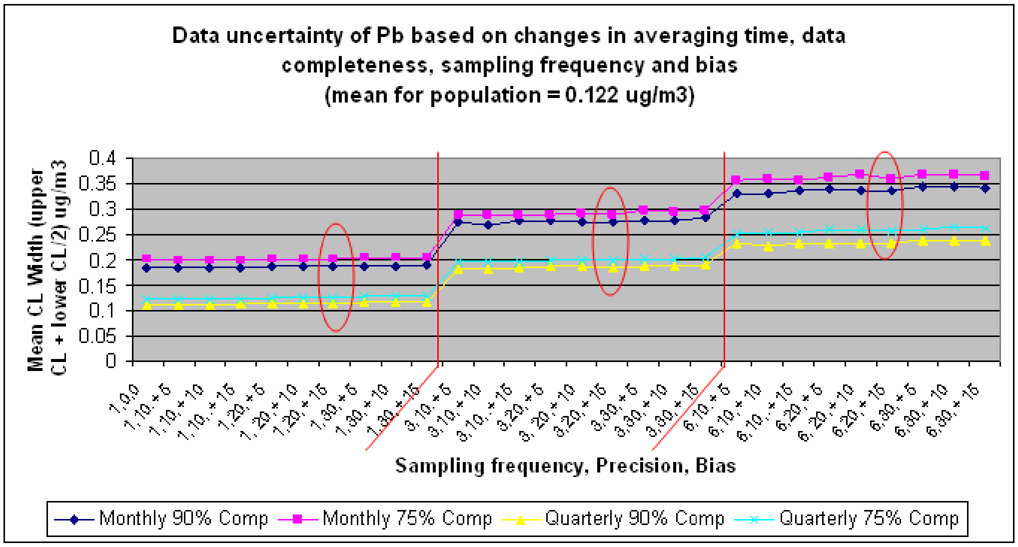


Figure 3. Summary of data uncertainty scenarios

As an example, lets take the scenario of the quarterly estimate with 90% completeness with everyday sampling at 10% precision and ± 5% bias. This scenario is represented by the second point on yellow line in Figure 3. The model predicted an:

- upper CL of 0.1976 μg/m^3 and a lower CL of 0.0797 μg/m^3 for a positive 5% bias, (width 0.1179 μg/m^3) and,
- an upper CL of 0.1787 μg/m^3 and a lower CL of 0.0723 μg/m^3 for a negative 5% bias (width 0.1064 μg/m^3)
- the average of the two CL widths (0.1179 + 0.1064 /2) is 0.1121 μg/m^3

Using an average width negates the affect of bias in this figure, but it does provide a better estimate of the width of the CLs for any scenario and provides a general assessment of magnitude of different variability scenarios. The effect of bias is shown later in this summary (see Fig. 4).

NOTE: the width of quarterly 95% CL with 0% precision and 0% bias (90% completeness) is 0.1115 μg/m^3 which is represented as the first point on the yellow line (as well as the first point of each line). The width is very close to the data points with increasing amounts of imprecision and bias which indicate that the majority of the variability is natural (day-to-day) temporal variability (see Fig. 2).

Observations

Figure 4 provides another way to illustrate the affects of the various data quality scenarios. Each box and whisker plot represents the distribution (both positive and negative bias) all the data quality scenarios associated with one particular variable. For example the first chart in Figure 4 represents the monthly and quarterly averaging times. The box and whisker represent the differing widths of the CL as one changes sampling frequency, completeness, precision and bias. Since all the scenarios are applied in the same manner for any one variable (e.g. averaging time) one can determine which variables have the most impact on uncertainty.

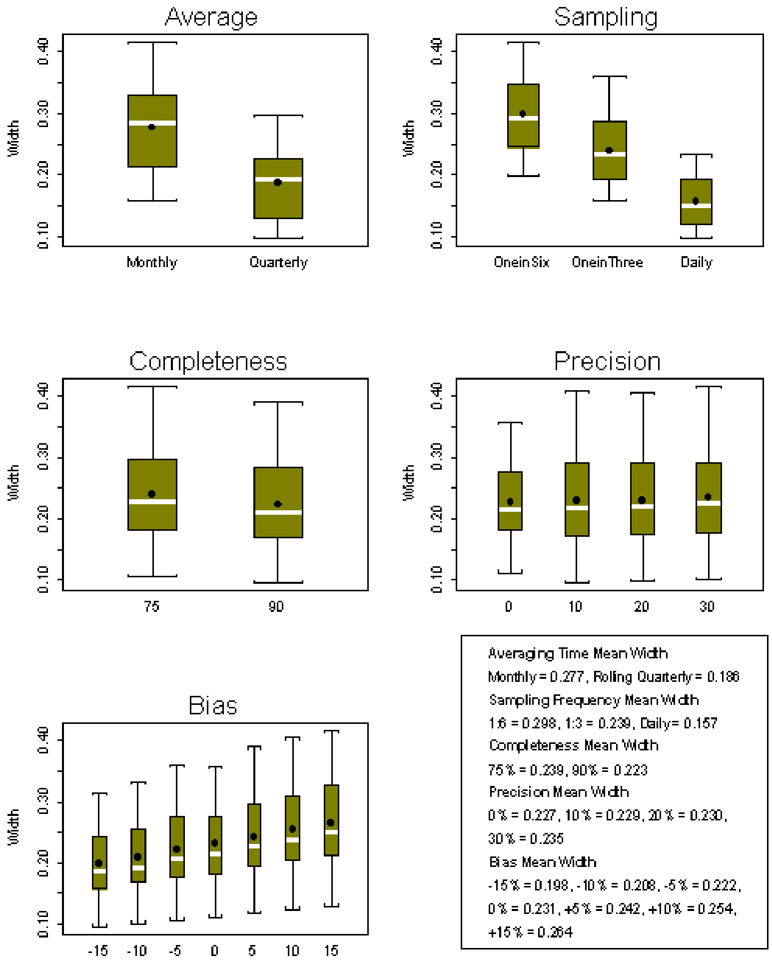


Figure 4 Effect of data quality scenarios on the data

- All independent variables- design value averaging time, sampling frequency, completeness, precision and bias- have a statistically significant impact on the width of the confidence interval for the mean (p-value < 0.0000).
- The design value averaging time and sampling frequency have the greatest impact on the width but the two variables interact. That is the decrease in the mean width with the increase in sampling frequency is greater for the monthly averaging design value than the rolling quarterly average.
- The change in design value averaging time also interacts with the change in bias. The decrease in the mean width with the decrease in bias is greater for the monthly averaging design value than the rolling quarterly average.
- The change in data completeness has less impact on mean width than the change in design value averaging time, sampling frequency and bias
- The change in precision has the least influence on mean width

Proposed Measurement Quality Objectives for Precision and Bias Data Quality Indicators

Precision --

Data quality indicators are quantitative statistics and qualitative descriptors that are used to interpret the degree of acceptability or utility of data to the user. The principle data quality indicators are precision, bias, completeness, comparability, representativeness and detectability. A measurement quality objective is a goal set by EPA guidance that represents a reasonable expectation of what one should be able to achieve for a specific data quality indicator in order to maintain acceptable levels of uncertainty. EPA reviewed precision data from various sources including routine Pb data from the SLAMS, National Air Toxics Trends Sites and Chemical Speciation Network Sites; this Pb data was collected by various sampling and analytical methods. Table 3 provides a comparison of this data. The data represent eight precision assessments due to use of either a different sampling method or a different analysis method. As with our other particulate-based criteria pollutants, EPA identifies a "cutoff" concentration value and precision estimates are made only using pairs of precision values that are equal to or above this cutoff value. At low concentrations, agreement between measurements of collocated values, expressed as relative percent difference, is understandably poor but at such low concentrations precision is not an important objective for air quality purposes. Prior to the new Pb NAAQS standard, the collocated precision cutoff value was 0.15 µg/m^3. With the lowering of the NAAQS, and improvements in sampling and analytical technologies, EPA feels this cutoff value can and should be lowered. The data in Table 3 was reviewed at a number of potential cutoff values; starting at 0.002 µg/m^3, which is the proposed method detection limit (MDL) for the XRF-based FRM for Pb-PM_{10}, and up to 0.02 µg/m^3. Some scenarios in Table 3 do not show the 0.01 or 0.02 µg/m^3 scenarios because there were not enough (or no) routine data concentrations in these ranges. Based on our evaluation, we believe that 0.02 µg/m^3 is an appropriate cutoff value for two reasons: 1) there has been an established concept of a "limit of quantitation" that is usually estimated at ten times the MDL, and 2) it is potentially one order of magnitude away from the NAAQS (if the NAAQS is set at 0.2 µg/m^3) and provides an adequate margin of safety for data review. As an alternative, EPA could consider 0.01 µg/m^3 as a cutoff but we do not recommend going below this concentration. Based on this cutoff value and reviewing the historical data in Table 1 at or above the 0.02 µg/m^3 cutoff value, EPA proposes a precision measurement quality objective of 20% for a 90% confidence limit coefficient of variation, aggregated over a 3-year period at the primary quality assurance organization level. This means that the large majority of paired precision data should show a difference below 20%; monitoring organizations that do not achieve this result would be advised of the problem and encouraged to investigate and resolve the causes of the disagreements.

Bias--

Estimates of Pb bias were evaluated by reviewing data collected through the $PM_{2.5}$ Chemical Speciation Network (CSN) and the National Air Toxics Trends Stations (NATTS) QA programs. Data was evaluated using the DASC Tool which calculates bias by current 40 CFR Appendix A methods. As with precision, cutoff values are also used for bias data. For the reasons provided in the precision section a cutoff value of 0.02 µg/m^3 is proposed.

CSN Data –

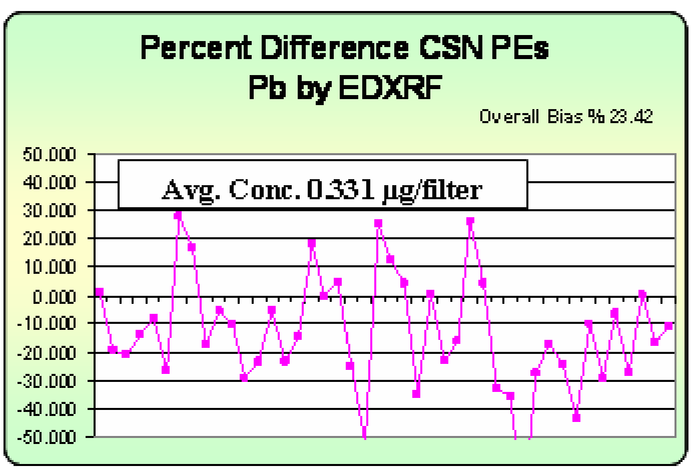


Figure 5. CSN XRF Bias information

The XRF bias estimates for the $PM_{2.5}$ CSN were obtained from data provided by the analysis of Performance Evaluation (PE) samples. Multiple laboratories provide XRF analyses in support of the CSN (Research Triangle Institute, California Air Resources Board, Oregon DEQ, Desert Research Institute). CSN PE samples consist of "real-world" particle filters collected over multiple days to ensure that an adequate amount of material is present for analysis. For XRF, 46.2-mm Teflon filters were collected and analyzed by an EPA reference lab prior to distribution. Figure 5 provides the individual percent difference results for 44 observations by the three laboratories. The average concentration in µg/filter was 0.331 µg/filter and the equivalent concentration in µg/m^3, based on 24 m^3 (16.7 Lpm sampling), was 0.0138 µg/m^3. The overall absolute bias upper bound for the 95% percentile is 23.42%.

NATTS Data —

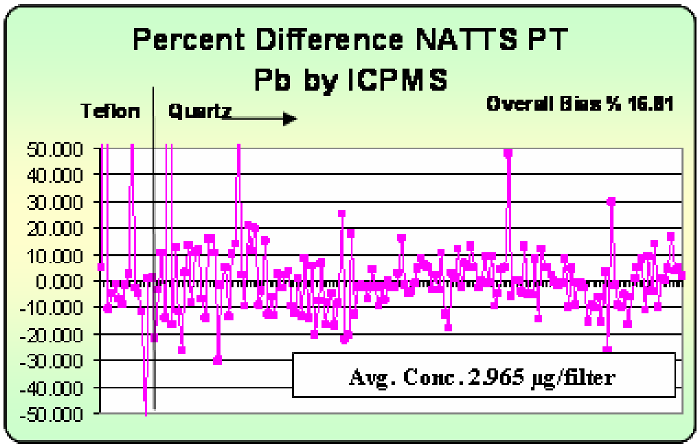


Figure 6. NATTS ICPMS Bias information

Bias estimates for the NATTS were obtained from data provided by the analysis of Performance Evaluation (PE) samples by ICP-MS. Several laboratories provide ICP-MS analyses in support of the NATTS. NATTS PE samples consisted mostly of 46.2-mm quartz fiber filters that are produced by the aerosolization and deposition of a Pb-salt solution onto each filter. The size distribution of the liquid aerosol was not controlled or characterized. Initially Teflon filters were used and then switched to quartz filters to match the filter material used by the NATTS. The filters were prepared and analyzed by ICP-MS at a reference lab prior to distribution. Figure 6 provides the individual percent difference results for 175 observations. The average concentration in µg/filter was 2.965 µg/filter and the equivalent concentration in µg/m^3, based on 24 m^3 (16.7 Lpm sampling) was 0.1236 µg/m^3. The overall absolute bias upper bound for the 95% percentile is 16.81%.

It is important to note the differences in the PE samples generated for each program as these differences have the potential to affect the bias estimates. The XRF bias estimate is based on $PM_{2.5}$ particles collected in the field and include any associated particle or sample "matrix" effects. For NATTS, the ICP-MS PEs samples are lab-generated liquid aerosols. In addition, the XRF PE samples are at a concentration level that is one order of magnitude lower than the ICP-MS PE samples (0.331 versus 2.965 µg/filter) and at an equivalent concentration (0.0138 µg/m3). It should be observed that this equivalent concentration is below the proposed cut off value. Therefore, one might expect for XRF bias results to comparable to the NATTS bias results if values above the proposed cutoff are used.

Based on this cutoff value and reviewing the CSN and NATTS data, EPA is proposing an overall absolute bias upper bound goal of 15%. The XRF bias estimate of 23.4% is expected to improve at concentrations 10 times higher than those evaluated. The ICP-MS bias estimate of 16.81% is in line with the proposed goal. This means that the large majority of bias data should show difference below 15%; monitoring organizations that do not achieve this result would be advised of the problem and encouraged to investigate and resolve the causes of the disagreements.

Summary

The report provides a summary of the influences of uncertainty around design value averaging times, sampling frequency, completeness, precision and bias. As is normally the case with environmental data, natural spatial and temporal variability represent the largest amount of uncertainty. Measurement uncertainty is then influenced, in order of largest to smallest effect, by averaging time, sampling frequency, bias, completeness and precision. Based on a review of precision and bias data from various sources, AAMG suggests that measurement quality objectives (MQOs) for precision be initially established at 20% and bias at $\pm$15%. This precision and bias combination for each completeness and sampling frequency scenario is represented by the red circles in Fig 3. For completeness, it appears 75% could be considered acceptable. Most data reviews show routine data completeness higher than 75% and EPA could provide stronger guidance (extra samplers available for key sites or collocated precision at key sites for data substitution) to ensure higher data completeness. AAMG is not in a position to make decisions on design value averaging times or sampling frequency but quarterly design value averaging times and higher sampling frequency significantly reduce data uncertainty. It is possible that sampling frequency be increased (through regulation) as data approach the NAAQS values thereby reducing data uncertainty as much a practicable.

Table 3. Precision Summary

Summary of Pb Precision Data from Collocated Samplers

90% CV = 90% Confidence Limit Coefficient of Variation **Ave Abs. PD** = Average Absoulte Value Percent Difference **Med. Abs. PD** = Median Absolute Percent Difference

1. PM10 NATTS Pb High-volume sampling (~113 LPM) Analysis ICP-MS

Years 2001-2006

	n	90%CV	Ave. Abs. PD	Med. Abs PD
PM10 Pb All	637	23.2	16.6	7.6
PM10 Pb > 0.002 ug/m3	521	19.4	13.4	6.6
PM10 Pb > 0.006 ug/m3	184	20.7	14.3	6.3
PM10 Pb > 0.01 ug/m3	90	11.2	8.7	5.1
PM10 Pb > 0.02 ug/m3	33	12.0	9.3	4.2

2. TSP Pb High-volume sampling (~113 LPM) Analysis ICP-MS

Years 2002-2006

	n	90%CV	Ave. Abs. PD	Med. Abs PD
TSP Pb All	657	22.8	10.6	0.0
TSP Pb > 0.002 ug/m3	614	13.0	7.0	0.0
TSP Pb > 0.006 ug/m3	530	11.8	5.4	0.0
TSP Pb > 0.01 ug/m3	475	11.7	4.4	0.0
TSP Pb > 0.02 ug/m3	314	6.7	2.5	0.0

3. TSP Pb High-volume sampling (~113 LPM) Analysis Atomic Absorption

Years 2002-2006

	n	90%CV	Ave. Abs. PD	Med. Abs PD
TSP Pb All	2239	17.6	12.7	3.6
TSP Pb > 0.002 ug/m3	2100	16.9	12.1	2.7
TSP Pb > 0.006 ug/m3	2050	16.8	12.0	2.6
TSP Pb > 0.01 ug/m3	1996	16.5	11.8	2.1
TSP Pb > 0.02 ug/m3	1749	15.0	11.0	3.0

4.TSP Pb High volume NY Data (Analysis Graphite Furnace AA)

	n	90%CV	Ave. Abs. PD	Med. Abs PD
TSP Pb All (not provided)				
TSP Pb > 0.002 ug/m3	61	9.4	9.0	7.4
TSP Pb > 0.006 ug/m3	47	8.8	8.8	8.0
TSP Pb > 0.01 ug/m3	34	8.1	7.5	6.5
TSP Pb > 0.02 ug/m3	22	9.0	7.9	5.8

5. TSP Pb Low-volume sampling Analysis XRF

Years 2002-2006

	n	90%CV	Ave. Abs. PD	Med. Abs PD
TSP Pb All	71	41.9	38.9	28.6
TSP Pb > 0.002 ug/m3	64	36.6	34.8	28.6
TSP Pb > 0.006 ug/m3	29	29.1	25.4	15.4
TSP Pb > 0.01 ug/m3	9	24.1	16.1	11.4

6. PM2.5 CSN Very-low-volume sampling (~6 & 7 LPM) Analysis XRF

Years 2001-2007

	n	90%CV	Ave. Abs. PD	Med. Abs PD
PM2.5 Pb All	2321	40.7	38.4	22.2
PM2.5 Pb > 0.002 ug/m3	1833	37.0	36.3	22.4
PM2.5 Pb > 0.006 ug/m3	711	36.1	40.5	28.9
PM2.5 Pb > 0.01 ug/m3	297	18.3	19.5	15.5
PM2.5 Pb > 0.02 ug/m3	80	14.0	14.4	10.0

7. PM2.5 CSN Texas Low-volume sampling (16.7 LPM) Analysis XRF

Years 2002-2007

	n	90%CV	Ave. Abs. PD	Med. Abs PD
PM2.5 Pb All	230	54.3	56.1	41.9
PM2.5 Pb > 0.002 ug/m3	83	23.5	24.3	21.3
PM2.5 Pb > 0.006 ug/m3	9	14.9	10.8	9.3

8. TSP Pb High-volume sampling (~113 LPM) Analysis ICAP

Years 2002-2006

	n	90%CV	Ave. Abs. PD	Med. Abs PD
TSP Pb All	827	15.5	11.3	6.1
TSP Pb > 0.002 ug/m3	826	15.5	11.3	6.1
TSP Pb > 0.006 ug/m3	808	15.4	11.1	6.1
TSP Pb > 0.01 ug/m3	759	15.4	10.5	5.6
TSP Pb > 0.02 ug/m3	559	16.4	11.9	6.9

SUMMARY

90% Coefficient of Variation Summary

Data Values	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Pb > 0.002 ug/m3	19.4	13.0	16.9	9.4	36.6	37.0	23.5	15.5
Pb > 0.006 ug/m3	20.7	11.8	16.8	8.8	29.1	36.1	14.9	15.4
Pb > 0.01 ug/m3	11.2	11.7	16.5	8.1	24.1	18.3		15.4
Pb > 0.02 ug/m3	12.0	6.7	15.0	9.0		14.0		16.4

1. PM10 NATTS Pb High-volume sampling (~113 LPM) Analysis ICP-MS
2. TSP Pb High-volume sampling (~113 LPM) Analysis ICP-MS
3. TSP Pb High-volume sampling (~113 LPM) Analysis Atomic Absorption
4.TSP Pb High volume NY Data Analysis Graphite Furnace AA
5. TSP Pb Low-volume sampling Analysis XRF
6. PM2.5 CSN Very-low-volume sampling (~6 & 7 LPM) Analysis XRF
7. PM2.5 CSN Texas Low-volume sampling (16.7 LPM) Analysis XRF
8. TSP Pb High-volume sampling (~113 LPM) Analysis ICAP

ATTACHMENT A

NEPTUNE AND COMPANY, INC.
Vicki Ann Lancaster, Ph.D.
724 Europe Street
Baton Rouge, LA 70802.6411
Voice: 225.383.1845
Fax: 886.498.1280
Email: vlancaster3@cox.net

TASK ORDER:	**05400**
WORK ORDER:	**0000**
SUBJECT:	**FINAL REPORT- Pb Data Quality Objective Assessment**
NUMBER OF PAGES:	**45**
DATE:	**08/26/2008**
REVISION:	

The Environmental Protection Agency (EPA) is in the process of revising the lead (Pb) National Ambient Air Quality Standards (NAAQS). As part of the process, the monitoring requirements for Pb are being reviewed. Monitoring options that are being reviewed include a:

- Change in *averaging time* of the indicator from the current quarterly average to a monthly average or rolling quarterly average; and a
- Change in *monitoring device* from the current Federal Reference Method (FRM), the high-volume TSP sampler, to the low-volume PM_{10} sampler.

This report provides the information needed to evaluate the consequences of a change in the averaging time and will not address the evaluation of the monitoring devices.

Monitoring Requirements: Change in Averaging Time

The following tasks were completed in order to explore how changes in averaging times, sampling frequency, data completeness, precision and bias effect data uncertainty:

- Selection of an Air Quality Standard (AQS) Pb data set with a measurement duration of 24 hours;
- From the above data set, selection of two data sets representing the hypothetical location types, source oriented and nonsource oriented;
- Evaluation of the hypothetical source and nonsource data sets, to determine whether separate tables of performance metrics should be constructed for each location type;
- Construction of a model to estimate performance metrics under the various combinations of sampling frequency, completeness, decrease in precision, and increase in bias;
- Calculation of the metrics to evaluate model performance under the various combinations of sampling frequency, completeness, and increase in the random (precision) and non-random (bias) components of measurement error, in order to compare design values.

Selection of an Air Quality Data Set

EPA provided an initial Pb-TSP data set containing 130 locations with measurements from the six years, 2001 to 2006. Descriptive statistics, time series plots and distribution plots were constructed for each of the 130 locations. This initial data analysis revealed many locations had few measurements (less than 75% completeness for a 1 in 6 sampling frequency) within a calendar year. In order to assure the data would capture seasonal trends, EPA made the decision to include only those locations that had at least 40 measurements (75% completeness for a 1 in 6 sampling frequency) for at least one calendar year. With the exception of measurements below the minimum detection limit, qualified measurements were dealt with on a measurement-by-measurement basis. All measurements identified as below the minimum detection limit (MDL) were used as recorded, in AQS, typically either a zero or one-half the MDL. Specific details on the construction the final data set are provided in Appendix A. The final data set used to select the hypothetical source and nonsource locations consisted of 41 source locations with 12,291 observations and 65 nonsource locations with 18,230 observations.

Construction of Hypothetical Source and Nonsource Locations

A hypothetical source location data set was constructed for use in model construction and design value comparisons. The goal was to construct a location based on the metric of Pb variability, where the within location variability (temporal variability) was greater than the average. The approach is similar to the one used to develop both the $PM_{2.5}$ DQO and ozone DQOs. To achieve this goal, the standard deviation for each of the 41 source locations was calculated using all the available data for each location. The descriptive statistics for the 41 source locations are provided in Appendix B, where the locations are sorted by standard deviation. The five source locations (480850007, 471870104, 171190010, 080010005, and 290930030) with standard deviations in the 80^{th} to 90^{th} percentile were selected to represent a hypothetical source location. These 1,730 observations were combined into a single data set used to represent a hypothetical source location. A plot of the source location means versus standard deviations is provided in Figure 1, where the data pairs with standard deviations in the 80^{th} to 90^{th} percentile are highlighted. The descriptive statistics for the hypothetical source population are provided in Table 1.

Figure 1. Source Location Means versus Standard Deviations (μg/m^3).
The five locations with standard deviations in the 80^{th} to 90^{th} percentile are filled circles, an intercept = 0 / slope = 1 line is provided for reference.

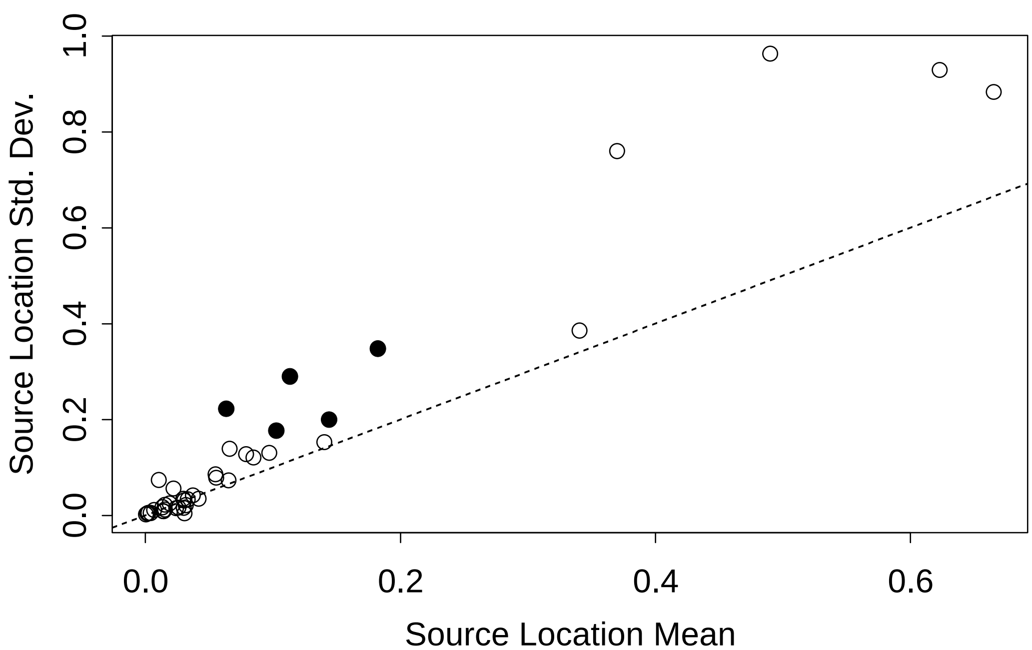


Table 1. Descriptive Statistics for the Hypothetical Source Location (μg/m^3).

N	Min.	Q(.25)	Med.	Mean	Q(.75)	Max.	SD	#0s	%0s
1730	0.0000	0.0100	0.0445	0.1214	0.1332	5.1590	0.2678	177	10.23

The hypothetical nonsource location was constructed in the same manner. The description statistics for the 65 nonsource locations are provided in Appendix B, where the locations are sorted by standard deviation. There were a total of seven source locations with standard deviations in the 80th to 90th percentile (060719004, 060371103, 170313301, 171193007, 060375001, 171630010, and 060371301). These 2,325 observations were combined into a single data set used to represent a hypothetical nonsource location. A plot of the nonsource location means versus standard deviations is provided in Figure 2, where the data pairs with standard deviations in the 80th to 90th percentile are highlighted. The descriptive statistics for the hypothetical nonsource population are provided in Table 2.

Figure 2. Nonsource Location Means versus Standard Deviations (µg/m^3).
The seven locations with standard deviations in the 80th to 90th percentile are filled circles, an intercept = 0 / slope = 1 line is provided for reference.

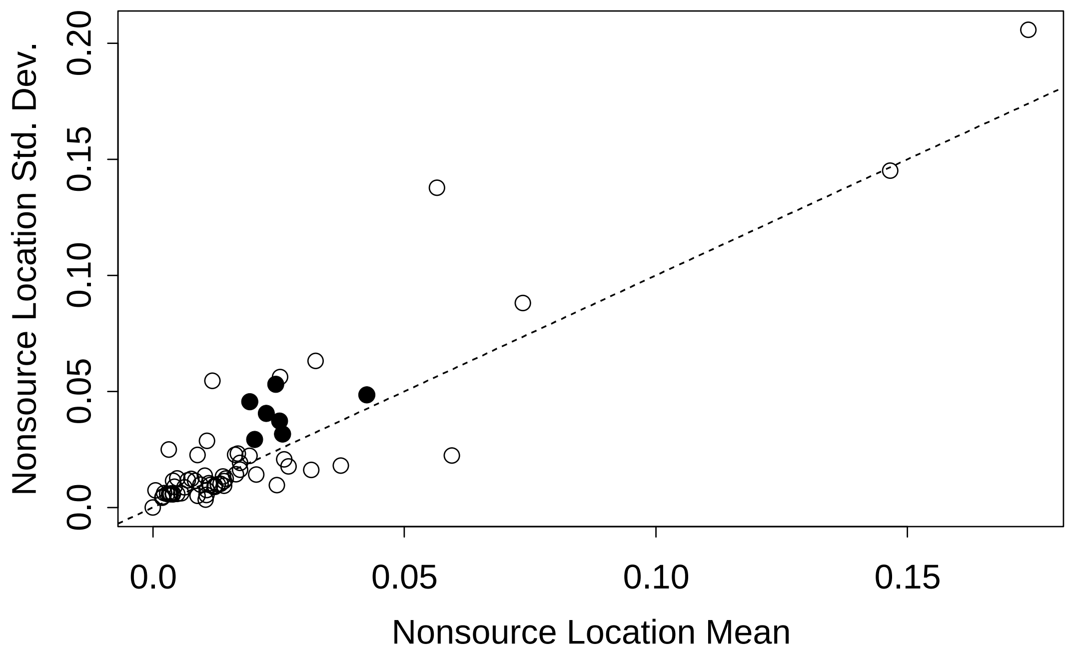


Table 2. Descriptive Statistics for the Hypothetical Nonsource Location (µg/m^3).

N	Min.	Q(.25)	Med.	Mean	Q(.75)	Max.	SD	#0s	%0s
2325	0.0000	0.0100	0.0200	0.02619	0.0300	0.9600	0.0429	106	4.56

Time series and distribution plots for the hypothetical source and nonsource locations are provided in Appendix C.

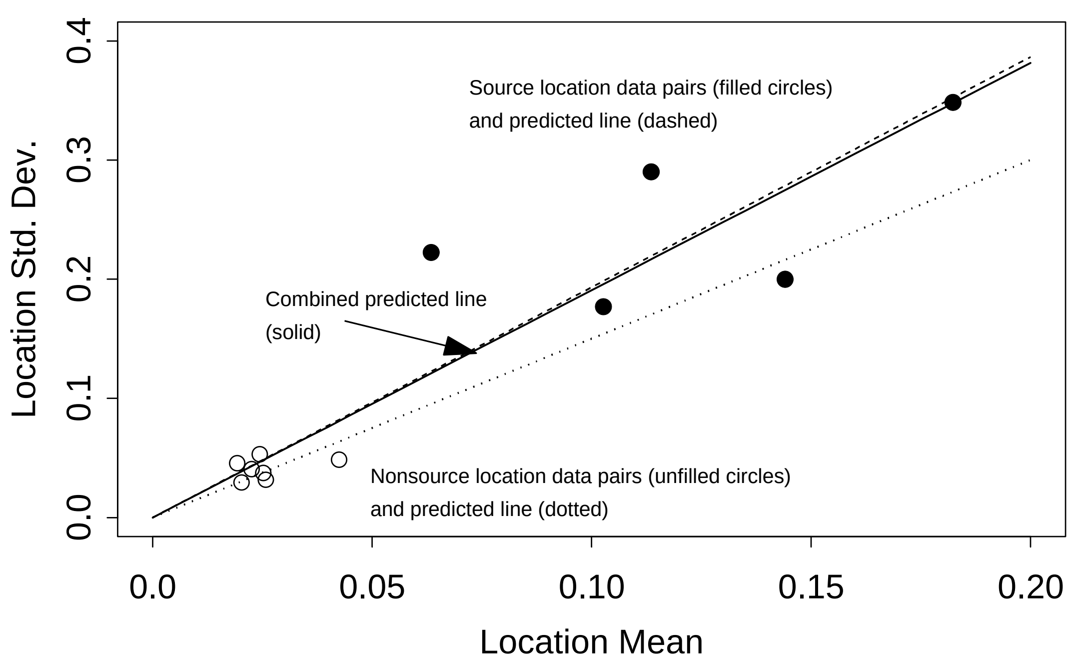
Hypothetical Source versus Nonsource Locations - One Model or Two?

The decision whether or not to construct a separate model for source and nonsource locations was made by looking at the relationship between the means and standard deviations for the locations selected to construct the two types of hypothetical locations. A separate simple linear regression model was fit to the standard deviation/mean pairs for the locations selected to construct the hypothetical source and nonsource locations. A figure of the observed data pairs and the predicted simple linear regression lines for both types of hypothetical locations is provided in Figure 3.

To evaluate source and nonsource differences, a beta-hat model was constructed to test the null hypothesis, H_o: $\beta_S = \beta_{NS}$, the source (S) and nonsource (NS) slopes are equivalent. Beta-hat models are used when the same form of model is used to describe data from more than one population, treatment, or treatment combinations. Questions about the models are answered by testing hypotheses and constructing confidence intervals about functions of parameters or a single parameter from each model. In this case a separate simple linear regression model was fit to the source and nonsource data pairs where for both models the intercept was constrained to zero (the assumption of a zero slope implies that when the standard deviation is zero the mean is zero). A figure of the observed data pairs and the predicted lines for both types of hypothetical locations is provided in Figure 3.

The results of the inferential test, H_o: $\beta_S = \beta_{NS}$, provided a *p*-value of 0.2428 (*F*-value = 1.5411 and mean square error = 0.0366) where the slope estimate for the source model is 1.9326 with a standard error of 0.2572 and a coefficient of simple determination (r^2) of 0.9339; and the slope estimate for the nonsource model is 1.4999 with a standard error of 0.1709 and a coefficient of simple determination of 0.9277. Based on the inferential test, the null hypothesis of equality was not rejected and it was concluded the relationship between the mean and standard deviation pairs for the two hypothetical locations are similar. That is, regardless of location if the mean increases by 0.1 $\mu g/m^3$ the standard deviation increases by approximately 0.2 $\mu g/m^3$, twice as much. Consequently, the data from the hypothetical source site was chosen to evaluate the design value averaging times since these location means are closer to the potential action level of 0.2 $\mu g/m^3$.

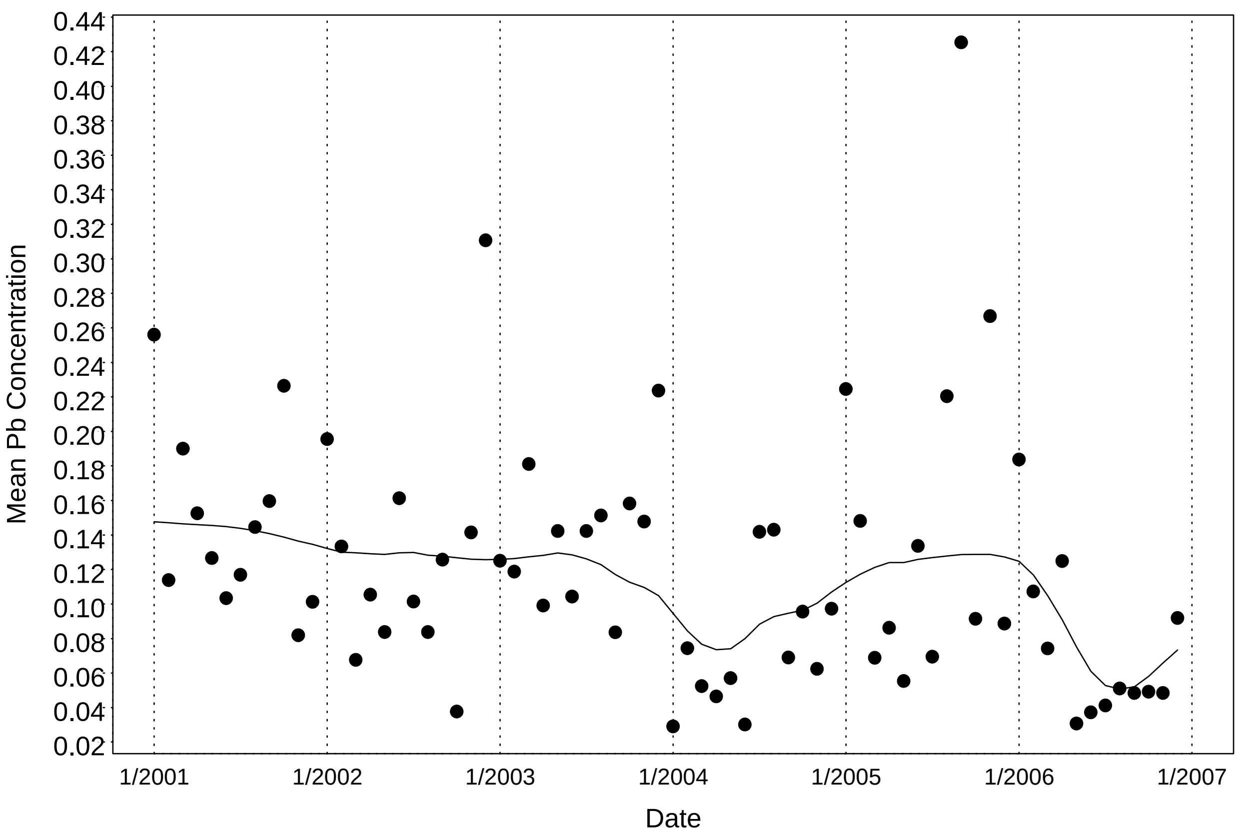
Figure 3. Hypothetical Source (filled circles) and Nonsource (unfilled circles) Location Means versus Standard Deviations ($\mu g/m^3$).
The predicted simple linear regression line for: source locations, dashed line; nonsource locations, dotted line; and both data sets combined solid line.



Model Construction

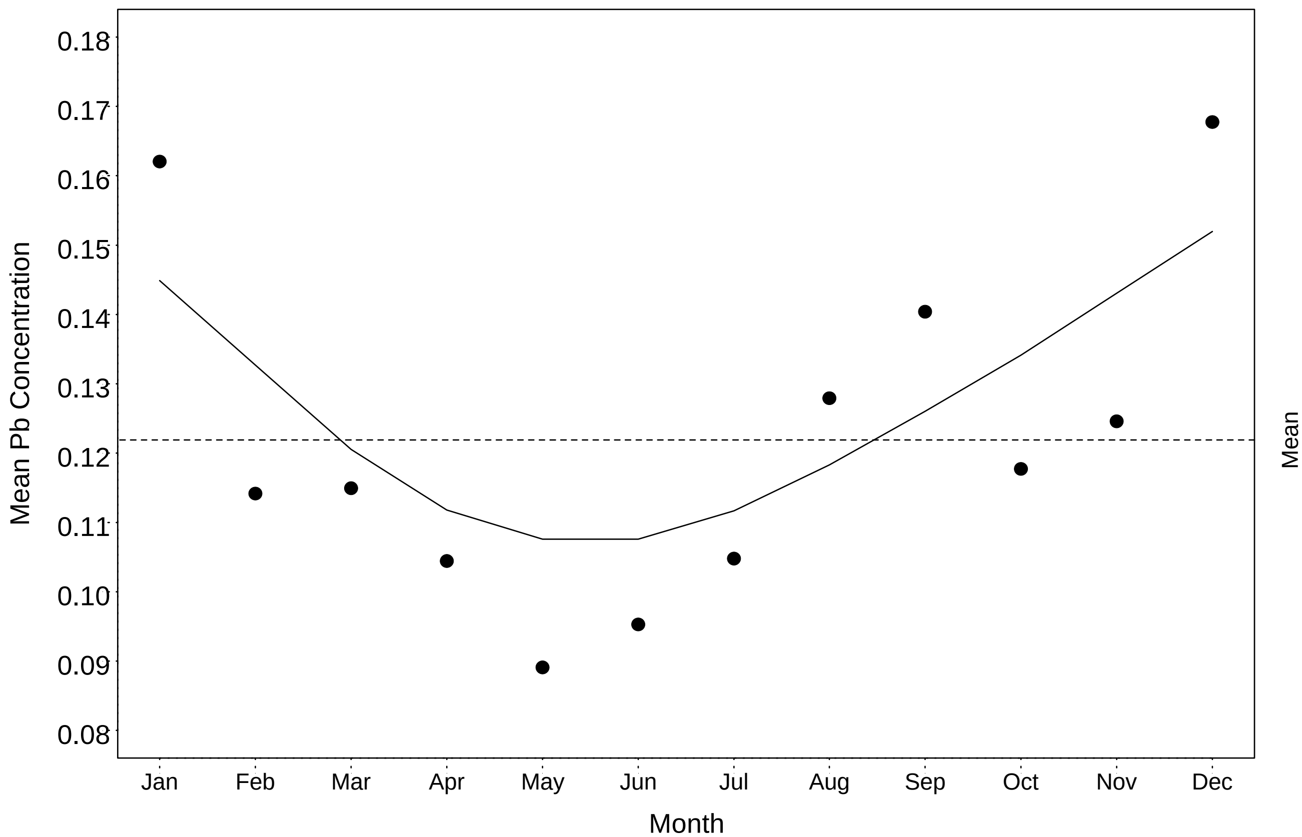
Prior to model construction the hypothetical source data set was evaluated for trends using time series plots. A plot of the year by month means for the hypothetical source location are displayed in Figure 4. A trend line is included in the figure, based on a symmetric k-nearest neighbor linear least squares procedure (k/2 data points on each side of x are used in a linear regression to predict the value at x2).

Figure 4. Hypothetical Source Location Year by Month Means with a Trend Line (μg/m[3]).



Based on Figure 4 which displays an absence of a significant trend over the time period 2001 to 2006, the decision was made to collapse the data by month to capture any seasonal trends. (Note, there is a slight decreasing trend when a simple linear regression model is fit to the data in Figure 4. But the slope of -0.0008 is not statistically different at a significance level of 0.05.) Seasonal trends are displayed in Figure 5 where the monthly means are plotted for the data combined over the years 2001 to 2006. A trend line (k/2 data points on each side of x are used in a linear regression to predict the value at x2) highlights the increase in mean concentration for the colder months.

Figure 5. Hypothetical Source Location Month Means with a Trend Line (μg/m³).



A bootstrap method was employed to evaluate the uncertainty associated with various combinations of sampling frequency, completeness, measurement bias and precision, on the estimate of monthly and rolling quarterly design values. The following bootstrap algorithm was constructed.

1. Generate 2500 bootstrap samples for each month over a three year period where the sample size is a function of the number of days in the month. Constructing monthly bootstrap samples allows for seasonal trends.
2. Select a random sample from each of the 2500 x 36 bootstrap samples where the sample size is a function of one of the six combinations of sampling frequency (daily, 1 in 3, 1 in 6) and completeness (90% and 75%).
3. Adjust *each* observation in the 2500 x 36 random samples for a particular combination of precision and bias, as follows.
    a. For precision, assume a normal random variable with a mean of zero and standard deviation of the Pb concentration times a decrease in precision of either 10%, 20%, or 30%.
    b. For bias, assume a fixed amount per measurement, add or subtract the Pb concentration times an increase in bias of either 5%, 10%, or 15%.
4. Estimate the design value:
    a. For a monthly averaging time calculate the 2500 x 36 monthly averages, $M_{i,j,k}$,

where i = 1, 2, 3, years, j = 1, 2, ..., 12, months, and k = 1, 2, ..., 2500 bootstrap samples;

   b. For a rolling quarterly averaging time calculate the 2500 x 36 rolling quarterly averages as [$Q_{1,1,1} = (M_{3,11,1} + M_{3,12,1} + M_{1,1,1}) / 3$].

5. Calculate the compliance rates for various action levels where the compliance rule is:
   a. For monthly averaging time – no more than one monthly average within a three year period greater than the action level (the design value is the second highest value within a three year period);
   b. For quarterly averaging time – no rolling quarterly average within a three year period greater than the action level (the design value is the highest value within a three year period).
6. Calculate the average of the descriptive statistics (mean, standard error, 95% confidence interval) for each three year period for 2500 bootstrap iterations.

The R-code used to generate the bootstrap samples is provided in Appendix D.

To provide a reference distribution, the averaging time distributions for a daily sampling schedule, 100% completeness, no change in measurement bias and precision, were constructed for both monthly and rolling quarterly averaging times. The descriptive statistics for these bootstrap distributions are provided in Table 3 for the monthly design value and Table 4 for the rolling quarterly design value. Box plots of the two distributions are provided in Figure 6.

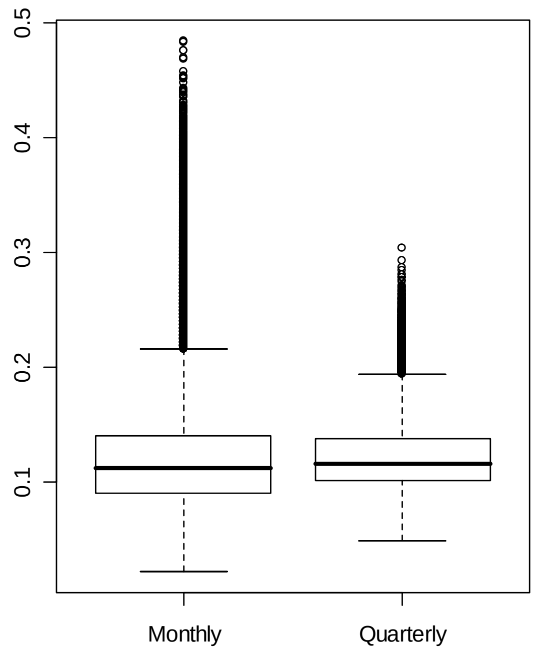
Table 3. Average Monthly Design Value Statistics for the Hypothetical Monitoring Site ($\mu g/m^3$).
(Daily Sampling, 100% Completeness, No Decrease in Precision or Increase in Bias)

N	**Min.**	**Q(.025)**	**Q(.25)**	**Med.**	**Mean**	**Q(.75)**	**Q(.975)**	**Max.**	**SE**
2500	0.0563	0.0563	0.0890	0.1119	0.1221	0.1385	0.2329	0.2810	0.0487

Table 4. Average Rolling Quarterly Design Value Statistics for the Hypothetical Monitoring Site ($\mu g/m^3$).
(Daily Sampling, 100% Completeness, No Decrease in Precision or Increase in Bias)

N	**Min.**	**Q(.025)**	**Q(.25)**	**Med.**	**Mean**	**Q(.75)**	**Q(.975)**	**Max.**	**SE**
2500	0.0776	0.0776	0.0997	0.1163	0.1221	0.1375	0.1843	0.1937	0.0297

Figure 6. Box plots of Monthly and Quarterly Design Value Distributions (µg/m^3).
(Daily Sampling, 100% Completeness, No Decrease in Precision or Increase in Bias)



Data Analysis

The data in Tables 9 to 20 were summarized using an analysis of variance. The model used was a completely randomized design with a factorial treatment structure (averaging with two levels, sampling frequency with three levels, completeness with two levels, precision with three levels and bias with six levels) where the derived response variable was the width of the scenario confidence interval (95% confidence interval for the mean). The analysis of variance table is provide in Table 6 for all statistically significant (p-value < 0.05) main and two-way interaction effects.

Table 6. Analysis of Variance Table for the Derived Response Variable the Ratio of the Width of the Scenario Confidence Interval to the Width of the Reference Confidence Interval

Source of Variation	**DF**	**Sum of Squares**	**Mean Squares**	**F-value**	**Prob ≥ (F-value)**
Average	1	0.4673	0.4673	8885.953	0.00000
Sampling	2	0.7677	0.3838	7297.906	0.00000
Completeness	1	0.0155	0.0155	295.369	0.00000
Precision	3	0.0014	0.0005	9.115	0.00001
Bias	5	0.1242	0.0248	472.231	0.00000
Average x Sampling	2	0.0083	0.0042	79.247	0.00000
Average x Bias	6	0.0047	0.0008	15.019	0.00000
Sampling x Completeness	2	0.0017	0.0009	16.398	0.00000
Sampling x Bias	12	0.0087	0.0007	13.816	0.00000
Experimental Error	193	0.0101	0.0001		

The main effects are displayed in Figure 7 using box plots.

All independent variables, design value averaging time, sampling frequency, completeness, precision and bias have a statistically significant impact on the width of the confidence interval for the mean (p-value < 0.0000). The design value averaging time and sampling frequency have the greatest impact on the width (design value averaging time: monthly = 0.2766, rolling quarterly = 0.1860; sampling frequency: 1:6 = 0.2980, 1:3 = 0.2392, daily = 0.1566 (μg/m^3)) but the change in sampling frequency impacts the two design values differently. The decrease in the mean width with the increase in sampling frequency is greater for the monthly average than the rolling quarterly average (Table 7).

Table 7. Design Value Averaging Time by Sampling Frequency Interaction Mean for the Width of the 95% Confidence Interval for the Mean (μg/m^3).

Design Value	**1:6 Sampling**	**1:3 Sampling**	**Daily Sampling**
Monthly	0.3496	0.2864	0.1937
Rolling Quarterly	0.2465	0.1920	0.1194

The change in design value averaging time also interacts with the change in bias. The decrease in the mean width with the decrease in bias is greater for the monthly averaging design value than the rolling quarterly average (Table 8).

Table 8. Design Value Averaging Time by Bias Interaction Means for the Width of the 95% Confidence Interval for the Mean (μg/m^3).

Design Value	**-15% Bias**	**-10% Bias**	**-5% Bias**	**0% Bias**	**+5% Bias**	**+10% Bias**	**+15% Bias**
Monthly	0.2361	0.2487	0.2683	0.2708	0.2892	0.3032	0.3158
Rolling Quarterly	0.1596	0.1679	0.1768	0.1827	0.1953	0.2046	0.2129

The change in data completeness has less impact on mean width than the change in design value averaging time, sampling frequency and bias (completeness: 75% = 0.2395, 90% = 0.2230 (μg/m^3)); and the change in precision has the least influence on mean width (10% = 0.2293, 20% = 0.2305, 30% = 0.2347 (μg/m^3)).

Figure 7. Box plots of the Main Effects for the
Width of the 95% Confidence Interval for the Mean (µg/m^3)
(Mean = black filled circle, Median = white horizontal line).

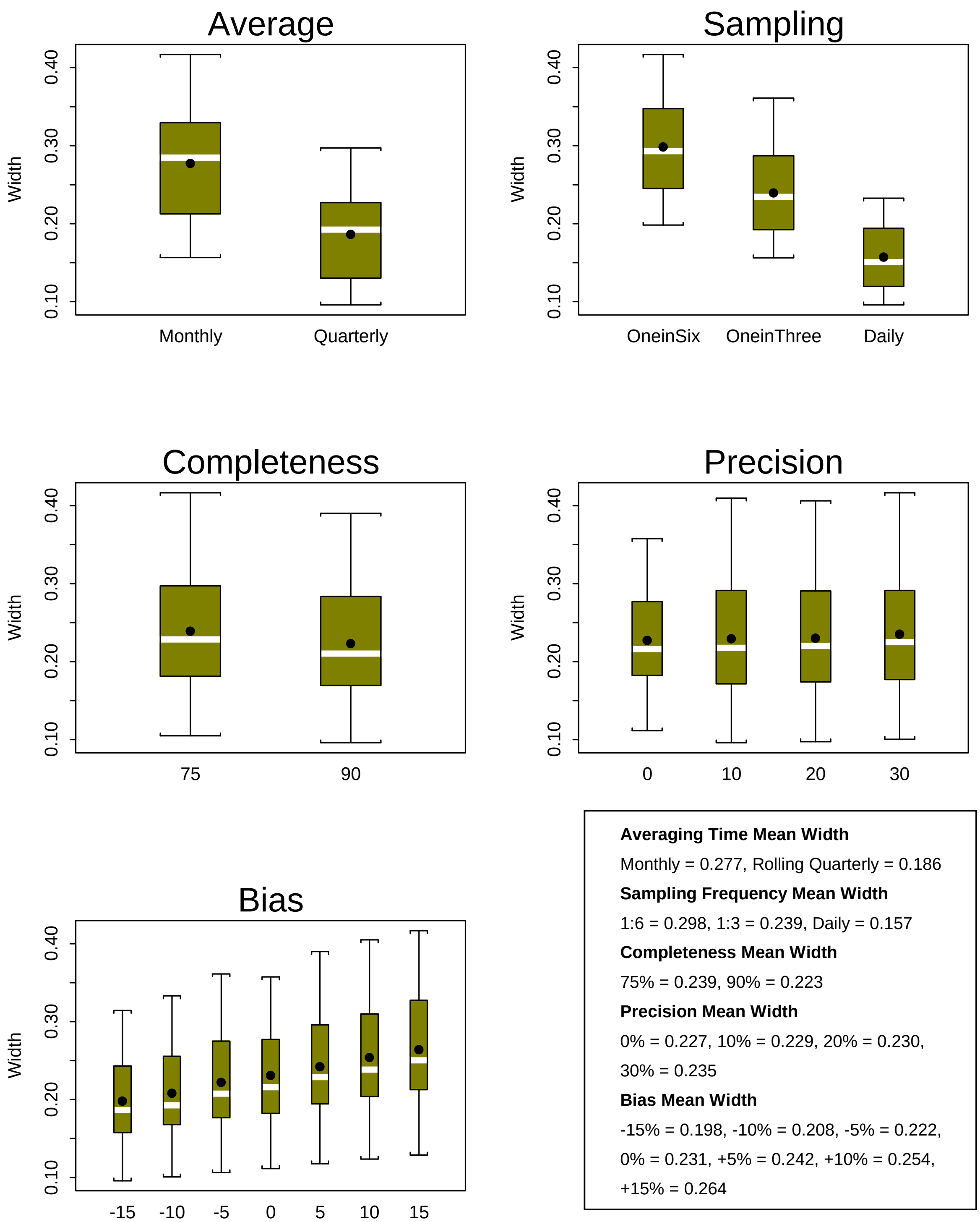


Table 9. Bootstrap Precision Estimates for Monthly Design Values over a Three Year Period
Source Locations with Daily Sampling and 90% Completeness

Monthly		**Measurement**		**Design Value Statistics for a Three Year Period**				**Percent in Compliance[1] For Action Levels**		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m³**	**0.2 µg/m³**	**0.3 µg/m³**
Daily	90%	0	0	0.1222	0.0513	0.0541	0.2391	0	22.60	91.92
		10%	+5%	0.1282	0.0540	0.0568	0.2509	0	14.60	85.24
		10%	+10%	0.1345	0.0567	0.0593	0.2632	0	9.24	76.16
		10%	+15%	0.1404	0.0590	0.0622	0.2748	0	5.44	68.92
		20%	+5%	0.1281	0.0546	0.0562	0.2512	0	14.88	83.16
		20%	+10%	0.1344	0.0574	0.0586	0.2630	0	9.28	76.52
		20%	+15%	0.1405	00597	0.0617	0.2751	0	5.36	68.52
		30%	+5%	0.1282	0.0558	0.0553	0.2513	0	14.16	83.52
		30%	+10%	0.1344	0.0585	0.0583	0.2641	0	8.68	76.52
		30%	+15%	0.1405	0.0609	0.0606	0.2761	0	4.92	69.56
	90%	10%	-5%	0.1160	0.0488	0.0514	0.2266	0	31.48	85.44
		10%	-10%	0.1100	0.0463	0.0486	0.2146	0	40.68	97.64
		10%	-15%	0.1037	0.0438	0.0457	0.2023	0	49.32	99.28
		20%	-5%	0.1160	0.0497	0.0508	0.2272	0	31.60	93.32
		20%	-10%	0.1100	0.0474	0.0480	0.2152	0	41.36	96.36
		20%	-15%	0.1038	0.0446	0.0452	0.2034	0	51.20	98.20
		30%	-5%	0.1161	0.0512	0.0496	0.2299	0	31.60	90.28
		30%	-10%	0.1099	0.0485	0.0466	0.2169	0	42.60	94.20
		30%	-15%	0.1038	0.0462	0.0440	0.2053	0	52.80	97.08

[1]Compliance for monthly design value: no more than one monthly average within a three year period greater than the action level

Table 10. Bootstrap Precision Estimates for Monthly Design Values over a Three Year Period
Source Locations with Daily Sampling and 75% Completeness

Monthly		**Measurement**		**Design Value Statistics for a Three Year Period**				**Percent in Compliance[1] For Action Levels**		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m^3**	**0.2 µg/m^3**	**0.3 µg/m^3**
Daily	75%	0	0	0.1223	0.0564	0.0499	0.2502	0	21.44	78.68
		10%	+5%	0.1281	0.0591	0.0521	0.2602	0	15.32	74.16
		10%	+10%	0.1343	0.0617	0.0550	0.2745	0	8.48	66.72
		10%	+15%	0.1405	0.0647	0.0575	0.2861	0	4.24	61.32
		20%	+5%	0.1283	0.0601	0.0516	0.2616	0	13.44	75.08
		20%	+10%	0.1343	0.0630	0.0541	0.2744	0	7.04	68.84
		20%	+15%	0.1404	0.0657	0.0569	0.2866	0	3.52	63.04
		30%	+5%	0.1284	0.0615	0.0508	0.2633	0	12.88	75.68
		30%	+10%	0.1345	0.0646	0.0533	0.2761	0	7.12	68.68
		30%	+15%	0.1404	0.0667	0.0559	0.2885	0	3.56	63.56
	75%	10%	-5%	0.1160	0.0537	0.0472	0.2377	0	29.16	85.64
		10%	-10%	0.1100	0.0510	0.0446	0.2251	0	37.24	91.68
		10%	-15%	0.1039	0.0483	0.0422	0.2123	0	48.48	95.72
		20%	-5%	0.1160	0.0548	0.0464	0.2380	0	28.56	84.84
		20%	-10%	0.1099	0.0519	0.0441	0.2254	0	38.36	90.36
		20%	-15%	0.1039	0.0492	0.0416	0.2133	0	46.52	94.32
		30%	-5%	0.1160	0.0560	0.0454	0.2395	0	27.04	85.60
		30%	-10%	0.1098	0.0543	0.0431	0.2273	0	36.48	89.24
		30%	-15%	0.1038	0.0509	0.0403	0.2151	0	47.36	92.36

[1]Compliance for monthly design value: no more than one monthly average within a three year period greater than the action level

Table 11. Bootstrap Precision Estimates for Monthly Design Values over a Three Year Period
Source Locations with 1 in 3 Sampling and 90% Completeness

Monthly		Measurement		Design Value Statistics for a Three Year Period				Percent in Compliance[1] For Action Levels		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m^3**	**0.2 µg/m^3**	**0.3 µg/m^3**
1 in 3	90%	0	0	0.1220	0.0854	0.0286	0.2974	0	6.24	70.36
		10%	+5%	0.1287	0.0910	0.0300	0.3196	0	3.04	59.24
		10%	+10%	0.1343	0.0947	0.0314	0.3284	0	1.60	54.12
		10%	+15%	0.1407	0.0997	0.0324	0.3501	0	0.80	43.92
		20%	+5%	0.1287	0.0917	0.0291	0.3175	0	3.52	58.88
		20%	+10%	0.1340	0.0959	0.0306	0.3329	0	1.96	51.76
		20%	+15%	0.1404	0.0998	0.0328	0.3458	0	0.56	44.64
		30%	+5%	0.1279	0.0931	0.0287	0.3177	0	3.04	56.76
		30%	+10%	0.1343	0.0973	0.0301	0.3332	0	1.80	48.08
		30%	+15%	0.1411	0.1036	0.0315	0.3542	0	0.92	38.69
	90%	10%	-5%	0.1163	0.0823	0.0270	0.2869	0	9.72	72.76
		10%	-10%	0.1098	0.0774	0.0257	0.2683	0	16.40	79.84
		10%	-15%	0.1042	0.0744	0.0241	0.2592	0	24.40	81.72
		20%	-5%	0.1166	0.0871	0.0268	0.2903	0	9.84	70.24
		20%	-10%	0.1101	0.0804	0.0249	0.2741	0	15.64	76.12
		20%	-15%	0.1038	0.0751	0.0235	0.2586	0	23.12	81.04
		30%	-5%	0.1163	0.0865	0.0253	0.2920	0	9.12	68.92
		30%	-10%	0.1098	0.0816	0.0240	0.2760	0	13.40	74.92
		30%	-15%	0.1093	0.0779	0.0222	0.2632	0	22.40	79.00

[1]Compliance for monthly design value: no more than one monthly average within a three year period greater than the action level

Table 12. Bootstrap Precision Estimates for Monthly Design Values over a Three Year Period
Source Locations with 1 in 3 Sampling and 75% Completeness

Monthly		Measurement		Design Value Statistics for a Three Year Period				Percent in Compliance[1] For Action Levels		
Sampling	Completeness	Precision	Bias	Average Mean	Average Standard Error	Average 2.5% Quantile	Average 97.5% Quantile	0.1 µg/m³	0.2 µg/m³	0.3 µg/m³
1 in 3	75%	0	0	0.1221	0.0934	0.0245	0.3100	0	4.40	62.48
		10%	+5%	0.1282	0.0988	0.0255	0.3293	0	2.64	54.68
		10%	+10%	0.1345	0.1024	0.0270	0.3435	0	1.56	45.64
		10%	+15%	0.1407	0.1075	0.0280	0.3602	0	0.48	36.72
		20%	+5%	0.1281	0.0989	0.0252	0.3273	0	1.36	53.36
		20%	+10%	0.1348	0.1050	0.0264	0.3477	0	1.52	43.40
		20%	+15%	0.1406	0.1081	0.0276	0.3608	0	0.68	34.84
		30%	+5%	0.1285	0.1015	0.0243	0.3341	0	2.72	48.24
		30%	+10%	0.1347	0.1065	0.0254	0.3516	0	1.48	41.24
		30%	+15%	0.1409	0.1111	0.0267	0.3684	0	0.36	31.48
	75%	10%	-5%	0.1162	0.0888	0.0228	0.2977	0	6.84	69.96
		10%	-10%	0.1101	0.0842	0.0217	0.2809	0	13.32	74.72
		10%	-15%	0.1041	0.0805	0.0206	0.2683	0	19.20	78.72
		20%	-5%	0.1158	0.0894	0.0225	0.2977	0	7.44	67.08
		20%	-10%	0.1102	0.0857	0.0214	0.2855	0	12.30	73.24
		20%	-15%	0.1039	0.0811	0.0199	0.2651	0	18.88	79.04
		30%	-5%	0.1161	0.0926	0.0215	0.3052	0	6.80	63.20
		30%	-10%	0.1098	0.0875	0.0202	0.2864	0	11.32	70.44
		30%	-15%	0.1040	0.0838	0.0191	0.02743	0	15.92	75.32

[1]Compliance for monthly design value: no more than one monthly average within a three year period greater than the action level

Table 13. Bootstrap Precision Estimates for Monthly Design Values over a Three Year Period
Source Locations with 1 in 6 and 90% Completeness

Monthly		Measurement		Design Value Statistics for a Three Year Period				Percent in Compliance[1] For Action Levels		
Sampling	Completeness	Precision	Bias	Average Mean	Average Standard Error	Average 2.5% Quantile	Average 97.5% Quantile	0.1 µg/m^3	0.2 µg/m^3	0.3 µg/m^3
1 in 6	90%	0	0	0.1217	0.1121	0.0153	0.3430	0	3.28	39.88
		10%	+5%	0.1285	0.1183	0.0163	0.3633	0	1.40	30.68
		10%	+10%	0.1338	0.1226	0.0167	0.3772	0	0.80	25.84
		10%	+15%	0.1407	0.1296	0.0173	0.4040	0	0.68	19.04
		20%	+5%	0.1279	0.1188	0.0155	0.3696	0	1.56	31.48
		20%	+10%	0.1344	0.1250	0.0165	0.3897	0	1.00	25.72
		20%	+15%	0.1399	0.1293	0.0171	0.3990	0	0.64	21.04
		30%	+5%	0.1283	0.1214	0.0153	0.3763	0	1.32	31.44
		30%	+10%	0.1346	0.1271	0.0159	0.3938	0	0.80	25.52
		30%	+15%	0.1398	0.1322	0.0167	0.4069	0	0.80	20.72
	90%	10%	-5%	0.1154	0.1062	0.0143	0.3257	0	4.48	47.88
		10%	-10%	0.1093	0.1000	0.0137	0.3110	0	7.60	58.08
		10%	-15%	0.1039	0.0960	0.0130	0.2967	0	11.56	63.60
		20%	-5%	0.1162	0.1080	0.0142	0.3365	0	4.20	44.72
		20%	-10%	0.1093	0.1022	0.0134	0.3134	0	8.16	56.64
		20%	-15%	0.1040	0.0986	0.0123	0.3047	0	10.36	61.76
		30%	-5%	0.1161	0.1109	0.0136	0.3440	0	5.40	44.48
		30%	-10%	0.1098	0.1063	0.0129	0.3241	0	6.92	50.04
		30%	-15%	0.1039	0.0995	0.0117	0.3046	0	9.76	58.28

[1]Compliance for monthly design value: no more than one monthly average within a three year period greater than the action level

Table 14. Bootstrap Precision Estimates for Monthly Design Values over a Three Year Period
Source Locations with 1 in 6 and 75% Completeness

Monthly		Measurement		Design Value Statistics for a Three Year Period				Percent in Compliance[1] For Action Levels		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m³**	**0.2 µg/m³**	**0.3 µg/m³**
1 in 6	75%	0	0	0.1224	0.1240	0.0117	0.3692	0	3.04	30.44
		10%	+5%	0.1281	0.1290	0.0123	0.3869	0	1.40	27.32
		10%	+10%	0.1345	0.1363	0.0128	0.4079	0	0.88	20.48
		10%	+15%	0.1403	0.1422	0.0135	0.4233	0	0.44	17.60
		20%	+5%	0.1285	0.1320	0.0119	0.3932	0	1.40	26.00
		20%	+10%	0.1357	0.1408	0.0127	0.4176	0	1.00	31.32
		20%	+15%	0.1398	0.1421	0.0133	0.4196	0	0.04	19.12
		30%	+5%	0.1287	0.1353	0.0115	0.4015	0	1.24	26.12
		30%	+10%	0.1347	0.1407	0.0122	0.4154	0	1.08	21.16
		30%	+15%	0.1403	0.1449	0.0129	0.4296	0	0.48	19.92
	75%	10%	-5%	0.1160	0.1173	0.0111	0.3500	0	3.84	38.40
		10%	-10%	0.1101	0.1120	0.0105	0.3328	0	6.12	45.80
		10%	-15%	0.1037	0.1050	0.0098	0.3143	0	9.88	55.00
		20%	-5%	0.1163	0.1194	0.0109	0.3559	0	4.00	39.04
		20%	-10%	0.1096	0.1117	0.0103	0.3395	0	6.28	46.94
		20%	-15%	0.1037	0.1074	0.0094	0.3189	0	9.76	52.84
		30%	-5%	0.1156	0.1199	0.0103	0.3578	0	4.24	38.52
		30%	-10%	0.1107	0.1173	0.0096	0.3427	0	6.08	43.48
		30%	-15%	0.1039	0.1106	0.0087	0.3230	0	8.86	51.60

[1]Compliance for monthly design value: no more than one monthly average within a three year period greater than the action level

Table 15. Bootstrap Precision Estimates for Rolling Quarterly Design Values over a Three Year Period Source Locations with Daily Sampling and 90% Completeness.

Quarterly		**Measurement**		**Design Value Statistics for a Three Year Period**				**Percent in Compliance[2] For Action Levels**		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m³**	**0.2 µg/m³**	**0.3 µg/m³**
Daily	90%	0	0	0.1222	0.0310	0.0761	0.1876	0	56.44	99.96
		10%	+5%	0.1282	0.0326	0.0797	0.1976	0	40.68	99.72
		10%	+10%	0.1345	0.0342	0.0836	0.2072	0	28.61	99.28
		10%	+15%	0.1404	0.0357	0.0873	0.2164	0	17.96	98.00
		20%	+5%	0.1281	0.0330	0.0794	0.1988	0	39.64	99.60
		20%	+10%	0.1344	0.0347	0.0830	0.2087	0	28.52	98.72
		20%	+15%	0.1405	0.0360	0.0870	0.2171	0	19.64	97.84
		30%	+5%	0.1282	0.0335	0.0786	0.2003	0	37.84	98.96
		30%	+10%	0.1344	0.0352	0.0824	0.2103	0	27.28	97.96
		30%	+15%	0.1405	0.0367	0.0865	0.2196	0	19.00	96.04
	90%	10%	-5%	0.1160	0.0295	0.0723	0.1787	0	70.24	99.96
		10%	-10%	0.1100	0.0280	0.0685	0.1693	0	82.36	100.00
		10%	-15%	0.1037	0.0265	0.0644	0.1603	0	90.60	100.00
		20%	-5%	0.1160	0.0300	0.0717	0.1801	0	67.44	100.00
		20%	-10%	0.1100	0.0285	0.0680	0.1715	0	78.12	100.00
		20%	-15%	0.1038	0.0269	0.0640	0.1613	0	88.40	100.00
		30%	-5%	0.1161	0.0307	0.0708	0.1827	0	64.16	99.84
		30%	-10%	0.1099	0.0290	0.0671	0.1727	0	75.40	99.96
		30%	-15%	0.1038	0.0277	0.0631	0.1636	0	84.12	100.00

[2]Compliance for quarterly design values: no rolling quarterly averages within a three year period greater than the action level

Table 16. Bootstrap Precision Estimates for Rolling Quarterly Design Values over a Three Year Period
Source Locations with Daily Sampling and 75% Completeness.

Quarterly		Measurement		Design Value Statistics for a Three Year Period				Percent in Compliance[2] For Action Levels		
Sampling	Completeness	Precision	Bias	Average Mean	Average Standard Error	Average 2.5% Quantile	Average 97.5% Quantile	0.1 $\mu g/m^3$	0.2 $\mu g/m^3$	0.3 $\mu g/m^3$
Daily	75%	0	0	0.1223	0.0336	0.0734	0.1951	0	43.64	99.48
		10%	+5%	0.1281	0.0352	0.0766	0.2050	0	31.84	98.72
		10%	+10%	0.1343	0.0367	0.0807	0.2141	0	23.84	97.16
		10%	+15%	0.1405	0.0386	0.0841	0.2243	0	16.68	94.40
		20%	+5%	0.1283	0.0358	0.0761	0.2067	0	31.40	97.76
		20%	+10%	0.1343	0.0375	0.0797	0.2162	0	23.48	96.20
		20%	+15%	0.1404	0.0391	0.0836	0.2258	0	16.52	94.20
		30%	+5%	0.1284	0.0365	0.0756	0.2087	0	31.40	96.96
		30%	+10%	0.1345	0.0384	0.0790	0.2195	0	23.44	94.40
		30%	+15%	0.1404	0.0395	0.0828	0.2274	0	17.56	91.60
	75%	10%	-5%	0.1160	0.0321	0.0693	0.1860	0	55.80	99.68
		10%	-10%	0.1100	0.0303	0.0658	0.1762	0	70.16	99.96
		10%	-15%	0.1039	0.0288	0.0620	0.1669	0	81.84	10.00
		20%	-5%	0.1160	0.0326	0.0686	0.1876	0	54.08	99.36
		20%	-10%	0.1099	0.0309	0.0650	0.1775	0	67.72	99.72
		20%	-15%	0.1039	0.0292	0.0613	0.1679	0	78.60	100.00
		30%	-5%	0.1160	0.0332	0.0680	0.1893	0	51.20	99.12
		30%	-10%	0.1098	0.0317	0.0639	0.1800	0	63.16	99.72
		30%	-15%	0.1038	0.0301	0.0604	0.1707	0	74.24	99.88

[2]Compliance for quarterly design values: no rolling quarterly averages within a three year period greater than the action level

Table 17. Bootstrap Precision Estimates for Rolling Quarterly Design Values over a Three Year Period Source Locations with 1 in 3 Sampling and 90% Completeness.

Quarterly		Measurement		Design Value Statistics for a Three Year Period				Percent in Compliance[2] For Action Levels		
Sampling	Completeness	Precision	Bias	Average Mean	Average Standard Error	Average 2.5% Quantile	Average 97.5% Quantile	0.1 µg/m³	0.2 µg/m³	0.3 µg/m³
1 in 3	90%	0	0	0.1220	0.0488	0.0568	0.2367	0	30.92	70.20
		10%	+5%	0.1287	0.0521	0.0595	0.2514	0	22.36	64.48
		10%	+10%	0.1343	0.0541	0.0619	0.2611	0	17.72	60.36
		10%	+15%	0.1407	0.0571	0.0643	0.2746	0	11.28	55.32
		20%	+5%	0.1281	0.0525	0.0587	0.2526	0	21.00	65.92
		20%	+10%	0.1340	0.0549	0.0611	0.2635	0	15.88	62.44
		20%	+15%	0.1404	0.0571	0.0643	0.2742	0	11.76	58.44
		30%	+5%	0.1279	0.0532	0.0576	0.2539	0	22.00	67.20
		30%	+10%	0.1343	0.0555	0.0604	0.2646	0	16.76	62.52
		30%	+15%	0.1411	0.0591	0.0631	0.2806	0	10.52	57.20
	90%	10%	-5%	0.1163	0.0470	0.0533	0.2263	0	36.68	77.48
		10%	-10%	0.1098	0.0444	0.0507	0.2142	0	42.72	83.52
		10%	-15%	0.1042	0.0425	0.0479	0.2040	0	47.72	55.96
		20%	-5%	0.1166	0.0481	0.0530	0.2306	0	3412	75.40
		20%	-10%	0.1101	0.0459	0.0498	0.2192	0.04	40.60	80.00
		20%	-15%	0.1038	0.0429	0.0469	0.2044	0	48.40	85.96
		30%	-5%	0.1163	0.0493	0.0543	0.2333	0	32.28	74.24
		30%	-10%	0.1098	0.0465	0.0486	0.2199	0	42.28	78.72
		30%	-15%	0.1039	0.0443	0.0460	0.2089	0	49.20	83.12

[2]Compliance for quarterly design values: no rolling quarterly averages within a three year period greater than the action level

Table 18. Bootstrap Precision Estimates for Rolling Quarterly Design Values over a Three Year Period Source Locations with 1 in 3 Sampling and 75% Completeness.

Quarterly		**Measurement**		**Design Value Statistics for a Three Year Period**				**Percent in Compliance[2] For Action Levels**		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 μg/m^3**	**0.2 μg/m^3**	**0.3 μg/m^3**
1 in 3	75%	0	0	0.1221	0.0532	0.0528	0.2492	0	27.68	65.00
		10%	+5%	0.1282	0.0562	0.0549	0.2622	0	20.92	61.60
		10%	+10%	0.1345	0.0584	0.0581	0.2731	0	13.92	59.44
		10%	+15%	0.1407	0.0613	0.0603	0.2850	0	10.44	56.44
		20%	+5%	0.1281	0.0565	0.0545	0.2630	0	20.24	63.80
		20%	+10%	0.1348	0.0595	0.0574	0.2763	0	14.24	60.08
		20%	+15%	0.1406	0.0615	0.0601	0.2862	0	11.16	56.24
		30%	+5%	0.1285	0.0578	0.0536	0.2663	0	18.52	64.40
		30%	+10%	0.1347	0.0605	0.0566	0.2785	0	12.92	60.72
		30%	+15%	0.1409	0.0633	0.0589	0.2920	0	9.92	55.12
	75%	10%	-5%	0.1162	0.0507	0.0496	0.2362	0	34.36	69.92
		10%	-10%	0.1101	0.0479	0.0475	0.2240	0	40.44	75.84
		10%	-15%	0.1041	0.0459	0.0447	0.2134	0	46.80	79.68
		20%	-5%	0.1158	0.0509	0.0492	0.2368	0	32.88	70.76
		20%	-10%	0.1102	0.0487	0.0468	0.2260	0	40.76	75.48
		20%	-15%	0.1039	0.0460	0.0440	0.2137	0	46.72	80.12
		30%	-5%	0.1161	0.0525	0.0481	0.2410	0	30.96	71.72
		30%	-10%	0.1098	0.0500	0.0452	0.2289	0	39.52	74.92
		30%	-15%	0.1040	0.0476	0.0427	0.2176	0	45.64	79.32

[2]Compliance for quarterly design values: no rolling quarterly averages within a three year period greater than the action level

Table 19. Bootstrap Precision Estimates for Rolling Quarterly Design Values over a Three Year Period
Source Locations with 1 in 6 Sampling and 90% Completeness.

Quarterly		**Measurement**		**Design Value Statistics for a Three Year Period**				**Percent in Compliance[2] For Action Levels**		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m³**	**0.2 µg/m³**	**0.3 µg/m³**
1 in 6	90%	0	0	0.1217	0.0632	0.0430	0.2747	0	20.76	64.56
		10%	+5%	0.1285	0.0670	0.0448	0.2904	0	13.40	60.88
		10%	+10%	0.1338	0.0693	0.0473	0.3017	0	10.12	59.24
		10%	+15%	0.1407	0.0734	0.0496	0.3187	0	6.76	53.16
		20%	+5%	0.1279	0.0669	0.0444	0.2893	0	13.80	61.92
		20%	+10%	0.1344	0.0705	0.0470	0.3043	0	10.64	57.84
		20%	+15%	0.1399	0.0731	0.0488	0.3163	0	6.76	53.36
		30%	+5%	0.1283	0.0686	0.0435	0.2942	0	13.60	60.60
		30%	+10%	0.1346	0.0718	0.0461	0.3077	0	9.68	56.04
		30%	+15%	0.1398	0.0744	0.0477	0.3197	0	8.36	51.92
	90%	10%	-5%	0.1154	0.0602	0.0407	0.2623	0	26.84	67.20
		10%	-10%	0.1093	0.0563	0.0386	0.2400	0	34.40	71.48
		10%	-15%	0.1039	0.0542	0.0366	0.2347	0.16	41.08	71.20
		20%	-5%	0.1162	0.0611	0.0405	0.2632	0	23.56	67.76
		20%	-10%	0.1093	0.0575	0.0379	0.2482	0.04	32.28	71.44
		20%	-15%	0.1040	0.0556	0.0358	0.2383	0.04	39.08	72.44
		30%	-5%	0.1161	0.0625	0.0392	0.2665	0	24.72	67.40
		30%	-10%	0.1098	0.0599	0.0367	0.2556	0	29.29	70.52
		30%	-15%	0.1039	0.0562	0.0347	0.2393	0.04	35.24	74.80

[2]Compliance for quarterly design values: no rolling quarterly averages within a three year period greater than the action level

Table 20. Bootstrap Precision Estimates for Rolling Quarterly Design Values over a Three Year Period
Source Locations with 1 in 6 Sampling and 75% Completeness.

Quarterly		**Measurement**		**Design Value Statistics for a Three Year Period**				**Percent in Compliance[2] For Action Levels**		
Sampling	**Completeness**	**Precision**	**Bias**	**Average Mean**	**Average Standard Error**	**Average 2.5% Quantile**	**Average 97.5% Quantile**	**0.1 µg/m³**	**0.2 µg/m³**	**0.3 µg/m³**
1 in 6	75%	0	0	0.1224	0.0700	0.0383	0.2933	0	15.64	63.24
		10%	+5%	0.1281	0.0726	0.0403	0.3044	0	11.40	59.12
		10%	+10%	0.1345	0.0770	0.0420	0.3231	0	8.48	55.36
		10%	+15%	0.1403	0.0804	0.0435	0.3367	0	5.40	49.84
		20%	+5%	0.1285	0.0748	0.0395	0.3121	0	10.92	57.48
		20%	+10%	0.1357	0.0793	0.0416	0.3306	0	8.40	50.88
		20%	+15%	0.1398	0.0803	0.0433	0.3362	0	5.96	48.80
		30%	+5%	0.1287	0.0759	0.0383	0.3147	0	11.12	56.16
		30%	+10%	0.1347	0.0793	0.0408	0.3302	0	8.04	50.64
		30%	+15%	0.1403	0.0815	0.0425	0.3396	0	6.68	46.52
	75%	10%	-5%	0.1160	0.0662	0.0361	0.2770	0.04	21.48	68.04
		10%	-10%	0.1101	0.0632	0.0345	0.2643	0	28.52	69.88
		10%	-15%	0.1037	0.0593	0.0326	0.2487	0.04	35.88	72.76
		20%	-5%	0.1163	0.0674	0.0353	0.2804	0	20.20	67.00
		20%	-10%	0.1096	0.0629	0.0335	0.2619	0	28.32	70.64
		20%	-15%	0.1037	0.0604	0.0315	0.2516	0	33.32	72.60
		30%	-5%	0.1156	0.0673	0.0347	0.2794	0	20.92	66.16
		30%	-10%	0.1107	0.0660	0.0328	0.2731	0	24.56	68.36
		30%	-15%	0.1039	0.0623	0.0303	0.2573	0	33.00	72.48

[2]Compliance for quarterly design values: no rolling quarterly averages within a three year period greater than the action level

Appendix A: Construct of the data set.

How qualified measurements or outlying values were handled:

- Location 06025-0005 - qualified observations were retained since they lie within the main body of the data.
- Location 06073-1007 - the qualified observations follow the same trend as the non-qualified; even though 77 of the 150 observations are qualified, the qualified observations were retained.
- Location 09009-2123 – the location was removed since 50% of the observations are qualified and there are less than 40 measurements for the location.
- Location 26163-0019 - the qualified observation was retained since the qualifier is a "Validated Value."
- Location 29093-0016 - the four qualified observations appear to follow the same trend as the non-qualified observations and since the qualifier is a "Validated Value" the qualified observations were retained.
- Location 29093-0021 - the two qualified observations appear to follow the same trend as the non-qualified observations and since the qualifier is a "Validated Value" the qualified observations were retained.
- Location 29093-0024 - the fourteen qualified observations appear to follow the same trend as the non-qualified observations and since the qualifier is a "Validated Value" the qualified observations were retained.
- Location 29099-0005 - all but one of the qualified observations appear to follow the same trend as the non-qualified, these qualified observations have the qualifier of "Validated Value" and were retained. The one observation outside the main body of the data was taken in 2001 and has the highest concentration in the entire data set (39.8 micrograms/cubic meter, the next highest is 15.9 micrograms/cubic meter). This observation was removed.
- Location 29099-0013 - the three qualified observations appear to follow the same trend as the non-qualified observations and since the qualifier is a "Validated Value" the qualified observations were retained.
- Location 4811-3057 - the 54 highest concentrations for 2001 are all from location 4811-3057, these observations are outside the main body of the data and were removed.

Locations with less than 40 measurements for at least one calendar year.
The following locations were deleted from the final data set:

- 08001-0001 measurements in only four months
- 09009-2123 only 24 measurements in a calendar year
- 11001-0027 measurement in only eight months
- 13089-0003
- 08031-0024 measurements in only five months
- 29099-0021 measurements in only three months

Measurement without a duration code of "7" (code for 24 hour measurement).
The following quarterly and monthly measurements were deleted:

- 86 monthly observations from 01109-0003 from the years 2001 to 2005;
- eight quarterly observations from 47093-0027 from the years 2001 and 2002; and
- 12 observations from 2616-3001 from the year 2001.

Special Studies

All four locations with the project label of Special Studies, 29093-0016, 29093-0021, 29099-0004, and 29099-0013, were removed.

Appendix B: Descriptive statistics for the AQS source and nonsource locations.

Table B-1. Descriptive Statistics for Source Locations sorted by Standard Deviation.

ID[1]	N	Min	Q(25)	Median	Mean	Q(75)	Max	Std. Dev.	%CV	# of 0s	%0s
450452002	443	0.00000	0.00000	0.00000	0.00067	0.00000	0.02200	0.00291	437	418	94.36
160790006	88	0.03000	0.03000	0.03000	0.03091	0.03000	0.06000	0.00419	14	0	0.00
450190003	441	0.00000	0.00000	0.00000	0.00161	0.00000	0.04100	0.00453	281	381	86.39
270370421	193	0.00000	0.00000	0.00000	0.00430	0.01000	0.02000	0.00556	129	116	60.10
270530968	143	0.00000	0.00000	0.00000	0.00252	0.00000	0.04000	0.00563	223	113	79.02
421010047	55	0.00000	0.01000	0.01000	0.01436	0.02000	0.04000	0.00877	61	6	10.91
170313103	358	0.01000	0.01000	0.01000	0.01394	0.01000	0.11000	0.00937	67	0	0.00
270530967	187	0.00000	0.00000	0.00000	0.00695	0.01000	0.11000	0.01149	165	94	50.27
180970076	350	0.00100	0.00700	0.01200	0.01535	0.02000	0.09200	0.01254	82	0	0.00
421010449	324	0.00000	0.01200	0.02200	0.02416	0.03025	0.08100	0.01512	63	10	3.09
170316003	335	0.00000	0.02000	0.03000	0.03027	0.04000	0.10000	0.01566	52	1	0.30
170310022	350	0.00000	0.02000	0.02000	0.02617	0.03000	0.12000	0.01659	63	4	1.14
470930027	165	0.00000	0.01000	0.01000	0.01352	0.01000	0.14000	0.01749	129	8	4.85
481410033	229	0.01000	0.01000	0.03000	0.03192	0.05000	0.09000	0.02210	69	0	0.00
471570044	200	0.00000	0.01000	0.01000	0.01540	0.01000	0.24000	0.02241	146	3	1.50
291892003	359	0.00000	0.00000	0.00000	0.01916	0.05000	0.13000	0.02493	130	223	62.12
471570045	43	0.01000	0.01000	0.02000	0.03070	0.03000	0.17000	0.03165	103	0	0.00
180890023	354	0.00000	0.01100	0.02250	0.03325	0.04100	0.24000	0.03411	103	3	0.85
170310026	350	0.00000	0.02000	0.03000	0.04194	0.05000	0.29000	0.03512	84	1	0.29
180970063	683	0.00000	0.00800	0.01700	0.03008	0.03900	0.23600	0.03525	117	7	1.02
490351001	283	0.01700	0.01800	0.01900	0.03741	0.03150	0.22900	0.04185	112	0	0.00
481130018	344	0.00000	0.00500	0.00500	0.02217	0.03000	0.92000	0.05615	253	68	19.77
420110005	285	0.03000	0.03000	0.04000	0.06533	0.06000	0.58000	0.07336	112	0	0.00
270370020	217	0.00000	0.00000	0.00000	0.01074	0.01000	1.09000	0.07420	691	120	55.30
471633002	354	0.01000	0.02000	0.04000	0.05579	0.06000	1.19000	0.07851	141	0	0.00
360713001	316	0.00300	0.03000	0.03000	0.05511	0.05000	0.78000	0.08599	156	0	0.00
420250105	283	0.03000	0.04000	0.04000	0.08495	0.09000	1.30000	0.12105	142	0	0.00
270370001	222	0.00000	0.02000	0.03000	0.07919	0.09000	0.87000	0.12798	162	5	2.25
360713002	308	0.00200	0.03000	0.05000	0.09719	0.10175	1.03000	0.13050	134	0	0.00
340231003	256	0.00700	0.00800	0.01200	0.06627	0.04400	0.93500	0.13950	211	0	0.00
420110717	338	0.00000	0.04000	0.09000	0.14047	0.18750	1.26000	0.15300	109	1	0.30
480850007	**344**	**0.00300**[2]	**0.00900**	**0.03350**	**0.10273**	**0.13300**	**2.04000**	**0.17693**	**172**	**0**	**0.00**
471870104	**209**	**0.00000**	**0.01000**	**0.07000**	**0.14411**	**0.20000**	**1.25000**	**0.19983**	**139**	**14**	**6.70**
171190010	**346**	**0.01000**	**0.01000**	**0.03000**	**0.06347**	**0.06000**	**3.73000**	**0.22227**	**350**	**0**	**0.00**
080010005	**420**	**0.00000**	**0.01583**	**0.05000**	**0.11359**	**0.12880**	**5.15920**	**0.28996**	**255**	**6**	**1.43**
290930030	**411**	**0.00000**	**0.00000**	**0.09300**	**0.18238**	**0.22000**	**4.98000**	**0.34828**	**191**	**157**	**38.20**
011090003	54	0.02400	0.07925	0.12600	0.34056	0.45750	1.39000	0.38605	113	0	0.00
471870100	233	0.00000	0.04000	0.14000	0.37000	0.41000	8.58000	0.76024	205	10	4.29
290930024	404	0.00000	0.08000	0.44550	0.66525	0.86250	6.72000	0.88295	133	64	15.84
120571066	353	0.00000	0.10000	0.30000	0.62295	0.70000	5.60000	0.92887	149	39	11.05
290990005	661	0.00000	0.03400	0.16000	0.49005	0.51000	8.62000	0.96313	197	137	20.73

[1]The first two numbers of the ID identify the state, the next three identify the county, and the last four identify the site.

[2]Bolded indicate locations used to construct the hypothetical location.

Table B-2. Descriptive Statistics for Nonsource Locations sorted by Standard Deviation.

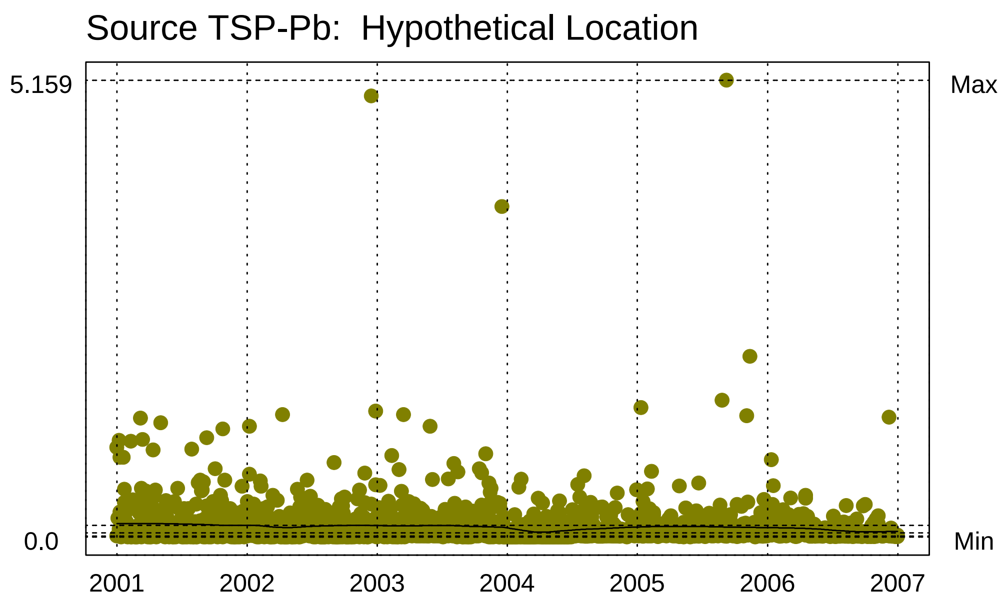
ID	N	Min	Q(25)	Median	Mean	Q(75)	Max	Std. Dev.	%CV	# of 0s	%0s
245100041	138	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	NA	138	100.00
170314201	119	0.01000	0.01000	0.01000	0.01050	0.01000	0.04000	0.00341	32	0	0.00
271377555	371	0.00000	0.00000	0.00000	0.00186	0.00000	0.02000	0.00416	224	306	82.48
270370423	227	0.00000	0.00000	0.00000	0.00198	0.00000	0.03000	0.00471	237	188	82.82
080310025	163	0.00070	0.00555	0.00770	0.00886	0.01135	0.03380	0.00499	56	0	0.00
340071007	52	0.00700	0.00700	0.00700	0.01063	0.01900	0.01900	0.00539	51	0	0.00
060375005	147	0.00000	0.00000	0.00000	0.00395	0.01000	0.02000	0.00556	141	94	63.95
270531007	265	0.00000	0.00000	0.00000	0.00321	0.01000	0.04000	0.00570	178	190	71.70
270370442	193	0.00000	0.00000	0.00000	0.00352	0.01000	0.03000	0.00578	164	133	68.91
270530966	239	0.00000	0.00000	0.00000	0.00485	0.01000	0.03000	0.00579	119	131	54.81
721270003	294	0.00000	0.00000	0.00000	0.00286	0.00000	0.05000	0.00588	206	221	75.17
450790006	177	0.00000	0.00000	0.00000	0.00346	0.00900	0.03800	0.00599	173	125	70.62
270530050	263	0.00000	0.00000	0.01000	0.00567	0.01000	0.03000	0.00601	106	127	48.29
270530963	262	0.00000	0.00000	0.00000	0.00397	0.01000	0.04000	0.00609	153	169	64.50
450450008	627	0.00000	0.00000	0.00000	0.00224	0.00000	0.08100	0.00615	274	519	82.78
121033005	365	0.00000	0.00000	0.00000	0.00055	0.00000	0.10000	0.00739	1349	363	99.45
261630019	225	0.00236	0.00597	0.00917	0.01072	0.01280	0.07667	0.00743	69	0	0.00
120310032	93	0.00000	0.00000	0.00400	0.00630	0.01000	0.05400	0.00865	137	43	46.24
450430009	456	0.00000	0.00000	0.00000	0.00430	0.00825	0.06800	0.00883	205	330	72.37
060731007	150	0.00070	0.00650	0.00985	0.01230	0.01500	0.05900	0.00889	72	0	0.00
170310001	347	0.00000	0.01000	0.01000	0.01415	0.02000	0.07000	0.00928	66	11	3.17
060651003	362	0.00000	0.01000	0.01000	0.01235	0.01000	0.06000	0.00931	75	54	14.92
360470122	174	0.00500	0.02000	0.02450	0.02466	0.03000	0.06000	0.00963	39	0	0.00
171170002	352	0.01000	0.01000	0.01000	0.01142	0.01000	0.17000	0.00968	85	0	0.00
260490021	351	0.00000	0.00500	0.00754	0.00948	0.01100	0.11610	0.00984	104	16	4.56
171430037	347	0.01000	0.01000	0.01000	0.01294	0.01000	0.09000	0.00997	77	0	0.00
110010039	80	0.00320	0.00700	0.01300	0.01346	0.02000	0.07600	0.01010	75	0	0.00
360850067	159	0.00000	0.00000	0.01000	0.01113	0.02000	0.05000	0.01043	94	51	32.08
250250002	224	0.00500	0.00500	0.01300	0.01416	0.01800	0.08800	0.01127	80	0	0.00
120310084	98	0.00000	0.00000	0.00700	0.00837	0.01200	0.09600	0.01140	136	34	34.69
270530964	109	0.00000	0.00000	0.00000	0.00404	0.01000	0.09000	0.01148	284	80	73.39
271231003	272	0.00000	0.00000	0.01000	0.00695	0.01000	0.13000	0.01171	169	126	46.32
482011034	352	0.00100	0.00600	0.00700	0.00766	0.00800	0.21000	0.01229	160	0	0.00
060658001	353	0.00000	0.01000	0.01000	0.01448	0.02000	0.07000	0.01242	86	60	17.00
270530965	256	0.00000	0.00000	0.00000	0.00484	0.01000	0.17000	0.01252	259	163	63.67
060371602	76	0.00000	0.01000	0.01000	0.01395	0.02000	0.09000	0.01337	96	15	19.74
270530053	122	0.00000	0.00000	0.01000	0.01033	0.01000	0.09000	0.01366	132	38	31.15
170310052	348	0.00000	0.01000	0.02000	0.02057	0.03000	0.12000	0.01409	68	8	2.30
421010004	49	0.00000	0.00000	0.02000	0.01653	0.03000	0.05000	0.01422	86	15	30.61
481410010	61	0.01000	0.01000	0.03000	0.03148	0.04000	0.06000	0.01611	51	0	0.00
060250005	351	0.00200	0.00695	0.01400	0.01734	0.02150	0.14200	0.01626	94	0	0.00
360713004	309	0.00000	0.03000	0.03000	0.02700	0.03000	0.27000	0.01764	65	1	0.32
420210808	358	0.03000	0.03000	0.04000	0.03737	0.04000	0.30000	0.01804	48	0	0.00
060711004	349	0.00000	0.01000	0.01000	0.01734	0.02000	0.29000	0.01914	110	35	10.03
060371601	253	0.00000	0.01000	0.02000	0.02609	0.03000	0.18000	0.02074	80	7	2.77
180892008	338	0.00000	0.00800	0.01300	0.01919	0.02400	0.23400	0.02221	116	9	2.66
481410002	109	0.01000	0.04000	0.06000	0.05945	0.07000	0.12000	0.02235	38	0	0.00

420030002	322	0.00000	0.00000	0.00000	0.00886	0.00000	0.11500	0.02256	255	278	86.34
060374002	378	0.00000	0.01000	0.01000	0.01638	0.02000	0.36000	0.02274	139	47	12.43
080650001	308	0.00000	0.00950	0.01190	0.01691	0.02000	0.30000	0.02311	137	20	6.49
271377001	279	0.00000	0.00000	0.00000	0.00315	0.00000	0.41000	0.02488	789	238	85.30
080410011	349	0.00000	0.00440	0.00900	0.01073	0.00980	0.31410	0.02872	268	40	11.46
060719004	**347**	**0.00000**	**0.01000**	**0.01000**	**0.02026**	**0.02000**	**0.35000**	**0.02927**	**144**	**36**	**10.37**
060371103	**374**	**0.00000**	**0.01000**	**0.02000**	**0.02578**	**0.03000**	**0.52000**	**0.03158**	**123**	**17**	**4.55**
170313301	**345**	**0.00000**	**0.01000**	**0.02000**	**0.02522**	**0.03000**	**0.44000**	**0.03728**	**148**	**1**	**0.29**
171193007	**358**	**0.01000**	**0.01000**	**0.01000**	**0.02257**	**0.02000**	**0.43000**	**0.04049**	**179**	**0**	**0.00**
060375001	**195**	**0.00000**	**0.01000**	**0.01000**	**0.01928**	**0.02000**	**0.60000**	**0.04558**	**236**	**32**	**16.41**
171630010	**356**	**0.01000**	**0.01000**	**0.02000**	**0.04253**	**0.05000**	**0.35000**	**0.04846**	**114**	**0**	**0.00**
060371301	**350**	**0.00000**	**0.01000**	**0.02000**	**0.02443**	**0.03000**	**0.96000**	**0.05300**	**217**	**20**	**5.71**
261630001	345	0.00109	0.00595	0.00781	0.01183	0.01095	1.01900	0.05457	461	0	0.00
080310002	341	0.00000	0.00930	0.01240	0.02532	0.02000	0.53500	0.05618	222	19	5.57
481130057	1659	0.00000	0.00500	0.00500	0.03235	0.04000	1.34000	0.06306	195	333	20.07
420070505	312	0.03000	0.03000	0.04000	0.07356	0.07000	0.67000	0.08801	120	0	0.00
220950003	112	0.00000	0.01100	0.02300	0.05649	0.05300	1.32700	0.13770	244	10	8.93
300490726	56	0.00000	0.03000	0.08000	0.14661	0.23250	0.53000	0.14512	99	1	1.79
420110003	59	0.03000	0.04000	0.08000	0.17407	0.24000	0.96000	0.20581	118	0	0.00

[1]The first two numbers of the ID identify the state, the next three identify the county, and the last four identify the site.
[2]Bolded indicate locations used to construct the hypothetical location.

Appendix C: Descriptive statistics, time series plots and distribution plots for the hypothetical source and nonsource locations.

Figure C-1. Time series plots of the hypothetical source and nonsource locations.



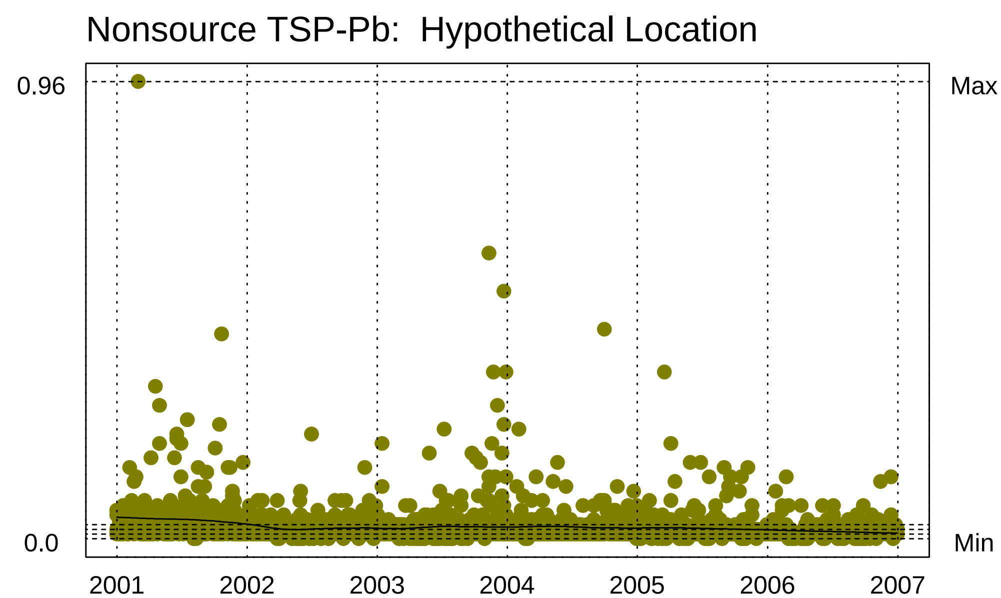


Figure C-2. Distribution plots and descriptive statistics for the hypothetical source location.

Descriptive Statistics:
N = 1730
Min. = 0
Q(.25) = 0.01
Median = 0.0445
Q(.75) = 0.1332
Max. = 5.159
Mean = 0.1214
SD = 0.2678
CS = 10.0738
CK = 160.2688
CV = 2.2059

D'Agostino & Pearson Test for Normality:
K2 = 2777.4439
p(K2) = 0
Z(CS) = 44.8599
p(Z(CS)) = 0
Z(CK) = 27.6592
p(Z(CK)) = 0

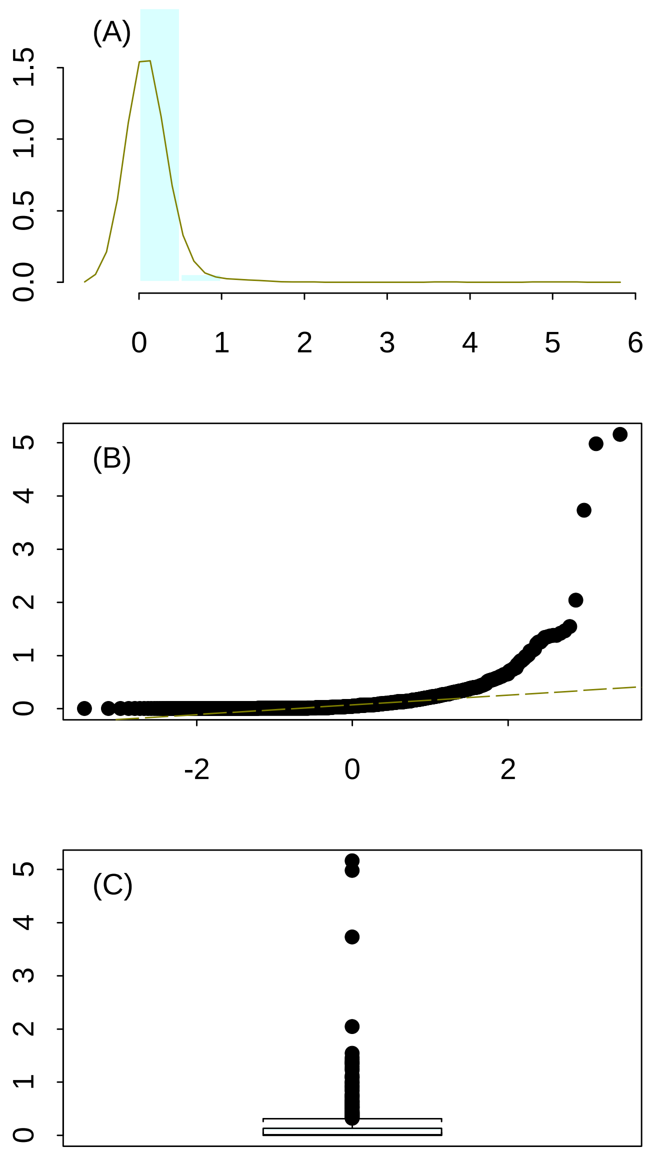
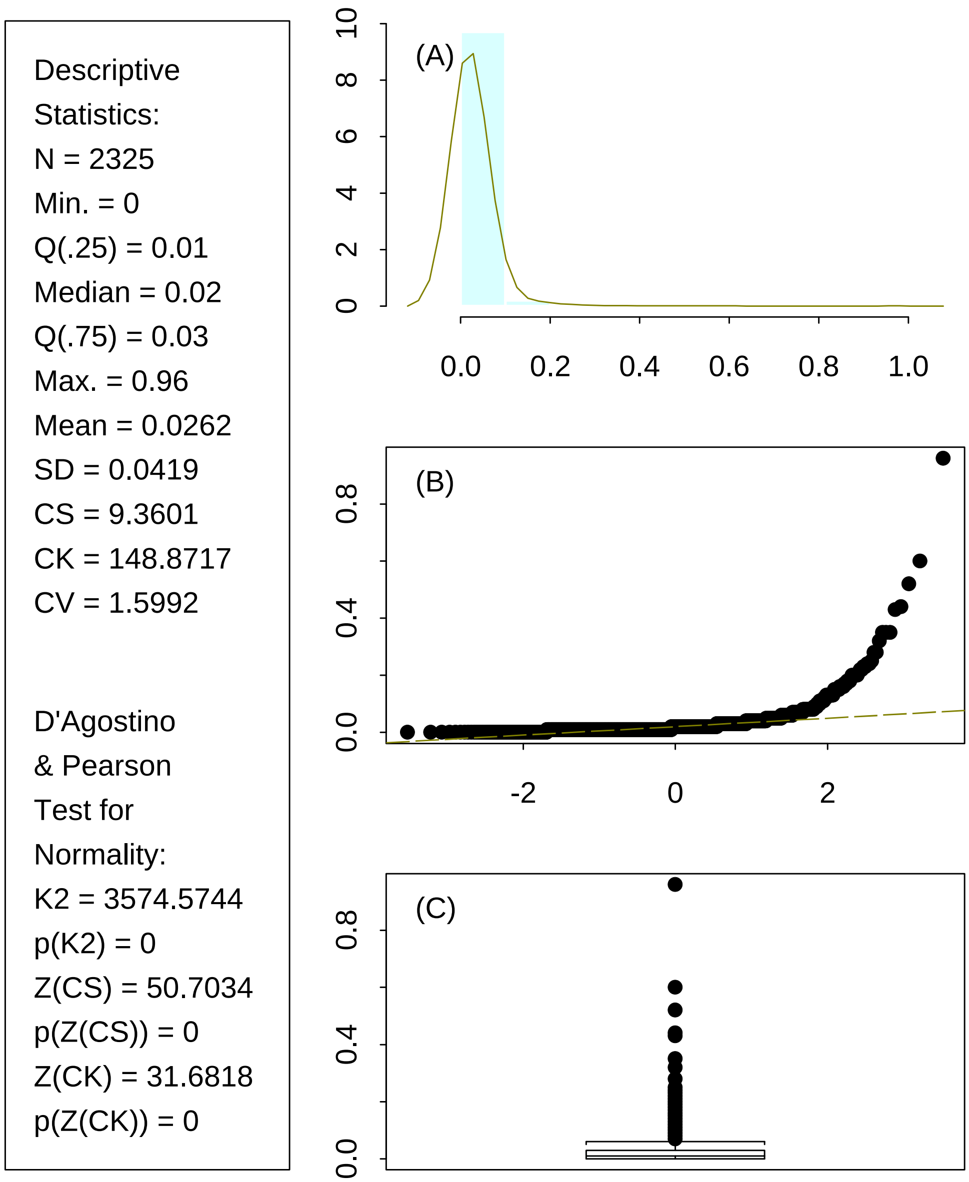


Figure C-2. Distribution plots and descriptive statistics for the hypothetical nonsource location.



Acronyms, description of statistics and how to interpret the figures in Appendix C.

Descriptive Statistics

N	sample size
Min.	minimum value of a set of observations
Q(.25)	25^{th} quantile; divides the data set such that one fourth of the observations fall below Q(.25) and three fourths lie above
Median	50^{th} quantile; divides the data set such that one half of the observations fall below Q(.50) and one half lie above
Q(.75)	75^{th} quantile; divides the data set such that three fourths of the observations fall below Q(.75) and one fourth lie above
Max.	maximum value of a set of observations
Mean	the arithmetic average of all the values in a set of observations; the mean is the most commonly used measure of central tendency.
SD	the standard deviation describe the dispersion relative to the center of a set of observations; the variance is the average of the squared deviation of each observation from the mean; the standard deviation is the square root of the variance
CS	coefficient of skewness; the third moment about the mean is a measure of asymmetry; symmetrical distributions will have a skewness of 0, distributions that are skewed to the left will have a skewness < 0, and distributions that are skewed to the right will have a skewness > 0
CK	coefficient of kurtosis; the fourth moment about the mean is a measure of curvature or kurtosis, which is the degree of flatness of a density near its center; values close to $3(n - 1)/(n + 1)$ indicate normality
CV	coefficient of variation; the mean divided by the standard deviation

D'Agostino & Pearson Test for Normality

K2	the test statistic for an omnibus test of normality based on the coefficients of skewness and kurtosis; the null hypothesis is that the data are normally distributed
p(Z(K2))	the probability of observing a value of Z(CS) or one greater
Z(CS)	the test statistic for an inferential test for detecting nonnormality due to skewness; the null hypothesis is the CS = 0
p(Z(CS))	the probability of observing a value of Z(CS) or one greater
Z(CK)	the test statistic for an inferential test for detecting nonnormality due to kurtosis; the null hypothesis is the CK = 3
p(Z(CK))	the probability of observing a value of Z(CK) or one greater

Figures

(A) A ***histogram*** partitions the range of the data into several nonoverlapping intervals of equal length, called bins, and counts the number observations in each bin. The number of counts in each bin can be displayed on a density scale, where the y-axis represents the probability; or a nondensity or frequency scale, where the y-axis represents the bin counts. The histogram is completely determined by two parameters, the *bin width* and the *bin origin*.

The histogram of a set of observations that are normally distributed will appear unimodal and symmetric.

(B) A ***normal quantile-quantile plot (Q-Q plot)*** is obtained by plotting the quantiles of the observed data against the corresponding quantiles of the normal distribution. If the quantiles of the empirical distribution and the quantiles of the normal distribution, fall on a straight line then the distributions are similar.

(C) A ***box plot*** is a rectangle, the top and bottom of the rectangle represent the upper and lower quartiles of the data, the horizontal line within the rectangle represents the median. Lines, in the shape of a "T", extend from the box to the nearest value not beyond a *standard span* from the quartiles. These lines are often referred to as whiskers. Values beyond the end of the whiskers are drawn individually.

The ***standard span*** is 1.5·Inter-Quartile Range (IQR) , where the *upper quartile* is the 75^{th} quantile, Q(.75), the *lower quartile* is the 25^{th} quantile, Q(.25) and the *IQR* = Q(.75) Q(.25).

The box plot of a set of observations that are normally distributed will be symmetric with the median in the center of the box.

Appendix D: Bootstrap R Code

```
#Bootstrap Samples for Source Locations
#Construct the bootstrap samples for each month assuming daily sampling and 100%
#completeness.   These bootstrap samples are used to construct the various combinations
#sampling frequency, completeness, precision and bias

SF <- read.csv(file="TSPSourceFinal.csv")

Conc <- SF$Conc
Month <- SF$Month

R <- 2500 #number of bootstrap samples for each month
sample.size <- c(31,29,31,30,31,30,31,31,30,31,30,31)

#Construct the 2500 bootstrap samples for each month for the first year
SJan.boot1 <- matrix(rep(0, R*sample.size[1]), nrow=R)
for (i in 1:R)
{
	SJan.boot1[i,]<- sample(Conc[Month==1], sample.size[1], replace = FALSE)
}

SFeb.boot1 <- matrix(rep(0, R*sample.size[2]), nrow=R)
for (i in 1:R)
{
	SFeb.boot1[i,]<- sample(Conc[Month==2], sample.size[2], replace = FALSE)
}

SMar.boot1 <- matrix(rep(0, R*sample.size[3]), nrow=R)
for (i in 1:R)
{
	SMar.boot1[i,]<- sample(Conc[Month==3], sample.size[3], replace = FALSE)
}

SApr.boot1 <- matrix(rep(0, R*sample.size[4]), nrow=R)
for (i in 1:R)
{
	SApr.boot1[i,]<- sample(Conc[Month==4], sample.size[4], replace = FALSE)
}

SMay.boot1 <- matrix(rep(0, R*sample.size[5]), nrow=R)
for (i in 1:R)
{
	SMay.boot1[i,]<- sample(Conc[Month==5], sample.size[5], replace = FALSE)
}

SJun.boot1 <- matrix(rep(0, R*sample.size[6]), nrow=R)
for (i in 1:R)
```

```
{
	SJun.boot1[i,]<- sample(Conc[Month==6], sample.size[6], replace = FALSE)
}

SJul.boot1<- matrix(rep(0, R*sample.size[7]), nrow=R)
for (i in 1:R)
{
	SJul.boot1[i,]<- sample(Conc[Month==7], sample.size[7], replace = FALSE)
}

SAug.boot1 <- matrix(rep(0, R*sample.size[8]), nrow=R)
for (i in 1:R)
{
	SAug.boot1[i,]<- sample(Conc[Month==8], sample.size[8], replace = FALSE)
}

SSep.boot1 <- matrix(rep(0, R*sample.size[9]), nrow=R)
for (i in 1:R)
{
	SSep.boot1[i,]<- sample(Conc[Month==9], sample.size[9], replace = FALSE)
}

SOct.boot1 <- matrix(rep(0, R*sample.size[10]), nrow=R)
for (i in 1:R)
{
	SOct.boot1[i,]<- sample(Conc[Month==10], sample.size[10], replace = FALSE)
}

SNov.boot1 <- matrix(rep(0, R*sample.size[11]), nrow=R)
for (i in 1:R)
{
	SNov.boot1[i,]<- sample(Conc[Month==11], sample.size[11], replace = FALSE)
}

SDec.boot1 <- matrix(rep(0, R*sample.size[12]), nrow=R)
for (i in 1:R)
{
	SDec.boot1[i,]<- sample(Conc[Month==12], sample.size[12], replace = FALSE)
}

#Construct the 2500 bootstrap samples for each month for the second year
SJan.boot2 <- matrix(rep(0, R*sample.size[1]), nrow=R)
for (i in 1:R)
{
	SJan.boot2[i,]<- sample(Conc[Month==1], sample.size[1], replace = FALSE)
}

SFeb.boot2 <- matrix(rep(0, R*sample.size[2]), nrow=R)
```

```
for (i in 1:R)
{
	SFeb.boot2[i,]<- sample(Conc[Month==2], sample.size[2], replace = FALSE)
}

SMar.boot2 <- matrix(rep(0, R*sample.size[3]), nrow=R)
for (i in 1:R)
{
	SMar.boot2[i,]<- sample(Conc[Month==3], sample.size[3], replace = FALSE)
}

SApr.boot2 <- matrix(rep(0, R*sample.size[4]), nrow=R)
for (i in 1:R)
{
	SApr.boot2[i,]<- sample(Conc[Month==4], sample.size[4], replace = FALSE)
}

SMay.boot2 <- matrix(rep(0, R*sample.size[5]), nrow=R)
for (i in 1:R)
{
	SMay.boot2[i,]<- sample(Conc[Month==5], sample.size[5], replace = FALSE)
}

SJun.boot2 <- matrix(rep(0, R*sample.size[6]), nrow=R)
for (i in 1:R)
{
	SJun.boot2[i,]<- sample(Conc[Month==6], sample.size[6], replace = FALSE)
}

SJul.boot2 <- matrix(rep(0, R*sample.size[7]), nrow=R)
for (i in 1:R)
{
	SJul.boot2[i,]<- sample(Conc[Month==7], sample.size[7], replace = FALSE)
}

SAug.boot2 <- matrix(rep(0, R*sample.size[8]), nrow=R)
for (i in 1:R)
{
	SAug.boot2[i,]<- sample(Conc[Month==8], sample.size[8], replace = FALSE)
}

SSep.boot2 <- matrix(rep(0, R*sample.size[9]), nrow=R)
for (i in 1:R)
{
	SSep.boot2[i,]<- sample(Conc[Month==9], sample.size[9], replace = FALSE)
}

SOct.boot2 <- matrix(rep(0, R*sample.size[10]), nrow=R)
```

```
for (i in 1:R)
{
	SOct.boot2[i,]<- sample(Conc[Month==10], sample.size[10], replace = FALSE)
}

SNov.boot2 <- matrix(rep(0, R*sample.size[11]), nrow=R)
for (i in 1:R)
{
	SNov.boot2[i,]<- sample(Conc[Month==11], sample.size[11], replace = FALSE)
}

SDec.boot2 <- matrix(rep(0, R*sample.size[12]), nrow=R)
for (i in 1:R)
{
	SDec.boot2[i,]<- sample(Conc[Month==12], sample.size[12], replace = FALSE)
}

#Construct the 2500 bootstrap samples for each month for the third year
SJan.boot3 <- matrix(rep(0, R*sample.size[1]), nrow=R)
for (i in 1:R)
{
	SJan.boot3[i,]<- sample(Conc[Month==1], sample.size[1], replace = FALSE)
}

SFeb.boot3 <- matrix(rep(0, R*sample.size[2]), nrow=R)
for (i in 1:R)
{
	SFeb.boot3[i,]<- sample(Conc[Month==2], sample.size[2], replace = FALSE)
}

SMar.boot3 <- matrix(rep(0, R*sample.size[3]), nrow=R)
for (i in 1:R)
{
	SMar.boot3[i,]<- sample(Conc[Month==3], sample.size[3], replace = FALSE)
}

SApr.boot3 <- matrix(rep(0, R*sample.size[4]), nrow=R)
for (i in 1:R)
{
	SApr.boot3[i,]<- sample(Conc[Month==4], sample.size[4], replace = FALSE)
}

SMay.boot3 <- matrix(rep(0, R*sample.size[5]), nrow=R)
for (i in 1:R)
{
	SMay.boot3[i,]<- sample(Conc[Month==5], sample.size[5], replace = FALSE)
}
```

```
SJun.boot3 <- matrix(rep(0, R*sample.size[6]), nrow=R)
for (i in 1:R)
{
	SJun.boot3[i,]<- sample(Conc[Month==6], sample.size[6], replace = FALSE)
}

SJul.boot3 <- matrix(rep(0, R*sample.size[7]), nrow=R)
for (i in 1:R)
{
	SJul.boot3[i,]<- sample(Conc[Month==7], sample.size[7], replace = FALSE)
}

SAug.boot3 <- matrix(rep(0, R*sample.size[8]), nrow=R)
for (i in 1:R)
{
	SAug.boot3[i,]<- sample(Conc[Month==8], sample.size[8], replace = FALSE)
}

SSep.boot3 <- matrix(rep(0, R*sample.size[9]), nrow=R)
for (i in 1:R)
{
	SSep.boot3[i,]<- sample(Conc[Month==9], sample.size[9], replace = FALSE)
}

SOct.boot3 <- matrix(rep(0, R*sample.size[10]), nrow=R)
for (i in 1:R)
{
	SOct.boot3[i,]<- sample(Conc[Month==10], sample.size[10], replace = FALSE)
}

SNov.boot3 <- matrix(rep(0, R*sample.size[11]), nrow=R)
for (i in 1:R)
{
	SNov.boot3[i,]<- sample(Conc[Month==11], sample.size[11], replace = FALSE)
}

SDec.boot3 <- matrix(rep(0, R*sample.size[12]), nrow=R)
for (i in 1:R)
{
	SDec.boot3[i,]<- sample(Conc[Month==12], sample.size[12], replace = FALSE)
}


#Precision Measures for Source Locations – Monthly & Quarterly

Bias.Choices <- c(0, 0.05, 0.10, 0.15, -0.05, -0.10, -0.15)
Precision.Choices <- c(0, 0.1, 0.2, 0.3)
Completeness.Choices <- c(1, 0.9, 0.75)
```

```
SamplingPlan.Choices <- c( 1, 0.333, 0.167)

Bias <- Bias.Choices[7]
Precision <- Precision.Choices[4]
Completeness <- Completeness.Choices[3]
Sampling.Plan <- SamplingPlan.Choices[2]

#Monthly Means for 3 Year 3
M1.1<-matrix(rep(0, R), ncol=1); M2.1<-matrix(rep(0, R), ncol=1)
M3.1<-matrix(rep(0, R), ncol=1); M4.1<-matrix(rep(0, R), ncol=1)
M5.1<-matrix(rep(0, R), ncol=1); M6.1<-matrix(rep(0, R), ncol=1)
M7.1<-matrix(rep(0, R), ncol=1); M8.1<-matrix(rep(0, R), ncol=1)
M9.1<-matrix(rep(0, R), ncol=1); M10.1<-matrix(rep(0, R), ncol=1)
M11.1<-matrix(rep(0, R), ncol=1); M12.1<-matrix(rep(0, R), ncol=1)

M1.2<-matrix(rep(0, R), ncol=1); M2.2<-matrix(rep(0, R), ncol=1)
M3.2<-matrix(rep(0, R), ncol=1); M4.2<-matrix(rep(0, R), ncol=1)
M5.2<-matrix(rep(0, R), ncol=1); M6.2<-matrix(rep(0, R), ncol=1)
M7.2<-matrix(rep(0, R), ncol=1); M8.2<-matrix(rep(0, R), ncol=1)
M9.2<-matrix(rep(0, R), ncol=1); M10.2<-matrix(rep(0, R), ncol=1)
M11.2<-matrix(rep(0, R), ncol=1); M12.2<-matrix(rep(0, R), ncol=1)

M1.3<-matrix(rep(0, R), ncol=1); M2.3<-matrix(rep(0, R), ncol=1)
M3.3<-matrix(rep(0, R), ncol=1); M4.3<-matrix(rep(0, R), ncol=1)
M5.3<-matrix(rep(0, R), ncol=1); M6.3<-matrix(rep(0, R), ncol=1)
M7.3<-matrix(rep(0, R), ncol=1); M8.3<-matrix(rep(0, R), ncol=1)
M9.3<-matrix(rep(0, R), ncol=1); M10.3<-matrix(rep(0, R), ncol=1)
M11.3<-matrix(rep(0, R), ncol=1); M12.3<-matrix(rep(0, R), ncol=1)

for(i in 1:R)
{
#January
S1.1<-sample(SJan.boot1[i,], round((sample.size[1]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S1.2<-sample(SJan.boot2[i,], round((sample.size[1]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S1.3<-sample(SJan.boot3[i,], round((sample.size[1]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M1.1[i] <- mean(S1.1 + rnorm(length(S1.1), 0, S1.1*Precision) + S1.1*Bias)
M1.2[i] <- mean(S1.2 + rnorm(length(S1.2), 0, S1.2*Precision) + S1.2*Bias)
M1.3[i] <- mean(S1.3 + rnorm(length(S1.3), 0, S1.3*Precision) + S1.3*Bias)

#February
S2.1<-sample(SFeb.boot1[i,], round((sample.size[2]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S2.2<-sample(SFeb.boot2[i,], round((sample.size[2]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S2.3<-sample(SFeb.boot3[i,], round((sample.size[2]*Completeness*Sampling.Plan), 0),
```

```
                    replace = FALSE)
M2.1[i] <- mean(S2.1 + rnorm(length(S2.1), 0, S2.1*Precision) + S2.1*Bias)
M2.2[i] <- mean(S2.2 + rnorm(length(S2.2), 0, S2.2*Precision) + S2.2*Bias)
M2.3[i] <- mean(S2.3 + rnorm(length(S2.3), 0, S2.3*Precision) + S2.3*Bias)

#March
S3.1<-sample(SMar.boot1[i,], round((sample.size[3]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S3.2<-sample(SMar.boot2[i,], round((sample.size[3]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S3.3<-sample(SMar.boot3[i,], round((sample.size[3]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
M3.1[i] <- mean(S3.1 + rnorm(length(S3.1), 0, S3.1*Precision) + S3.1*Bias)
M3.2[i] <- mean(S3.2 + rnorm(length(S3.2), 0, S3.2*Precision) + S3.2*Bias)
M3.3[i] <- mean(S3.3 + rnorm(length(S3.3), 0, S3.3*Precision) + S3.3*Bias)

#April
S4.1<-sample(SApr.boot1[i,], round((sample.size[4]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S4.2<-sample(SApr.boot2[i,], round((sample.size[4]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S4.3<-sample(SApr.boot3[i,], round((sample.size[4]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
M4.1[i] <- mean(S4.1 + rnorm(length(S4.1), 0, S4.1*Precision) + S4.1*Bias)
M4.2[i] <- mean(S4.2 + rnorm(length(S4.2), 0, S4.2*Precision) + S4.2*Bias)
M4.3[i] <- mean(S4.3 + rnorm(length(S4.3), 0, S4.3*Precision) + S4.3*Bias)

#May
S5.1<-sample(SMay.boot1[i,], round((sample.size[5]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S5.2<-sample(SMay.boot2[i,], round((sample.size[5]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S5.3<-sample(SMay.boot3[i,], round((sample.size[5]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
M5.1[i] <- mean(S5.1 + rnorm(length(S5.1), 0, S5.1*Precision) + S5.1*Bias)
M5.2[i] <- mean(S5.2 + rnorm(length(S5.2), 0, S5.2*Precision) + S5.2*Bias)
M5.3[i] <- mean(S5.3 + rnorm(length(S5.3), 0, S5.3*Precision) + S5.3*Bias)

#April
S6.1<-sample(SJun.boot1[i,], round((sample.size[6]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S6.2<-sample(SJun.boot2[i,], round((sample.size[6]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
S6.3<-sample(SJun.boot3[i,], round((sample.size[6]*Completeness*Sampling.Plan), 0),
                    replace = FALSE)
M6.1[i] <- mean(S6.1 + rnorm(length(S6.1), 0, S6.1*Precision) + S6.1*Bias)
M6.2[i] <- mean(S6.2 + rnorm(length(S6.2), 0, S6.2*Precision) + S6.2*Bias)
M6.3[i] <- mean(S6.3 + rnorm(length(S6.3), 0, S6.3*Precision) + S6.3*Bias)
```

```
#July
S7.1<-sample(SJul.boot1[i,], round((sample.size[7]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S7.2<-sample(SJul.boot2[i,], round((sample.size[7]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S7.3<-sample(SJul.boot3[i,], round((sample.size[7]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M7.1[i] <- mean(S7.1 + rnorm(length(S7.1), 0, S7.1*Precision) + S7.1*Bias)
M7.2[i] <- mean(S7.2 + rnorm(length(S7.2), 0, S7.2*Precision) + S7.2*Bias)
M7.3[i] <- mean(S7.3 + rnorm(length(S7.3), 0, S7.3*Precision) + S7.3*Bias)

#August
S8.1<-sample(SAug.boot1[i,], round((sample.size[8]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S8.2<-sample(SAug.boot2[i,], round((sample.size[8]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S8.3<-sample(SAug.boot3[i,], round((sample.size[8]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M8.1[i] <- mean(S8.1 + rnorm(length(S8.1), 0, S8.1*Precision) + S8.1*Bias)
M8.2[i] <- mean(S8.2 + rnorm(length(S8.2), 0, S8.2*Precision) + S8.2*Bias)
M8.3[i] <- mean(S8.3 + rnorm(length(S8.3), 0, S8.3*Precision) + S8.3*Bias)

#September
S9.1<-sample(SSep.boot1[i,], round((sample.size[9]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S9.2<-sample(SSep.boot2[i,], round((sample.size[9]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S9.3<-sample(SSep.boot3[i,], round((sample.size[9]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M9.1[i] <- mean(S9.1 + rnorm(length(S9.1), 0, S9.1*Precision) + S9.1*Bias)
M9.2[i] <- mean(S9.2 + rnorm(length(S9.2), 0, S9.2*Precision) + S9.2*Bias)
M9.3[i] <- mean(S9.3 + rnorm(length(S9.3), 0, S9.3*Precision) + S9.3*Bias)

#October
S10.1<-sample(SOct.boot1[i,], round((sample.size[10]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S10.2<-sample(SOct.boot2[i,], round((sample.size[10]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S10.3<-sample(SOct.boot3[i,], round((sample.size[10]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M10.1[i] <- mean(S10.1 + rnorm(length(S10.1), 0, S10.1*Precision) + S10.1*Bias)
M10.2[i] <- mean(S10.2 + rnorm(length(S10.2), 0, S10.2*Precision) + S10.2*Bias)
M10.3[i] <- mean(S10.3 + rnorm(length(S10.3), 0, S10.3*Precision) + S10.3*Bias)

#November
S11.1<-sample(SNov.boot1[i,], round((sample.size[11]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
```

```
S11.2<-sample(SNov.boot2[i,], round((sample.size[11]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S11.3<-sample(SNov.boot3[i,], round((sample.size[11]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M11.1[i] <- mean(S11.1 + rnorm(length(S11.1), 0, S11.1*Precision) + S11.1*Bias)
M11.2[i] <- mean(S11.2 + rnorm(length(S11.2), 0, S11.2*Precision) + S11.2*Bias)
M11.3[i] <- mean(S11.3 + rnorm(length(S11.3), 0, S11.3*Precision) + S11.3*Bias)

#December
S12.1<-sample(SDec.boot1[i,], round((sample.size[12]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S12.2<-sample(SDec.boot2[i,], round((sample.size[12]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
S12.3<-sample(SDec.boot3[i,], round((sample.size[12]*Completeness*Sampling.Plan), 0),
          replace = FALSE)
M12.1[i] <- mean(S12.1 + rnorm(length(S12.1), 0, S12.1*Precision) + S12.1*Bias)
M12.2[i] <- mean(S12.2 + rnorm(length(S12.2), 0, S12.2*Precision) + S12.2*Bias)
M12.3[i] <- mean(S12.3 + rnorm(length(S12.3), 0, S12.3*Precision) + S12.3*Bias)
}

dataQ <- cbind(
          apply(cbind(M11.1, M12.1, M1.1), 1, mean),
          apply(cbind(M12.1, M1.1, M2.1), 1, mean),
          apply(cbind(M1.1, M2.1, M3.1), 1, mean),
          apply(cbind(M2.1, M3.1, M4.1), 1, mean),
          apply(cbind(M3.1, M4.1, M5.1), 1, mean),
          apply(cbind(M4.1, M5.1, M6.1), 1, mean),
          apply(cbind(M5.1, M6.1, M7.1), 1, mean),
          apply(cbind(M6.1, M7.1, M8.1), 1, mean),
          apply(cbind(M7.1, M8.1, M9.1), 1, mean),
          apply(cbind(M8.1, M9.1, M10.1), 1, mean),
          apply(cbind(M9.1, M10.1, M11.1), 1, mean),
          apply(cbind(M10.1, M11.1, M12.1), 1, mean),
          apply(cbind(M11.2, M12.2, M1.2), 1, mean),
          apply(cbind(M12.2, M1.2, M2.2), 1, mean),
          apply(cbind(M1.2, M2.2, M3.2), 1, mean),
          apply(cbind(M2.2, M3.2, M4.2), 1, mean),
          apply(cbind(M3.2, M4.2, M5.2), 1, mean),
          apply(cbind(M4.2, M5.2, M6.2), 1, mean),
          apply(cbind(M5.2, M6.2, M7.2), 1, mean),
          apply(cbind(M6.2, M7.2, M8.2), 1, mean),
          apply(cbind(M7.2, M8.2, M9.2), 1, mean),
          apply(cbind(M8.2, M9.2, M10.2), 1, mean),
          apply(cbind(M9.2, M10.2, M11.2), 1, mean),
          apply(cbind(M10.2, M11.2, M12.2), 1, mean),
          apply(cbind(M11.3, M12.3, M1.3), 1, mean),
          apply(cbind(M12.3, M1.3, M2.3), 1, mean),
          apply(cbind(M1.3, M2.3, M3.3), 1, mean),
```

```
          apply(cbind(M2.3, M3.3, M4.3), 1, mean),
          apply(cbind(M3.3, M4.3, M5.3), 1, mean),
          apply(cbind(M4.3, M5.3, M6.3), 1, mean),
          apply(cbind(M5.3, M6.3, M7.3), 1, mean),
          apply(cbind(M6.3, M7.3, M8.3), 1, mean),
          apply(cbind(M7.3, M8.3, M9.3), 1, mean),
          apply(cbind(M8.3, M9.3, M10.3), 1, mean),
          apply(cbind(M9.3, M10.3, M11.3), 1, mean),
          apply(cbind(M10.3, M11.3, M12.3), 1, mean) )

     DVQ1<-rep(0,R); DVQ2<-rep(0,R); DVQ3<-rep(0,R)

     for (i in 1:R)
     {
          DVQ1[i]<- length(dataQ[i,][dataQ[i,]>0.10])
          DVQ2[i]<- length(dataQ[i,][dataQ[i,]>0.20])
          DVQ3[i]<- length(dataQ[i,][dataQ[i,]>0.30])
     }

dataM <- cbind(M1.1,M2.1,M3.1,M4.1,M5.1,M6.1,M7.1,M8.1,M9.1,M10.1,M11.1,M12.1,
        M1.2,M2.2,M3.2,M4.2,M5.2,M6.2,M7.2,M8.2,M9.2,M10.2,M11.2,M12.2,
        M1.3,M2.3,M3.3,M4.3,M5.3,M6.3,M7.3,M8.3,M9.3,M10.3,M11.3,M12.3)

     DVM1<-rep(0,R); DVM2<-rep(0,R); DVM3<-rep(0,R)

     for (i in 1:R)
     {
          DVM1[i]<- length(dataM[i,][dataM[i,]>0.10])
          DVM2[i]<- length(dataM[i,][dataM[i,]>0.20])
          DVM3[i]<- length(dataM[i,][dataM[i,]>0.30])
     }


#Transform the non-source design values so the DV with daily sampling, 100% #completeness,
0% bias, and a precision of one
     #var.adj <- 0.0257/sqrt(var(DVNS))
     #mn.adj <- 0.2000 - mean(DVNS*var.adj)
     #DVT <- DV*var.adj + mn.adj

     upper95<-function(x) {sort(x)[round(0.975*36, 0)]}
     lower95<-function(x) {sort(x)[round(0.025*36, 0)]}

#win.graph(height=11, width=8.5)
#par(mfrow=c(2,1))
DQ <- sample(dataQ, 2500)
#Check the distribution against a lognormal
par(pty = "s")
pel <- fitdistr(DQ, "lognormal")
```

```
rl <- rlnorm(2500,  pel[1]$estimate[[1]], pel[1]$estimate[[2]])
qqplot(DQ, qlnorm(ppoints(rl),
        meanlog = pel[1]$estimate[[1]],
        sdlog = pel[1]$estimate[[2]]), xlab = “Quarterly Empirical”, ylab = “”)
abline(0, 1, col = 2, lty = 2)

par(pty = “s”)
DM <- sample(dataM, 2500)
pel <- fitdistr(DM, "lognormal")
rl <- rlnorm(2500,  pel[1]$estimate[[1]], pel[1]$estimate[[2]])
qqplot(DM, qlnorm(ppoints(rl),
        meanlog = pel[1]$estimate[[1]],
        sdlog = pel[1]$estimate[[2]]), xlab = “Monthly Empirical”, ylab = “”)
abline(0, 1, col = 2, lty = 2)

#Precision estimates for Monthly
        round(mean(apply(dataM, 1, mean)), 4)
        round(mean(sqrt(apply(dataM, 1, var))), 4)
        round(mean(apply(dataM, 1, lower95)), 4)
        round(mean(apply(dataM, 1, upper95)), 4)
        round((length(DVM1[DVM1 < 2])/2500)*100, 2)
        round((length(DVM2[DVM2 < 2])/2500)*100, 2)
        round((length(DVM3[DVM3 < 2])/2500)*100, 2)

#Precision estimates for Quarterly
        round(mean(apply(dataQ, 1, mean)), 4)
        round(mean(sqrt(apply(dataQ, 1, var))) , 4)
        round(mean(apply(dataQ, 1, lower95)), 4)
        round(mean(apply(dataQ, 1, upper95)), 4)
        round((length(DVQ1[DVQ1==0])/2500)*100, 2)
        round((length(DVQ2[DVQ2==0])/2500)*100, 2)
        round((length(DVQ3[DVQ3==0])/2500)*100, 2)
```