UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF WATER

MEMORANDUM

DATE: August 11, 2011

SUBJECT: Determination of Effluent Limits for Flue Gas Desulfurization (FGD) Wastewater at PSNH Merrimack Station

TO: Sharon DeMeo
EPA Region 1
Industrial Permits Branch

FROM: Ronald Jordan and Cuc Schroeder
Engineering and Analysis Division (4303T)

The attached report, "Determination of Effluent Limits for Flue Gas Desulfurization (FGD) Wastewater at PSNH Merrimack Station Based on the Performance of Physical-Chemical Treatment Followed by Biological Treatment," presents the results of statistical analyses performed on treatment system performance data to calculate effluent limitations for inclusion in Merrimack Station's NPDES permit. Specifically, the attached report presents "daily maximum" and "monthly average" effluent limits for the following parameters: arsenic, chromium, copper, mercury, selenium, and zinc. These effluent limits are based on analyses of self-monitoring data collected by two power plants over a period of several years. Both power plants operate treatment systems that incorporate physical-chemical treatment (i.e., chemical precipitation/iron coprecipitation) followed by anoxic/anaerobic biological treatment of the FGD wastewater.

Section I of the report provides an overview of the objectives for the statistical and engineering analyses we conducted, and presents a general description of the power plants and FGD wastewater treatment systems associated with the treatment system performance data. Section I also provides an overview of the statistical methods used to evaluate the data and calculate effluent limits. Section II describes the data that were used to calculate effluent limits, and explains the rationale for excluding certain observations. Section III provides the statistical methodology used to calculate the limits, including a description of the statistical models and equations used, and explains how the data for multiple plants are combined to derive a single set of effluent limits. Section IV provides plots of the data, summary statistics,

long term averages, and daily and monthly variability factors for each pollutant from each plant. Section V presents the daily maximum and monthly average effluent limits based on the combined results for the two plants.

In addition to the five sections described above, there are six appendices in the attached report. Appendix 1 contains plots of all available data for the pollutants for the limits were calculated, including each data point that was ultimately excluded from the analysis. Appendix 2 contains plots similar to those in Appendix 1, with the addition of the smoothed curves superimposed over the data, but without making a distinction between quantified and non-quantified values. Appendix 3 contains a list of all observations that in a statistical screening test were deemed to be extreme observations (i.e., potential data outliers that may indicate treatment system upset or other abnormalities), and identifies which of those observations were actually excluded and the reasons for the exclusion. Data from the physical-chemical and bioreactor effluent sampling locations were subjected to this screening process. Appendix 4 contains plots of the natural logarithm of the data used to evaluate treatment system performance and calculate the effluent limits (i.e., excluded data are not shown on the plots). Appendix 5 presents plots of the same data shown in Appendix 4 (but not log-transformed), with the actual concentration results plotted on a logarithmic scale. Appendix 6 presents the results of analyses to determine whether the effluent data used to calculate effluent limits are autocorrelated.

DETERMINATION OF EFFLUENT LIMITS FOR FLUE GAS DESULFURIZATION (FGD) WASTEWATER AT PSNH MERRIMACK STATION BASED ON THE PERFORMANCE OF PHYSICAL-CHEMICAL TREATMENT FOLLOWED BY BIOLOGICAL TREATMENT

U.S. Environmental Protection Agency
Office of Water
Engineering and Analysis Division (4303T)

August 11, 2011

TABLE OF CONTENTS

I. INTRODUCTION

The sections below provide the purpose of the report, descriptions of Duke Energy's Allen and Belews Creek Stations and their flue gas desulfurization (FGD) wastewater treatment systems, and an overview of the statistical methods used in setting the limits.

1. Purpose

EPA's Engineering and Analysis Division prepared this report in response to requests for information about the efficacy of treatment systems employing a combination of physical-chemical treatment and biological treatment for FGD wastewater, and for a statistical evaluation of treatment system performance and achievable effluent levels for inclusion in NPDES permits.

This report presents characterization data for certain pollutants present in untreated and treated FGD wastewater from two coal-fired power plants that operate treatment systems incorporating physical-chemical treatment (i.e., chemical precipitation/iron coprecipitation) followed by anoxic/anaerobic biological treatment. In addition, this report presents "daily maximum" and "monthly average" effluent limits based on statistical analyses of data for the treatment system effluent, and describes how those effluent limits were calculated. Finally, the data for untreated and treated wastewater are described in detail, using various data plots and tables of summary statistics, long term averages, and variability factors.

2. Overview of Allen and Belews Creek Stations and Their Flue Gas Desulfurization (FGD) Wastewater Treatment Systems

a. Overview of Allen and Belews Creek Stations

Duke Energy's Allen Station and Belews Creek Station are coal-fired power plants that burn bituminous coal to generate electricity. Both plants operate limestone forced oxidation wet flue gas desulfurization (FGD) systems to reduce sulfur dioxide (SO_2) emissions, producing a commercial-grade gypsum byproduct. Allen Station operates selective non-catalytic reduction (SNCR) systems on all five generating units to reduce emissions of nitrogen oxides (NO_x); Belews Creek Station operates selective catalytic reduction (SCR) systems on both units to reduce NO_x. Both plants operate a wastewater treatment system designed to remove pollutants from FGD wastewater prior to discharge using a combination of physical-chemical treatment (i.e., one-stage chemical precipitation/iron co-precipitation) and anoxic/anaerobic biological treatment.

Allen Station operates five coal-fired steam electric generating units. Units 1 and 2 came on line in 1957, and each has a capacity of 165 megawatts (MW). Units 3, 4, and 5 came on line in 1959, 1960, and 1961, respectively, and each has a capacity of 275 MW. The electric generating units at Allen Station are operated on a cycling basis and are not baseload units.

Two FGD scrubber systems (Absorbers 1 and 3) are used to control the SO_2 emissions from the five generating units at Allen Station. Units 1, 2, and 5 exhaust to FGD Absorber 1 (which began operation in February 2009) and Units 3 and 4 exhaust to FGD Absorber 3 (which began operation in May 2009).

Belews Creek Station operates two baseload steam electric units. Units 1 and 2 came on line in 1974 and 1975, respectively, and each has a generating capacity of 1,200 MW. The Unit 1 FGD system was installed in February 2008, and the Unit 2 FGD system was installed in May 2008.

b. FGD Wastewater Treatment System

Both power plants operate treatment systems that incorporate physical-chemical treatment (i.e., chemical precipitation/iron coprecipitation) followed by anoxic/anaerobic biological treatment of the FGD wastewater. The treatment systems at the two plants are nearly identical in design and operation (except for design flow capacity) and include the following treatment operations in sequence: Equalization; pH adjustment (lime addition); Ferric chloride addition; Polymer addition; Clarification; Filtration; and Anoxic/anerobic biological treatment (designed to optimize removals of selenium compounds).

The FGD wastewater (i.e., FGD purge) is transferred to the equalization tank at the beginning of the wastewater treatment system. The equalization tank also receives several recycle streams from the treatment system, including sumps and backwash streams from sand filters and the bioreactor. From the equalization tank, the wastewater enters three reaction tanks in series: reaction tank 1, in which lime is added to increase the pH; reaction tank 2, designed for organosulfide addition if needed, but which is currently only used to provide additional mixing (i.e., no organosulfide is being added); and reaction tank 3, in which ferric chloride is added.

The overflow from reaction tank 3 is sent to two clarifiers operating in parallel. Each clarifier is capable of handling 100 percent of the waste stream flow, but periodically one is taken out of service for preventative maintenance. A polymer is added to the wastewater just before it enters the clarifiers. Hydrochloric acid is added to the wastewater at a point between the clarifiers and the sand filters to neutralize the pH prior to the biological treatment stage. After the hydrochloric acid addition, the overflow from the two clarifiers is split and sent to three sand filters in parallel, each of which can handle 75 percent of the waste stream capacity.

The sludge collected in the clarifiers is transferred to a sludge holding tank. A portion of the clarifier sludge is recycled from the sludge holding tank to reaction tank 1, while the majority is pumped to a filter press for dewatering. The dewatered treatment solids are sent to a landfill and the filtrate is recycled to the equalization tank.

The bioreactor portion of the treatment train consists of bioreactor cells containing activated carbon media and microbes which reduce selenium to its elemental form and precipitate other metals as sulfide complexes. The microbes also reduce the concentration of nitrogen present in the wastewater.

This report presents characterization data for certain pollutants present in the FGD wastewater at the two power plants, collected at the following sample locations: (1) FGD purge to the wastewater treatment system (i.e., untreated wastewater); (2) physical-chemical treatment effluent (i.e., an intermediate sample point in the treatment system); and (3) bioreactor effluent (i.e., final treated effluent for the FGD wastewater treatment system). The data presented in this report are self-monitoring data collected over several years of operation.

3. Overview of the Statistical Methods Used

A brief summary is given below of the analyses performed to derive long term average, variability factors, and the daily maximum and monthly average effluent limits.

First, data were thoroughly reviewed and data not appropriate for use were excluded before calculating the limits. Generally, EPA will exclude certain observations when deriving effluent limits if the data are associated with a treatment system upset or other situation not representative of a well-operated system. Other potential reasons for excluding data are instances of laboratory analysis or reporting errors; sample contamination during collection, transport, or analysis; observations associated with inappropriate analytical methods; or data entry errors. The specific data excluded from this analysis, and the reasons for their exclusion, are described below in section II.2.

Second, a summary of the data and presentation of the results are given (after the necessary data exclusions):

- Concentrations over time were plotted for the three sampling locations (FGD purge, physical-chemical effluent, and bioreactor effluent). The plots help to visualize how the concentrations change over time at each of the sampling locations. Also, the plots provide a visual comparison of how the pollutant concentrations measured at the three sampling locations relate to one another at a discrete point in time or over extended periods.
- Summary statistics were given for the three sampling locations for (i) only those observations that were quantified, (ii) the combination of quantified and non-quantified observations, and (iii) the numbers of quantified and non-quantified samples together with the sample-specific quantitation level.

Third, the appropriate statistical model was selected and used for calculating the following parameters for each plant: (i) long term average, (ii) daily variability factor, and (iii) monthly variability factor. See section III.1 for complete definitions of these parameters.

- Based on review of the data, it was determined that the modified delta-lognormal distribution should be used to model the effluent concentrations to obtain the long term averages, daily variability factor, and monthly variability factor (except for mercury collected at Allen and chromium collected at both Allen and Belews Creek). See section III for a description of the modified delta-lognormal model and why it was selected.

- Since all mercury observations at Allen were quantified, the lognormal distribution was used to obtain the long term averages and variability factors. The lognormal distribution is a special case of the modified delta-lognormal distribution used when all observations are quantified.
- For chromium, the lognormal and modified delta-lognormal models were not appropriate models for the data since there was only one effluent value above the quantitation limit for Belews Creek and all values were non-quantified for Allen. The lognormal and modified delta-lognormal models require at least two distinct values above the quantitation limit in order to estimate the variance of the distribution. In cases such as this, the long term average for each plant is defined as the arithmetic average of the data and the variability factor is not calculated. See section III.6.

Fourth, for each plant, the values for the long term average and daily and monthly variability factor were determined, assuming the data followed a lognormal or modified delta-lognormal distribution (except for chromium). For chromium, only the long term average is obtained (see the explanation in the third step above).

Finally, the combined long term average, the combined variability, and the limits are obtained using the results for the two plants as follows:

- The combined long term average is calculated as the median (which in this instance is the same as the mean since there were only two plants) of the long term averages obtained from the Allen and Belews Creek plants.
- The combined variability factor is calculated as the mean of the variability factors obtained from the Allen and Belews Creek plants.
- The limit is defined as the product of the combined long term average and the combined variability factor.

II. DESCRIPTION OF THE DATA USED TO CALCULATE THE LIMITS

The sections below provide a summary of the Allen and Belews Creek monitoring data used for calculating effluent limits. Also, this section provides the rationale for the exclusions from the monitoring data made prior to calculating the limits.

1. Summary of the Data

a. Allen Plant

Duke Energy provided sampling data collected for Allen Station at the following sampling locations: FGD purge, physical-chemical effluent (i.e., bioreactor influent), and bioreactor effluent. The data contain results for a variety of pollutants; however, this report focuses only on the following pollutants for which the limits are calculated: arsenic, chromium, copper, mercury, selenium, and zinc. Sampling data were available for these pollutants from March 03,

2009 to May 23, 2011 (except for copper, which was collected November 11, 2009 to May 23, 2011). Note that the monitoring data were not collected at a single, consistent fixed interval over the entire period. The bullets below give a brief summary by pollutant of the dates for which there are sampling results:

- For arsenic, chromium, selenium, and zinc the collection intervals varied from 1 to 49 days, with an average of 6 days (median of 2 days) between samples for all sampling locations.
- For copper, the collection intervals varied from 1 to 35 days, with an average of 9 days (median of 8 days) between samples for all sampling locations.
- For mercury, at the FGD purge sampling location, the collection intervals varied from 1 to 56 days, with an average of 6 days (median of 2 days) between samples. At the physical-chemical sampling location, the collection intervals varied from 1 to 82 days, with an average of 7 days (median of 2 days) between samples. At the bioreactor sampling location, the collection intervals varied from 1 to 49 days, with an average of 6 days (median of 2 days) between samples.

The following analytical methods were used at Allen Station to analyze the pollutants discussed in this report:

- EPA Method 200.8 (ICP-MS in conjunction with collision cell instrumentation: ICP-CRC-MS) was used to analyze the following pollutants: arsenic, chromium, copper, selenium, and zinc.
- EPA Methods 1631E and 245.1 were used to analyze mercury.

b. Belews Creek

Duke Energy provided sampling data collected for the Belews Creek Station at the following sampling locations: FGD purge, physical-chemical effluent, and bioreactor effluent. The data provided contains results for a variety of pollutants; however, this report focuses only on the following pollutants for which the limits are calculated: arsenic, chromium, copper, mercury, selenium, and zinc. Sampling data were available for these pollutants from February 06, 2008 to May 25, 2011 (except for copper which was collected December 10, 2008 to May 25, 2011). Included in the self-monitoring data are samples collected for use by plant staff to evaluate treatment system operation, and samples collected during a study conducted by the Electric Power Research Institute (EPRI). Note that the monitoring data were not collected at a single, consistent fixed interval over the entire period. The bullets below give a brief summary by pollutant of the dates for which there are sampling results:

- For arsenic, at the FGD purge, the collection intervals varied from 1 to 45 days, with an average of 7 days (median of 2 days) between samples. At the physical-chemical sampling location, the collection intervals varied from 1 to 85 days, with an average of 11 days (median of 7 days) between samples. At the bioreactor effluent sampling location, the collection intervals varied from 1 to 45 days, with an average of 6 days (median of 1 day) between samples.

- For chromium, at the FGD purge and bioreactor effluent sampling locations, the collection intervals varied from 1 to 585 days with an average of 9 days (median of 1 day) between samples. At the physical-chemical sampling location, the collection interval varied from 1 to 662 days, with an average of 16 days (median of 2 days) between samples.
- For copper, the collection intervals varied from 1 to 103 days, with an average of 19 days (median of 14 days) between samples for all three sampling locations.
- For mercury, at the FGD purge, the collection intervals varied from 1 to 45 days, with an average of 8 days (median of 4 days) between samples. At the physical-chemical sampling location, the collection intervals varied from 1 to 132 days, with an average of 8 days (median of 1 day) between samples. At the bioreactor sampling location, the collection intervals varied from 1 to 96 days, with an average of 7 days (median of 1 day) between samples.
- For selenium, the collection intervals varied from 1 to 45 days, with an average of 6 days (median of 2 days for FGD purge and 1 day for physical-chemical and bioreactor) between samples for all three sampling locations.
- For zinc, at the FGD purge and bioreactor sampling locations, the collection interval varied from 1 to 423 days, with an average of 7 days (median of 1 day) between samples. At the physical-chemical sampling location, the collection interval varied from 1 to 423 days, with an average of 12 days between samples (median of 7 days) between samples.

The following analytical methods were used at Belews Creek Station to analyze the pollutants discussed in this report:

- EPA Method 200.8 (ICP-MS in conjunction with collision cell instrumentation: ICP-CRC-MS) was used to analyze the following pollutants: arsenic, chromium, copper, selenium, and zinc.
- EPA Methods 1631E and 245.1 were used to analyze mercury.

2. Data Excluded and Rationale for the Exclusions

EPA reviewed the Allen and Belews Creek data on a pollutant-by-pollutant basis to identify data values that appeared unusual and could warrant exclusion because they do not represent optimal operation of the treatment system. For example, EPA eliminated data that were collected while a facility was experiencing treatment system upset or experiencing other atypical incidents with the treatment system. In developing effluent limits, it would also be appropriate to exclude data associated with time periods when a facility is in violation of its NPDES permit although no such instances were identified for these data.

EPA conducted an initial review of the data for both plants to identify results that appeared to be data entry errors, based on aberrations in patterns observed for the dataset as a whole. For example, certain data recorded as non-quantified measurements more closely resembled

quantified measurements in terms of quantity and/or format. Similarly, certain data reported as quantified values exhibited characteristics that more closely resembled results reported as non-quantified. EPA corrected only those results confirmed by Duke Energy as a data entry error, prior to conducting further analyses or excluding data.

Next, the general approach for identifying data warranting exclusion involved a combination of statistical screening analyses and engineering reviews, as well as reviewing potentially abnormal results with Duke Energy representatives. The statistical analyses, described below, were used to flag results that appeared unusually high or low, or otherwise deviated from general trends in the data. No results were excluded merely because they were flagged as extreme observations in the statistical analysis; instead, EPA then conducted engineering reviews of the data to assess whether they were representative of typical operation for a well-designed and well-operated treatment system. For both the flagged data and the data sets as a whole, the engineering reviews sought to identify aberrations in data patterns such as isolated extremely high concentrations surrounded by relatively low concentrations, inconsistencies between the reported values for the bioreactor effluent relative to the physical-chemical or FGD purge values, and indicators of treatment system upsets. As appropriate, questionable results were reviewed with Duke Energy to identify reasons for the reported values.

a. Excluding Data Associated with Treatment System Commissioning Period

EPA generally excludes data associated with the initial commissioning period for a new wastewater treatment system because it tends to exhibit relatively high variability not representative of typical operation. During the commissioning period, the treatment system undergoes a variety of testing to demonstrate that hydraulic flows through the system and equipment such as pumps, valves, and other equipment operate as designed. During this period, it is common for changes in process operations to occur as the treatment system undergoes performance verification testing (sometimes referred to as acceptance testing) and the operation is modified to identify approaches to optimize (i.e., minimize) operational costs or make adjustments to improve pollutant reductions. Refinements that may take place during the commissioning period include adjustments to chemical feed locations and/or feed rates; evaluating different pump cycles, filter backwash cycles, clarifier overflow rates, or sludge removal cycles; and changes to the chemicals used. Generally, this initial period of operation also serves as a time when the operators gain familiarity with the intricacies of the treatment system operation. During this acclimation and optimization period the effluent concentration values may exhibit higher variability than would typically be observed for a well-operated treatment system, producing occasional extreme values (high and low concentrations). After this initial adjustment period, a well-operated system should operate at steady state (or more precisely, a "quasi steady state") with relatively low variability around a long term average.

Since the data for Allen and Belews Creek Stations included the initial startup and commissioning period for the treatment systems, EPA needed to determine the duration of the commissioning period and to exclude these data points before the analyses were performed. The commissioning period for Allen and Belews Creek plants was determined from an

engineering perspective based on technical expertise and by looking at the longitudinal plots of all data in Appendix 1. In addition, the smooth curves superimposed on the individual data plots (see Appendix 2) were examined to help determine when the treatment systems appeared to stabilize. This smoothing technique (LOWESS) makes minimal assumptions about the distribution of the data and is meant to highlight patterns of the data (see Cleveland (1979, 1981) for more details).

After thoroughly examining effluent data from the Allen and Belews Creek plants from an engineering perspective, EPA determined that the initial six months of operation represented a reasonable estimate of the commissioning period for the treatment systems. The data associated with the commissioning periods were excluded from the data used to calculate the limits. The data excluded from each of the plants due to the commissioning period are:

- For Allen Station, data from March 03, 2009 through August 26, 2009 were excluded.
- For Belews Creek Station, data from February 6, 2008 through July 31, 2008 were excluded.

b. Data Excluded Due to Insufficiently Sensitive Analytical Method

The data analyzed by this report are self-monitoring data collected by Duke Energy to evaluate and optimize the operation of the treatment systems; they are not NPDES compliance monitoring samples. In some cases, Duke Energy analyzed the samples for mercury using EPA Method 1631E, while other samples were analyzed for mercury using EPA Method 245.1. Method 245.1 was approved by EPA in 1974 and can achieve measurements of mercury down to 200 parts per trillion (ppt), although the Method 245.1 data for both plants set the quantitation limit for mercury in the physical-chemical and bioreactor samples at 1 part per billion (ppb). EPA approved Method 1631E in 2002. This newer method has a quantitation level of 0.5 ppt, making it 400 times more sensitive than Method 245.1.

In a memorandum issued by EPA's Office of Wastewater Management, EPA directed permitting authorities to ensure that sufficiently-sensitive analytical methods are used when deciding whether to set a permit limitation for mercury and for sampling and analysis of mercury pursuant to the monitoring requirements within a NPDES permit.[1] One measure for determining whether a method for mercury is "sufficiently sensitive" is that its method quantitation level is low enough relative to the amount of mercury in a facility's discharge that the method detects and quantifies the level of mercury in the discharge. Since these self-monitoring samples were not collected for the purpose of demonstrating compliance with NPDES permit limits, Duke Energy had the discretion to choose whichever analytical method it wished to use. However, the quantitation level for Method 245.1 is not low enough to quantify the concentration of mercury present in the physical-chemical and bioreactor samples, making it inappropriate for use in calculating permit limitations. Therefore, all mercury results that

[1] James A. Hanlon, Director, Office of Wastewater Management. "Analytical Methods for Mercury in National Pollutant Discharge Elimination System (NPDES) Permits," August 23, 2007.

were analyzed using Method 245.1 for physical-chemical and bioreactor samples were excluded.

c. Data Excluded Due to Abnormal Conditions or Treatment System Upset

EPA reviewed the data for Allen and Belews Creek Stations to identify treatment system upsets or other instances of suboptimal operation that indicated the treatment system was not operating in a manner representative of typical operation for a well-designed and well-operated treatment system. This engineering evaluation uses a variety of approaches to identify apparent occurrences of atypical operation including assessments of: sudden spikes in pollutant concentrations (manifested as isolated occurrences or extended periods of elevated concentrations); instances where bioreactor effluent concentrations exceeded pollutant concentrations for the FGD purge or physical-chemical treatment stage (e.g., bioreactor effluent exceeding the bioreactor influent concentrations); departures from demonstrated data patterns; and unusual data patterns or deviations that seemingly do not follow expected treatment efficacy, based on best professional judgment and knowledge of pollutant treatability for similar wastewater treatment processes.

- Allen Station: EPA observed periodic occurrences in early 2010 of unusually high concentrations of arsenic in the bioreactor effluent relative to effluent from the physical-chemical treatment stage (i.e., bioreactor influent). According to Duke Energy personnel, the treatment system suffered an upset condition beginning March 2, 2010 that particularly impacted arsenic concentrations exiting the bioreactor. The treatment system suffered pipe failures in the bioreactor on March 2 and March 11, dislodging a portion of the carbon within the bioreactor cells. It took approximately four weeks for repairs to be completed, and additional time for the treatment system performance to stabilize. According to Duke Energy, elevated arsenic concentrations in the bioreactor effluent during the period March through May 2010 are likely attributable to the pipe breakages and related repair activities. EPA excluded all arsenic data for the Allen Station bioreactor effluent for the period collected from the date of the first pipe failure until the end of May (i.e., sample collection dates 3/8/10, 3/22/10, 4/5/10, 4/26/10, 5/10/10, and 5/25/10). EPA's review of the arsenic data also noted that the arsenic concentrations appeared to remain somewhat elevated during June 2010; however, since it is uncertain whether these values are attributable to the pipe failures, EPA did not exclude data for the samples collected in June.
- Belews Creek Station: EPA's engineering review identified unusual results reported for numerous pollutants on January 17, 2011, including arsenic, cadmium, chromium, nickel, selenium, silver, and zinc. For example, results for the physical-chemical and bioreactor effluent sampling locations for these parameters were reported as non-quantified, but at significantly elevated levels relative to other results in the dataset. These results indicate that the laboratory experienced difficulties while analyzing the samples, or perhaps mishandled the samples during analysis. In response to EPA's queries, Duke Energy stated that there appeared to be a quality control issue that had required higher dilution for 1/17/11 samples. Since these results indicated an abnormal

situation was occurring either in the laboratory analyses, or perhaps with the treatment system, EPA excluded the data for all pollutants for samples collected on 1/17/11.

d. Exclusions of Extreme Observations

After making the exclusions noted in parts a-c above, the remaining data were examined to check for extreme observations before the summary statistics, LTAs, and variability factors were calculated. The extreme observations are those observations with unusually high or low values compared to all other observations in the data. These extreme observations may have occurred for different reasons, for example, due to a system upset, contamination of samples during collection, errors during laboratory analyses, or data entry errors. Such observations need to be excluded before the analyses are performed since they do not represent the typical values that are seen under normal operating conditions. After excluding these observations, the statistical analysis results will not be distorted by unrealistic values.[2]

The extreme observations were identified by (i) an engineering assessment of the individual observations; (ii) reviewing the longitudinal plots for apparent aberrations, and (iii) use of a statistical method to systematically detect extreme observations. This third approach was used as a tool to supplement the first two approaches to make sure that all extreme observations were accounted for. Note that only data from physical-chemical and bioreactor effluent (but not FGD purge) sampling locations were subjected to this screening process because the purge from FGD systems is known to be highly variable and the large concentrations of suspended solids typically present in the FGD purge (on the order of 0.5 to 5 percent total suspended solids) exacerbates this variability. Although boxplots of the FGD purge were not created, EPA did review the FGD purge data (both longitudinal plots and individual concentration values) in conjunction with engineering reviews of the physical-chemical and bioreactor effluent data.

The statistical method used to detect extreme observations was the boxplot (Box-and-Whisker plot). This outlier test was applied to the set of all quantified observations. For the boxplot method, the observations that are considered as potential outliers are those observations with values lying outside the ends of the vertical lines (or whiskers) that extend above and below the box. Thus, an observation is considered to be a potential outlier if its value is either greater than Q3 + 1.5*(Q3 –Q1) or less than Q1 – 1.5*(Q3 – Q1), where Q1 is the 25th percentile of the data (i.e., the value such that 25% of the data are below it) and Q3 is the 75th percentile of the data. This method for identifying potential outliers provides a reasonable and objective approach since it does not require any assumption about the distribution of the data (e.g., does not require that the data be normally distributed). Because no assumption is made about the distribution of the data, the method can be used in a wide variety of cases that may have different distributions (e.g., for examining a number of different pollutants).

[2] The 5th Circuit Court determined that EPA could reasonably exclude an isolated and extremely high discharge "because the EPA has considerable expertise in interpreting sampling data. . . ." See Chemical Mfrs. Ass'n v. U.S.EPA, 885 F.2d 253, 264 (5th Cir. 1989).

Appendix 3 of this report contains a list of all the observations that were deemed outliers (i.e., extreme values) at the physical-chemical and bioreactor sampling locations. The appendix also identifies which of the outlying observation were actually excluded. Recognizing that some degree of variability around a long-term average is inherent in wastewater discharges, even at well-operated facilities, EPA has used an abundance of caution when excluding reported results. The statistical analyses such as the boxplot were used to flag results that appeared unusually high or low, or otherwise deviated from general trends in the data, but no results were excluded merely because they were flagged as extreme observations by the boxplots or the longitudinal plots. Each instance of extreme observation received careful engineering review to assess whether the observation was consistent with patterns EPA has observed for other data from these and other facilities in the steam electric power generating industry, as well as patterns observed for similar wastewaters or treatment technologies in other industrial sectors. In addition, each of the extreme observations that EPA ultimately excluded were reviewed with Duke Energy personnel to determine reasons for the unusual results. This approach toward data review resulted in EPA retaining most of the extreme observations, and excluding only those observations that conclusively were not representative of typical operation for a well-designed and well-operated treatment system.

- Belews Creek Station: The boxplots flagged the October 5, 2009 mercury results for the physical-chemical and bioreactor effluent sampling points as extreme observations. EPA's engineering review noted that the bioreactor effluent results for this day were 1-2 orders of magnitude higher than any other bioreactor effluent result collected in the six months preceding or following the October 5 sample. Similar results were observed for the physical-chemical sampling location. The mercury concentrations in the FGD purge on October 5 were not atypical. Duke Energy has suggested that the elevated mercury concentrations are due to the plant switching to a different coal source. However, Belews Creek Station obtains coal from approximately 100 sources and this affect on results for the physical-chemical and bioreactor sampling points is not observed at any other time during the several years of operation for the facility, including during the initial commissioning period for the treatment system. Further, although the concentration of dissolved mercury in the FGD purge on 10/5/09 is relatively high compared to most other samples collected in 2009, the October sample is only approximately three times higher than the 8/3/09 sample and it is approximately half the concentration in the 11/2/09 sample; yet, the samples collected on those dates for the physical-chemical and bioreactor sampling locations contained substantially lower concentrations of mercury than observed in the October sample. Therefore, EPA excluded the October 5, 2009 mercury results for the physical-chemical and bioreactor effluent sampling points as extreme observations. In addition, EPA concludes that even if the mercury truly was present at these levels in October (and the reported values are not due to sample contamination or laboratory error), then it indicates that the treatment system was overloaded and not performing to its fullest capabilities. The treatment system is designed with the capability to add organosulfide to optimize the precipitation and removal of mercury; however, Belews Creek does not currently do so because they are able to comply with NPDES permit limits without the chemical addition. Thus, if high concentrations of dissolved mercury are present in the FGD

purge, much of the dissolved mercury would likely pass through the physical-chemical stage untreated and potentially overload the bioreactor. Based on information EPA has reviewed regarding the treatment of FGD wastewater, the addition of organosulfide (which did not occur) likely would have effectively reduced the mercury concentration in the bioreactor effluent.

- Belews Creek Station: EPA's review of the boxplots and longitudinal plots noted extreme observations for the period May through October 2010, during which most results for mercury in the bioreactor effluent were higher than the levels in the bioreactor influent (i.e., physical-chemical sample location). Such results are inconsistent with the pollutant removal efficacy for the bioreactor, based on EPA's expertise evaluating treatment technologies for FGD wastewater from power plants and treatment technologies for wastewaters from other industry sectors, and it is also contrary to the long-term performance for the bioreactor at Belews Creek and bioreactors of similar design at other power plants. EPA also identified an extreme observation for 10/7/10 for the physical-chemical effluent, where the mercury concentration spiked upward by an order of magnitude. Due to the unusual results, the following observations were excluded as extreme observations: bioreactor effluent mercury on 5/26/10, 6/9/10, 8/11/10, 9/8/10, 10/7/10; and physical-chemical effluent mercury on 10/7/10.
- Belews Creek Station: The bioreactor effluent selenium result for July 14, 2010 was identified as an extreme observation. This observation is much higher than observed for the remainder of the dataset, by 1 to 2 orders of magnitude. EPA evaluated the selenium contributions in the FGD purge and bioreactor influent (total and dissolved selenium, as well as concentrations of selenate and selenite), and the values for total selenium, selenite, and selenate in the bioreactor effluent (there are no dissolved selenium results for the bioreactor effluent on that date). In addition, EPA evaluated the concentrations for total selenium, dissolved selenium, selenite, and selenate at all three sampling locations for the months preceding and following the July 14 sampling event. Duke Energy personnel reported no indication of abnormal events or activities at time of sample collection. They were unable to explain the result and noted that samples of the ash basin (which receives treated FGD wastewater) collected around the same time period did not exhibit a similar increase in selenium concentrations. Even taking into account the concentration of selenate present in the FGD purge and physical-chemical samples, the bioreactor effluent is substantially higher than would be expected based on the demonstrated performance of the bioreactors at both Allen and Belews Creek Stations. For example, on July 28 the FGD purge contained even higher concentrations of selenium and the physical-chemical effluent contained comparable levels of selenium, yet on that date the bioreactor effluent concentration was an order of magnitude lower. Since no plausible explanation validating the extreme observation could be determined, EPA excluded the bioreactor effluent selenium observation for July 14, 2010.

III. STATISTICAL METHODOLOGY

The sections below provide the following: (i) the definitions and background for the long term average, daily variability factor, and monthly variability factor; (ii) a brief overview of the modified delta-lognormal distribution used to model the effluent data collected from the Allen and Belews Creek plants; (iii) adjustment for autocorrelation, if data are deemed correlated; (iv) the calculation of the long term average (LTA), daily variability factor, and monthly variability factor (assuming four samples per month will be collected) assuming the data follow a modified delta-lognormal distribution (or lognormal distribution if all observations are quantified); (v) data requirements in the calculation of the variability factors; and (vi) the calculation of the combined long term average, combined variability factors, and daily and monthly limits using data from the Allen and Belews Creek plants. Also, this section describes how the limit was set in a special case where nearly all observations were non-quantified.

1. Definitions and Background

In developing the effluent limitations, a statistical procedure is used that involves fitting effluent data to statistical distribution models and estimating specified upper percentiles of these fitted distributions. The bullets below describe in detail the important quantities that are typically estimated based on the effluent data.

- Long term average (LTA) is the average concentration that is achieved over a period of time. The long term average is the mean of the underlying statistical distribution of the daily effluent values.
- A daily maximum limitation is the highest allowable discharge in any one day. The daily maximum limitation is the estimate of the 99^{th} percentile (P_{99}) of the distribution of the daily effluent values.
- The daily variability factor is the ratio of the 99^{th} percentile to the LTA. This ratio represents the relationship between the average treatment performance level and large values that a well-designed and well-operated treatment system should be capable of achieving at all times.
- A monthly average limitation is the highest allowable average of discharges calculated from the effluent data collected over a calendar month (or period of time specified in the permit). The monthly average limitation is the estimate of the 95^{th} percentile (P_{95}) of the distribution of the monthly averages of the daily effluent values.
- A monthly variability factor is the ratio of the 95^{th} percentile of the distribution of the monthly averages of the daily effluent values to the long term average.

2. Statistical Model Selected for the Data

After examining the data for each of the pollutants from the Allen and Belews Creek plants, the approaches used to calculate the long term averages and variability factors were as follows:

- For the majority of the pollutants for these two plants, the modified delta-lognormal distribution was selected to model the effluent concentrations to obtain the long term averages and variability factors. The modified delta-lognormal distribution is appropriate for these pollutants since the effluent data contained both quantified and non-quantified values. Also, the modified delta-lognormal distribution allows for non-quantified values that occur at multiple sample-specific quantitation limits. Note that when all values are quantified, then the modified delta-lognormal distribution reduces to the lognormal distribution. This distribution was used for mercury at Allen since all values were quantified.
- For chromium where the modified delta-lognormal distribution was determined not applicable, (since all values were below quantitation limits for Allen, and only one value above the quantitation limit for Belews Creek), a different method was used to set the limit. This method consisted of setting the daily maximum limitation at the highest-reported quantitation limit. The long term average is calculated as the arithmetic average of all observations (i.e., all sample specific quantitation limits). Further, no variability factors or monthly limits are calculated (see sections 6 and 7 below for further discussion).

3. Details of the Modified Delta-lognormal Distribution

The modified delta-lognormal distribution is a modification of the "delta distribution" originally developed by Aitchison and Brown (1969). The delta-lognormal distribution is a distribution that consists of a mixture of a continuous density portion with a discrete-valued spike at zero. EPA modified the distribution by replacing a single spike located at zero with a discrete distribution made up of multiple spikes to account for the multiple distinct reported quantitation limits, which typically is present in effluent data. It is worth noting that when all values are quantified, then the modified delta-lognormal distribution is identical to the lognormal distribution. See Kahn and Rubin (1989) for technical details of the lognormal distribution.

If U is a random variable having a modified delta-lognormal distribution, then the cumulative probability function of U (the probability that U can hold value less than or equal to some specified value of u) takes the following form:

$$P[U \le u] = \begin{cases} \sum_{i:D_i<u} \delta_i + (1-\delta)\Phi[(\log(u)-\mu)/\sigma] & \text{if } u < D_k \\ \delta + (1-\delta)\Phi[(\log(u)-\mu)/\sigma] & \text{if } u \ge D_k \end{cases}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and D_k denotes the largest of the k observed quantitation limits among the non-quantified values.

The expected value and variance of *U* are as follows:

$$E[U] = \sum_{i=1}^{k} \delta_i D_i + (1-\delta)\exp(\mu + 0.5\sigma^2)$$

$$Var[U] = \frac{1}{\delta}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\delta_i\delta_j(D_i - D_j)^2 + (1-\delta)\exp(2\mu+\sigma^2)(\exp(\sigma^2)-1) + \delta(1-\delta)\left[\frac{1}{\delta}\sum_{i=1}^{k}\delta_i D_i - \exp(\mu+0.5\sigma^2)\right]^2$$

The parameters in the mean and variance expressions above, i.e., δ_i, δ, μ, and σ^2, can be estimated from the data as follows:

$$\hat{\delta}_i = \frac{1}{n}\sum_{j=1}^{n_d} I(d_j = D_i)$$

$$\hat{\delta} = \frac{n_d}{n}$$

$$\hat{\mu} = \frac{1}{n_c}\sum_{i=1}^{n_c}\log(x_i)$$

$$\hat{\sigma}^2 = \frac{1}{n_c - 1}\sum_{i=1}^{n_c}(\log(x_i) - \hat{\mu})^2$$

where,

δ_i is the probability of a non-quantified value at the quantitation limit D_i, i = 1,...,k;

d_j is the quantitation limit associated with the j^{th} non-quantified values (j = 1,...,n_d);

n_d is the number of non-quantified observations;

δ is the probability of observing a non-quantified observation;

n is the number of effluent observations (including both quantified and non-quantified values);

μ and σ^2 are the parameters of the lognormal distribution;

x_i is the i^{th} quantified effluent value, i = 1,...,n_c;

Note that the "hats" over each of the parameters discussed above denote estimated quantities.

4. Autocorrelation Analysis

Effluent concentrations that are collected over time may be autocorrelated. When the data are said to be positively autocorrelated, it means that measurements taken at specific time intervals (such as 1 day or 2 days apart) are similar. For example, positive autocorrelation is

present in the data if the effluent concentration was relatively high one day and was likely to remain high on the next and possibly succeeding days. When the data are deemed autocorrelated, the variance estimates for the data need to be adjusted. The variance estimate is adjusted as follows:

$$\hat{\sigma}_A^2 = \frac{\hat{\sigma}^2}{g(R)}$$

where $g(R) = 1 - \left(\frac{2}{n(n-1)}\right)\left(\frac{\rho_1}{1-\rho_1}\right)\left((n-1) - \frac{\rho_1(1-\rho_1^{n-1})}{1-\rho_1}\right)$, and ρ_1 is the correlation of the natural logarithm of measurements from successive sampling events. Note that if the daily values are independent (autocorrelation is not present in the data), then *g(R)* = 1 and the variance estimate is unadjusted. This adjustment for the correlation in the data produces an unbiased estimate of the variance. An autoregressive lag-1 time series (AR(1)) model was used to obtain ρ_1. The AR(1) model assumes that the measurements from successive samples are equally spaced.

EPA examined autocorrelation and partial autocorrelation plots to determine whether it was necessary to adjust for correlation when setting the limits.

In order to minimize the bias associated with analyzing short time series for autocorrelation, EPA looked for autocorrelation in data sets that had at least 25 quantified observations. In this instance, there were sufficient data at either plant for the following analytes: arsenic at Allen, mercury at Allen, selenium at Allen, mercury at Belews Creek, and selenium at Belews Creek. The highest number of quantified readings in any other data set was five. In order to deal with differing time gaps that existed between collected samples in each of the analyte data sets, EPA inserted null values to make the time series consistent. Since the samples were collected approximately on a weekly basis, the null values were inserted at weekly intervals. This allows PROC ARIMA in SAS to treat observations with large gaps between sample collection dates differently than those with smaller gaps.

An autocorrelation function (ACF) plot shows how values in a time series are correlated with past values of the series. ACF plots show the degree of correlation with past values in the series as a function of the number of periods in the past (i.e., the lag) at which the correlation is computed. If an autoregressive model is appropriate for a set of data, then the ACF plot should show a gradually decreasing correlation for each successive lag.

However, based on the data for the five analyte/plant combinations mentioned above, none of the ACF plots suggested that an autocorrelation model would be appropriate. The plots (to the first ten lags) can be found in Appendix 6.

EPA also examined partial autocorrelation function (PACF) plots for evidence of a need to model the data using an autoregressive (AR) model. The partial autocorrelation of lag-*k* is the

correlation between an observation and the observation recorded *k* times before it that is not accounted for by lags *1* to *k-1*. An AR model is considered AR(*k*) if the partial autocorrelation of lag-*k* is significantly large and all other partial autocorrelations of higher lag are sufficiently close to zero. A PACF plot displays these correlations and aids in the selection of an appropriate AR model.

However, based on the log-transformed data for the five analyte/plant combinations mentioned above, none of the PACF plots suggested that an autocorrelation model would be appropriate. The plots (to the first ten lags) can be found in Appendix 6.

In summary, after examining the ACF and PACF plots described above (and given in Appendix 6), EPA found no evidence that an autocorrelation model would be appropriate for the data being used to set the limits. As such, EPA did not incorporate an autocorrelation factor in the calculation of the numeric limitations.

5. Calculation of the Long Term Average (LTA), Daily Variability Factors, and Monthly Variability Factors Using the Modified Delta-lognormal Distribution

Below describes how the long term average and variability factors were obtained based on the assumption that the data follow a modified delta-lognormal distribution (or lognormal distribution if all observations are quantified). Note that σ^2 is used since the analysis is not adjusted for autocorrelation.

a. Calculating the Long Term Average (LTA)

The long term average for each facility is the mean of the modified delta-lognormal distribution. The long term average is calculated as follows:

$$\text{LTA} = \hat{E}[U] = \sum_{i=1}^{k} \hat{\delta}_i D_i + (1-\hat{\delta}) \exp(\hat{\mu} + 0.5\hat{\sigma}^2)$$

In the special case where all values quantified, then the LTA is found under the assumption of a lognormal distribution where

$$\text{LTA} = \hat{E}[U] = \exp(\hat{\mu} + 0.5\hat{\sigma}^2)$$

In the case where there are less than two distinct quantified values then the variance in the above formula cannot be calculated. Thus, the long term average is calculated as the arithmetic mean.

b. Calculating Daily Variability Factor (VF_{daily})

The daily variability factor was calculated separately for each parameter and each facility. The daily variability factor for each facility is the ratio of the 99th percentile to the long term average. The daily variability factor is calculated as follows:

$$VF_{daily} = \frac{\hat{P}_{99}}{\hat{E}[U]} = \frac{\hat{P}_{99}}{LTA}$$

Below describes how the 99th percentile of the modified delta-lognormal distribution is estimated. The following steps also illustrate how multiple quantitation limits are accounted for when estimating the 99th percentile.

Under the modified delta-lognormal distribution, if $D_1 < D_2 < \ldots < D_k$ are the k observed quantitation limits expressed in increasing order, then for $j = 1, \ldots, k$, let

$$\hat{q}_j = \sum_{i=1}^{j} \hat{\delta}_i + (1-\hat{\delta})\Phi\left(\frac{\log(D_j) - \hat{\mu}}{\hat{\sigma}}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. If all k values of $\hat{q}_j$ are below 0.99, then

$$\hat{P}_{99} = \exp\left(\hat{\mu} + \hat{\sigma} \cdot \Phi^{-1}\left(\frac{0.99 - \hat{\delta}}{1-\hat{\delta}}\right)\right)$$

where $\Phi^{-1}(p)$ is the p^{th} percentile of the standard normal distribution (i.e., the inverse of the standard normal cumulative distribution function). Otherwise, let D_m denote the smallest quantitation limit for which $\hat{q}_m \geq 0.99$, and let $q^* = \hat{q}_m - \hat{\delta}_m$. Then the 99th percentile is found by the following:

$$\hat{P}_{99} = \begin{cases} D_m & \text{if } \hat{q}^* < 0.99 \\ \exp\left(\hat{\mu} + \hat{\sigma} \cdot \Phi^{-1}\left(\dfrac{0.99 - \sum_{i=1}^{m-1} \hat{\delta}_i}{1-\hat{\delta}}\right)\right) & \text{if } \hat{q}^* \geq 0.99 \end{cases}$$

In the special case where all observations are quantified, the 99th percentile is found under the assumption of a lognormal distribution, where

$$\hat{P}_{99} = \exp\left(\hat{\mu} + 2.326\hat{\sigma}\right)$$

c. Calculating the Monthly Variability Factor (Assuming Four Samples Will be Collected)

The monthly variability factor was calculated separately for each parameter and each facility. The monthly variability factor for each facility is the ratio of a 95th percentile to the long term average. In this case, the percentile EPA seeks is the 95th percentile of the distribution of $\overline{U}_4$, which represents the average of four samples for a given facility, which would be the case for a facility that monitors its effluent for the limited parameters on a weekly basis. The monthly variability factor is calculated as follows:

$$VF_4 = \frac{\hat{P}_{95}}{\hat{E}[U]}$$

Below describes how the 95th percentile is estimated assuming $\overline{U}_4$ has a modified delta-lognormal distribution. The following steps also show how multiple quantitation limits (for non-quantified values) were accounted for when estimating the 95th percentile.

First, it is important to describe the discrete component of the distribution of $\overline{U}_4$ that corresponds to its possible non-quantified values, or scenarios in which all four sample measurements in $\overline{U}_4$ are non-quantified. The discrete component of this distribution will depend on the *k* distinct quantitation limits associated with the observed daily measurements. It is assumed that each of the four daily measurements in $\overline{U}_4$ can be linked to any of the *k* quantitation limits with equal probability. When $\overline{U}_4$ is labeled as non-quantified, it can hold one of *k** possible values. For example, if *k* = 2, there are *k**=5 possible non-quantified values for $\overline{U}_4$. These *k** possible values correspond to different combinations of the two quantitation limits (D_1 and D_2) which can occur for any of the four values being considered. These five possible values are denoted by D_i^* (*I* = 1, ..., 5), and each has a corresponding probability of occurrence (denoted by δ_i^*). The following table summarizes these quantities for the case of *k* = 2:

I	Value of $\overline{U}_4$ (D_i^*)	Probability of occurrence (δ_i^*)
1	D_1	δ_1^4
2	$(3D_1 + D_2)/4$	$4\delta_1^3\delta_2$
3	$(2D_1 + 2D_2)/4$	$6\delta_1^2\delta_2^2$
4	$(D_1 + 3D_2)/4$	$4\delta_1\delta_2^3$
5 (=*k**)	D_2	δ_2^4

A similar table could be constructed for other values of k. When $k \geq 3$, the D_i^* values will be weighted averages of the k quantitation limits ($D_1 < D_2 < \ldots < D_k$), and the δ_i^* values will depend on δ_i (I = 1,...,k) and will be based on a multinomial distribution. Note that if $k = 3$, then k^* would be 15; if $k = 4$, then k^* would be 35; and if $k = 5$, then k^* would be 70.

The remaining approach to estimating P_{95} is similar to the approach used to estimate P_{99} in the calculation of one-day variability factors, as described in section c above. For $j = 1, \ldots, k^*$, let

$$\hat{q}_j = \sum_{i=1}^{j} \hat{\delta}_i^* + (1-\hat{\delta}^4)\Phi\left(\frac{\log(D_j^*) - \hat{\mu}_4}{\hat{\sigma}_4}\right)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and μ_4 and σ_4^2 are defined as follows:

$$\mu_4 = \log\left[\frac{(1-\delta^3)\sum_{i=1}^{k}\delta_i D_i + (1-\delta)\exp(\mu + 0.5\sigma^2)}{(1-\delta^4)}\right] - 0.5\sigma_4^2$$

$$\sigma_4^2 = \log\left[1 + \frac{1}{4\eta^2}\left[(1-\delta^4)Var[U] - (1-\delta^4)\delta^2\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\delta_i\delta_j(D_i - D_j)^2 - 4\delta^2\left(\sum_{i=1}^{k}\delta_i D_i(1-\delta^4) - \delta\eta\right)^2\right]\right]$$

where $\eta = (1-\delta^3)\sum_{i=1}^{k}\delta_i D_i + (1-\delta)\exp(\mu + 0.5\sigma^2)$.

Now, if all k values of $\hat{q}_j$ defined above are less than 0.95, then the 95th percentile is defined as

$$\hat{P}_{95} = \exp\left(\hat{\mu}_4 + \hat{\sigma}_4 \cdot \Phi^{-1}\left(\frac{0.95 - \hat{\delta}^4}{1-\hat{\delta}^4}\right)\right),$$

where $\Phi^{-1}(p)$ is the p^{th} percentile of the standard normal distribution (i.e., the inverse of the standard normal cumulative distribution function). Otherwise, let D_m^* denote the smallest of the k^* values of D_i^* for which $\hat{q}_m \geq 0.95$, and let $\hat{q}^* = \hat{q}_m - \hat{\delta}_m^*$. Then, the 95th percentile is defined by the following:

$$\hat{P}_{95} = \begin{cases} D_m^* & \text{if } \hat{q}^* < 0.95 \\ \exp\left(\hat{\mu}_4 + \hat{\sigma}_4 \cdot \Phi^{-1}\left(\dfrac{0.95 - \sum_{i=1}^{m-1}\hat{\delta}_i^*}{1-\hat{\delta}^4}\right)\right) & \text{if } \hat{q}^* \geq 0.95 \end{cases}$$

6. Data Required to Calculate Variability Factors

The lognormal or modified delta-lognormal model requires at least two distinct quantified values in order to estimate the variance of the distribution (i.e., to allow the variability factor to be calculated). Generally, EPA has been reluctant to estimate a variability factor from plants with a small number (at least two) of quantified observations of effluent concentrations. The minimum number of observations required to calculate the variability factor has been evaluated on case-by-case basis. For example, for Organic Chemicals, Plastics and Synthetic Fibers (OCPSF) rulemaking, a variability factor was estimated for a plant if it had at least 7 daily observations with at least 3 observations above the quantitation limit. In the Iron and Steel rulemaking, a variability factor was estimated for a plant if it had at least three observations, at least two of which were above the quantitation limits.

After examining the data for the pollutants for the Allen and Belews Creek plants, it was determined that the number of quantified observations and overall number of observations (quantified and non-quantified) for all the pollutants were adequate to estimate the variability factors, except for chromium. Variability factors were not calculated for chromium because all values were below the quantitation limits for Allen and only one value was above the quantitation limit for Belews Creek.

7. Calculating the Combined Long Term Average (LTA), Combined Variability Factors, and Effluent Limits Using Data from Multiple Plants

In general, when setting effluent limitations, EPA may use data from a single plant, if it is deemed appropriate. In other cases, EPA may use data from multiple plants to set the limitations. When there are multiple plants, EPA typically calculates the combined long term average, combined variability factor, and limits in the following manner.

- A long term average is obtained for each plant.
- Both daily and monthly variability factors are obtained for each plant.
- The combined long term average is defined as the median of the long term averages calculated for the multiple plants.
- The combined daily (monthly) variability factor is defined as the mean of the daily (monthly) variability factors calculated for the multiple plants.
- The effluent limit is defined as the product of the combined long term average and the combined variability factor.

Special Case – Effluent Limitations Based on a Quantitation Limit

Although the percentile estimates described in section III.5 play an important role in determining daily maximum and monthly average effluent limitations, they are not a requirement under the Clean Water Act and are not always used as the basis for effluent

limitations. In situations when there is a preponderance of non-quantified results and few, if any, quantified results, the statistical models may not be appropriate for calculating effluent limits. In such instances EPA has instead established effluent limits based on a quantitation limit. The data for chromium exhibit such characteristics, with all 64 bioreactor effluent samples at Allen Station reported as non-quantified, and all but one bioreactor effluent sample at Belews Creek (31 out of 32 samples) reported as non-quantified. The LTAs for each plant were calculated as the arithmetic average of all observations, using the sample-specific quantitation limit for the non-quantified results. The effluent limit is based on the highest reported sample-specific quantitiation level reported for the observations at both plants (i.e., 10 ug/L). It is possible that a more stringent effluent limit is technologically achievable, and it should be noted that the only quantified observation (3.6 ug/L) and the calculated LTAs (3.3 and 4.1 ug/L) are lower than the highest-reported quantitation limit. However, EPA is unable to determine, based on the available data, whether more stringent limitations would be achievable. Since the effluent limit is based on the highest quantitation limit for the reported observations, the limit is established as the daily maximum limit and no monthly average effluent limit is established. The purpose of a monthly average limitation is to require continuous dischargers to provide better control, on a monthly basis, than required by the daily maximum limitation. However, for chromium, EPA has determined that it would not be appropriate to establish a monthly average limitation lower than the highest quantitiation limit in the observations.

IV. DATA PLOTS, SUMMARY STATISTICS, LONG TERM AVERAGE (LTA), AND VARIABILITY FACTORS FOR ALLEN AND BELEWS CREEK PLANTS

The sections below provide the following for each of the pollutants at the Allen and Belews Creek plants: longitudinal plots of the data, summary statistics, long term averages, and variability factors. The bullet points below provide some important points to keep in mind about the data when considering the results presented in this section.

- The analysis is based on only the selected data used to calculate the limits (i.e., after excluding certain data, as explained in section II.2).
- The monitoring data were not collected according to a fixed interval or schedule over the entire period.
- For all values reported as non-quantified, their values are set equal to the quantitation limits and identified as non-quantified (i.e., censored).
- The long term averages and variability factors were derived assuming the data follow a modified delta-lognormal distribution, except for (i) mercury at Allen where all observations were quantified, and (ii) chromium (see description in the next bullet). For mercury at Allen, the lognormal distribution was assumed as this is a special case of the modified delta-lognormal distribution when all observations are quantified.
- For chromium, since all the values were non-quantified for Allen and only one value was above the quantitation limit for Belews Creek, the daily maximum limit was set equal to the highest reported sample-specific quantitation limit. Although not used to establish the effluent limit, long term averages were calculated as an arithmetic average of all

observations. Variability factors were not calculated for chromium because the modified delta-lognormal model requires at least two distinct values above the quantitation limit in order to estimate the variance of the distribution.

- No autocorrelation adjustment was used in calculating the limits. Observations were assumed to be independent.

The explanation for the plots and tables presented below will be given for arsenic only. The plots and tables for the other pollutants can be interpreted in a manner similar to that of arsenic.

1. Arsenic

a. Plots of Concentrations Over Time

Below are the longitudinal plots of the arsenic concentrations from the Allen and Belews Creek plants collected at FGD purge, physical-chemical treatment, and bioreactor effluent sampling locations. The plot on the left shows the concentrations for the Allen plant from September 9, 2009 to May 23, 2011. The plot on the right shows the concentrations for the Belews Creek plant from August 1, 2008 to May 25, 2011. Plots on the upper row show the concentrations collected all three sampling locations, while the plots on the bottom row show the concentrations only at the physical-chemical and bioreactor sampling locations. The tick-marks on the plots reflect sample collection dates. The quantified values are plotted as circles, and the non-quantified values are plotted as plus (+) symbols.

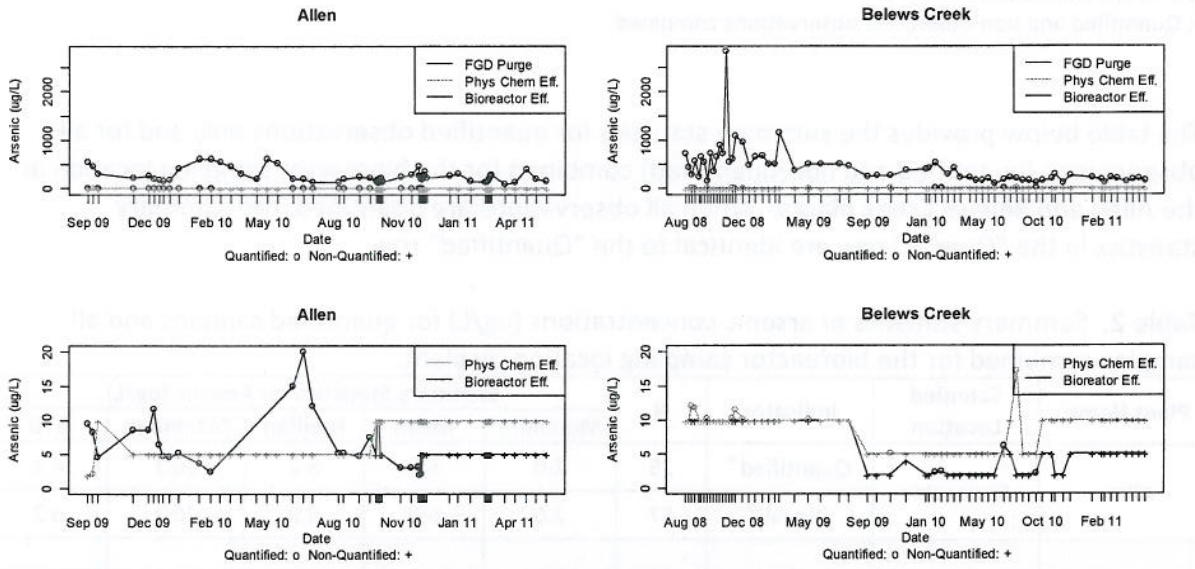

Figure 1. Plot of longitudinal data for arsenic (ug/L) for the Allen and Belews Creek plants after all necessary exclusions.

b. Summary Statistics

The table below provides the summary statistics for quantified observations only and for all observations (quantified and non-quantified) combined for at the FGD purge and physical-chemical sampling locations for the Allen and Belews Creek plants.

It is important to note that when all observations were quantified, then the summary statistics in the "Overall" row will be identical to the "Quantified" row.

Table 1. Summary statistics of arsenic concentrations (ug/L) for quantified samples and all samples combined for FGD purge and physical-chemical effluent locations by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Arsenic (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	FGD Purge	Quantified [a]	64	105.0	260.1	218.0	619.0	132.1
		Overall [b]	64	105.0	260.1	218.0	619.0	132.1
	Phys Chem Eff.	Quantified [a]	1	2.1	2.1	2.1	2.1	.
		Overall [b]	64	2.0	8.0	10.0	10.0	2.6
Belews Creek	FGD Purge	Quantified [a]	64	96.4	427.8	291.5	2,820.0	389.6
		Overall [b]	64	96.4	427.8	291.5	2,820.0	389.6
	Phys Chem Eff.	Quantified [a]	9	5.1	10.9	10.3	17.1	3.1
		Overall [b]	65	5.0	8.7	10.0	17.1	2.6

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

The table below provides the summary statistics for quantified observations only and for all observations (quantified and non-quantified) combined for the bioreactor sampling location at the Allen and Belews Creek plants. When all observations are quantified, the summary statistics in the "Overall" row are identical to the "Quantified" row.

Table 2. Summary statistics of arsenic concentrations (ug/L) for quantified samples and all samples combined for the bioreactor sampling location by plant.

Plant Name	Sampled Location	Indicator	N	Summary Statistics for Arsenic (ug/L)				
				Minimum	Mean	Median	Maximum	STD
Allen	Bioreactor	Quantified [a]	25	2.0	6.6	5.2	20.1	4.4
		Overall [b]	57	2.0	6.0	5.0	20.1	3.2
Belews Creek	Bioreactor	Quantified [a]	4	2.2	3.3	2.4	6.1	1.9
		Overall [b]	65	2.0	6.3	5.0	10.0	3.5

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

The table below provides summary statistics for the numbers of quantified and non-quantified samples together with the sample-specific quantitation limits for the Allen and Belews Creek plants. For example, for the Allen plant shown in Table 3 below, there were 25 quantified and 32 non-quantified observations. Of those 32 non-quantified observations, 28 were non-quantified at the quantitation limit of 5 ug/L and 4 were non-quantified at the quantitation limit of 10 ug/L.

Table 3. Number of quantified, non-quantified, and sample-specific quantitation limits for arsenic by plant.

Plant Name	Sampled Location	Indicator	# of Observations	Quantitation Limits for Arsenic (ug/L)			
				2	4	5	10
Allen	Bioreactor	Quantified	25	-	-		-
		Non-Quantified	32	0	0	28	4
Belews Creek	Bioreactor	Quantified	4	-	-		
		Non-Quantified	61	18	1	13	29

c. Long Term Average, Daily Variability Factor, and Monthly Variability Factor

The table below provides the long term average, daily variability factor, and monthly variability for Allen and Belews Creek. These values were estimated assuming the data follow a modified delta-lognormal distribution. The calculations for the monthly variability factor assume that samples will be collected four times per month.

Table 4. Long term average and variability factors at bioreactor sampling location for arsenic (ug/L) by plant.

Plant Name	Sampled Location	N[a]	Long Term Average	Daily Variability Factor		Monthly Variability Factor
Allen	Bioreactor	57	6.05	3.08		1.50
Belews Creek		65	6.28	1.59		1.06

[a]: Quantified and non-quantified observations combined.

2. Chromium

a. Plots of Concentrations Over Time

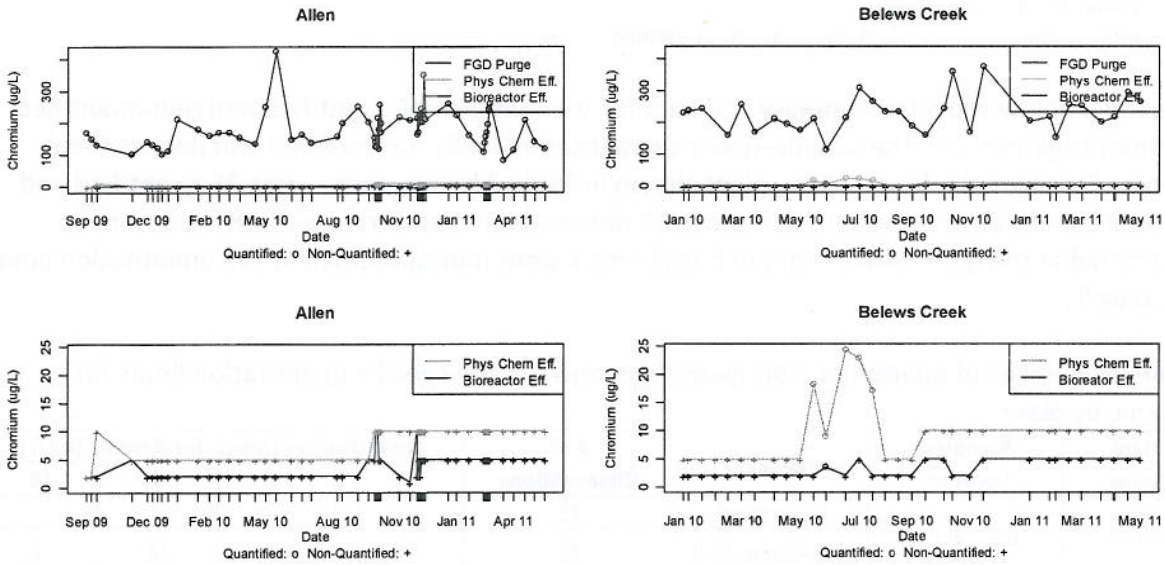

Figure 2. Plot of longitudinal data for chromium (ug/L) for the Allen and Belews Creek plants after all necessary exclusions.

b. Summary Statistics

Table 5. Summary statistics of chromium concentrations (ug/L) for quantified samples and all samples combined for FGD purge and physical-chemical effluent sampling locations by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Chromium (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	FGD Purge	Quantified [a]	64	80.8	180.4	168.0	422.0	59.1
		Overall [b]	64	80.8	180.4	168.0	422.0	59.1
	Phys Chem Eff.	Quantified [a]	0	-	-	-	-	-
		Overall [b]	64	2.0	8.0	10.0	10.0	2.6
Belews Creek	FGD Purge	Quantified [a]	31	113.0	225.4	218.0	374.0	56.9
		Overall [b]	31	113.0	225.4	218.0	374.0	56.9
	Phys Chem Eff.	Quantified [a]	5	8.9	18.3	18.2	24.4	6.1
		Overall [b]	32	5.0	9.3	10.0	24.4	5.1

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 6. Summary statistics of chromium concentrations (ug/L) for quantified samples and all samples combined for bioreactor effluent sampling location by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Chromium (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	Bioreactor	Quantified [a]	0	-	-	-	-	-
		Overall [b]	64	1.0	4.1	5.0	10.0	2.3
Belews Creek	Bioreactor	Quantified [a]	1	3.6	3.6	3.6	3.6	.
		Overall [b]	32	2.0	3.3	2.0	5.0	1.5

[a]: Only quantified observations.
[b]: Quantified and Non-quantified observations combined.

Table 7. Number of quantified, non-quantified, and sample-specific quantitation limits for chromium by plant.

Plant Name	Sampled Location	Indicator	# of Observations	Quantitation Limits for Chromium (ug/L)			
				1	2	5	10
Allen	Bioreactor	Quantified	0	-	-	-	-
		Non-Quantified	64	1	27	31	5
Belews Creek	Bioreactor	Quantified	1	-	-	-	-
		Non-Quantified	31	0	18	13	0

c. Long Term Average, Daily Variability Factor, and Monthly Variability Factor

For chromium, the long term average is calculated as the arithmetic average of all observations (since all values were below quantitation limits for Allen, and only one value was reported as quantified for Belews Creek). Further, no variability factors are calculated (see sections 6 and 7 above for further discussion).

Table 8. Long term average and variability factors at bioreactor effluent sampling location for chromium (ug/L) by plant.

Plant Name	Sampled Location	N[a]	Long Term Average
Allen	Bioreactor	64	4.1*
Belews Creek		32	3.3*

[a]: Quantified and Non-quantified observations combined.
*Arithmetic mean.

3. Copper

a. Plots of Concentrations Over Time

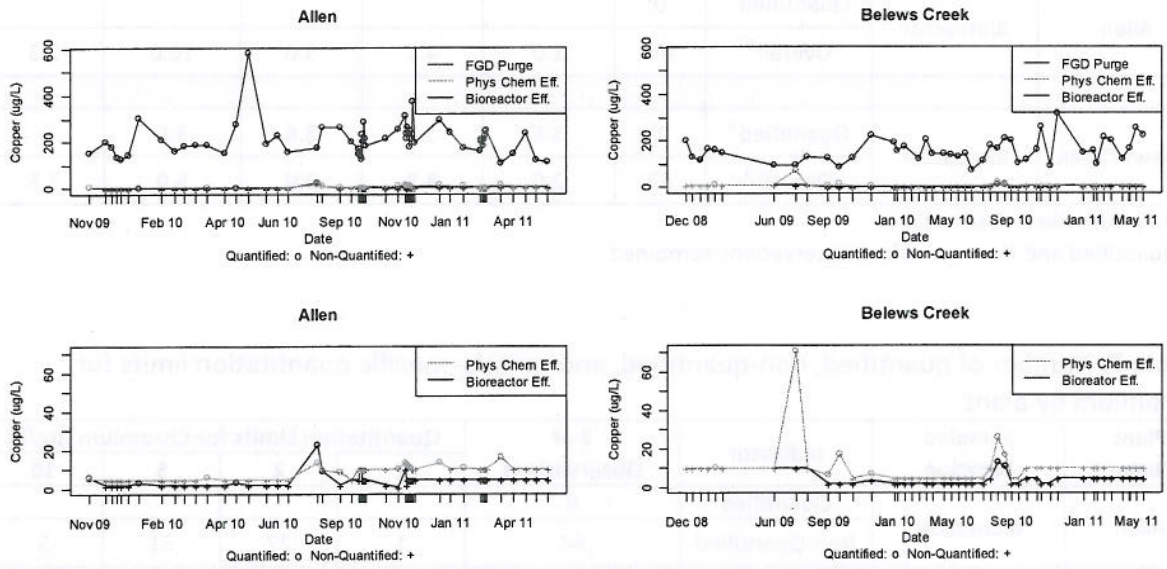

Figure 3. Plot of longitudinal data for copper (ug/L) for the Allen and Belews Creek plants after all necessary exclusions.

b. Summary Statistics

Table 9. Summary statistics of copper concentrations (ug/L) for quantified samples and all samples combined for FGD purge and physical-chemical effluent sampling locations by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Copper (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	FGD Purge	Quantified [a]	61	107.0	210.1	199.0	581.0	71.9
		Overall [b]	61	107.0	210.1	199.0	581.0	71.9
	Phys Chem Eff.	Quantified [a]	17	5.6	11.2	11.0	16.4	2.7
		Overall [b]	61	5.0	8.9	10.0	16.4	2.9
Belews Creek	FGD Purge	Quantified [a]	44	73.0	157.4	149.0	315.0	51.3
		Overall [b]	44	73.0	157.4	149.0	315.0	51.3
	Phys Chem Eff.	Quantified [a]	7	6.5	22.1	16.9	70.9	22.6
		Overall [b]	44	5.0	10.1	10.0	70.9	10.2

[a]: Only quantified observations.

[b]: Quantified and non-quantified observations combined.

Table 10. Summary statistics of copper concentrations (ug/L) for quantified samples and all samples combined for bioreactor effluent sampling location by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Copper (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	Bioreactor	Quantified [a]	5	3.0	8.8	6.1	22.5	8.1
		Overall [b]	61	1.0	4.7	5.0	22.5	3.3
Belews Creek	Bioreactor	Quantified [a]	3	10.5	11.4	10.9	12.8	1.2
		Overall [b]	44	2.0	5.0	5.0	12.8	3.4

[a]: Only Quantified observations.

[b]: Quantified and non-quantified observations combined.

Table 11. Number of quantified, non-quantified, and sample-specific quantitation limits for copper by plant.

Plant Name	Sampled Location	Indicator	# of Observations	Quantitation Limits for Copper (ug/L)				
				1	2	4	5	10
Allen	Bioreactor	Quantified	5	-	-	-	-	-
		Non-Quantified	56	1	19	0	31	5
Belews Creek	Bioreactor	Quantified	3	-	-	-	-	-
		Non-Quantified	41	0	19	1	13	8

c. Long Term Average, Daily Variability Factor, and Monthly Variability Factor

Table 12. Long term average and variability factors at bioreactor effluent sampling location for copper (ug/L) by plant.

Plant Name	Sampled Location	N[a]	Long Term Average	Daily Variability Factor		Monthly Variability Factor
Allen	Bioreactor	61	4.76	3.64		1.67
Belews Creek		44	5.03	2.52		1.50

[a]: Quantified and non-quantified observations combined.

4. Mercury

a. Plots of Concentrations Over Time

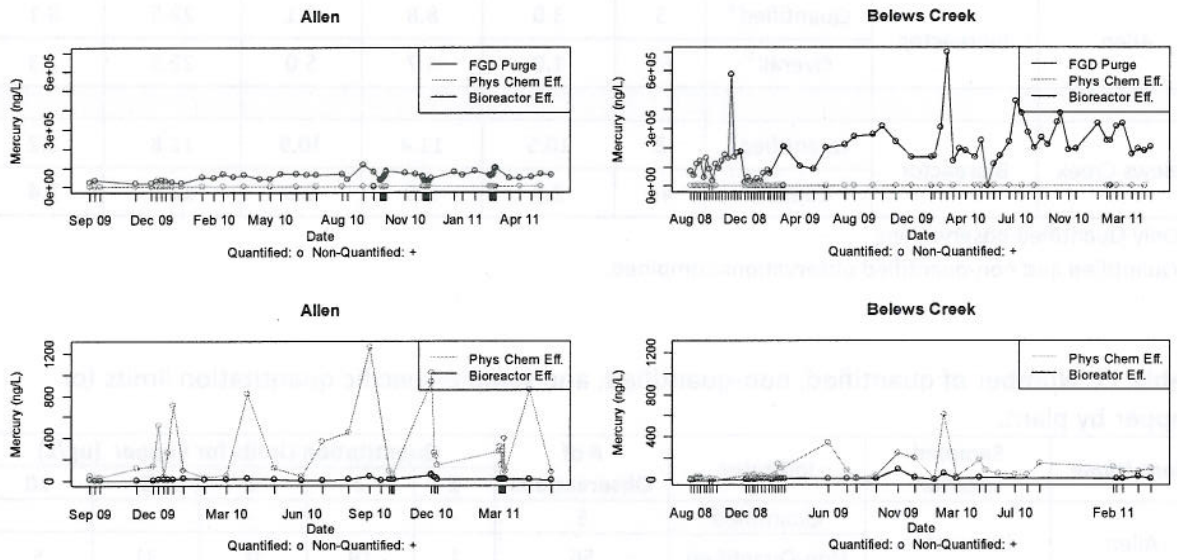

Figure 4. Plot of longitudinal data for mercury (ng/L) for the Allen and Belews Creek plants after all necessary exclusions.

b. Summary Statistics

Table 13. Summary statistics of mercury concentrations (ng/L) for quantified samples and all samples combined for FGD purge and physical-chemical effluent sampling locations by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Mercury (ng/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	FGD Purge	Quantified [a]	63	18,500.0	51,746.0	51,000.0	116,000.0	21,390.2
		Overall [b]	63	18,500.0	51,746.0	51,000.0	116,000.0	21,390.2
	Phys Chem Eff.	Quantified [a]	36	22.7	290.1	139.0	1,260.0	325.3
		Overall [b]	36	22.7	290.1	139.0	1,260.0	325.3
Belews Creek	FGD Purge	Quantified [a]	69	1,000.0	184,869.6	174,000.0	697,000.0	125,572.0
		Overall [b]	69	1,000.0	184,869.6	174,000.0	697,000.0	125,572.0
	Phys Chem Eff.	Quantified [a]	40	2.7	80.8	39.2	618.0	114.2
		Overall [b]	40	2.7	80.8	39.2	618.0	114.2

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 14. Summary statistics of mercury concentrations (ng/L) for quantified samples and all samples combined for bioreactor effluent sampling location by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Mercury (ng/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	Bioreactor	Quantified [a]	37	1.9	12.2	8.8	54.6	12.1
		Overall [b]	37	1.9	12.2	8.8	54.6	12.1
Belews Creek	Bioreactor	Quantified [a]	34	2.0	11.4	6.3	95.7	18.2
		Overall [b]	36	1.0	10.8	6.1	95.7	17.9

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 15. Number of quantified, non-quantified, and sample-specific quantitation limits for mercury by plant.

Plant Name	Sampled Location	Indicator	# of Observations	Quantitation Limits for Mercury (ng/L)
				1
Allen	Bioreactor	Quantified	37	-
		Non-Quantified	0	-
Belews Creek	Bioreactor	Quantified	34	-
		Non-Quantified	2	2

c. Long Term Average, Daily Variability Factor, and Monthly Variability Factor

Table 16. Long term average and variability factors at bioreactor effluent sampling location for mercury (ng/L) by plant.

Plant Name	Sampled Location	N[a]	Long Term Average	Daily Variability Factor		Monthly Variability Factor
Allen	Bioreactor	37	12.15	4.82		1.93
Belews Creek		36	9.46	5.32		2.04

[a]: Quantified and non-quantified observations combined.

5. Selenium

a. Plots of Concentrations Over Time

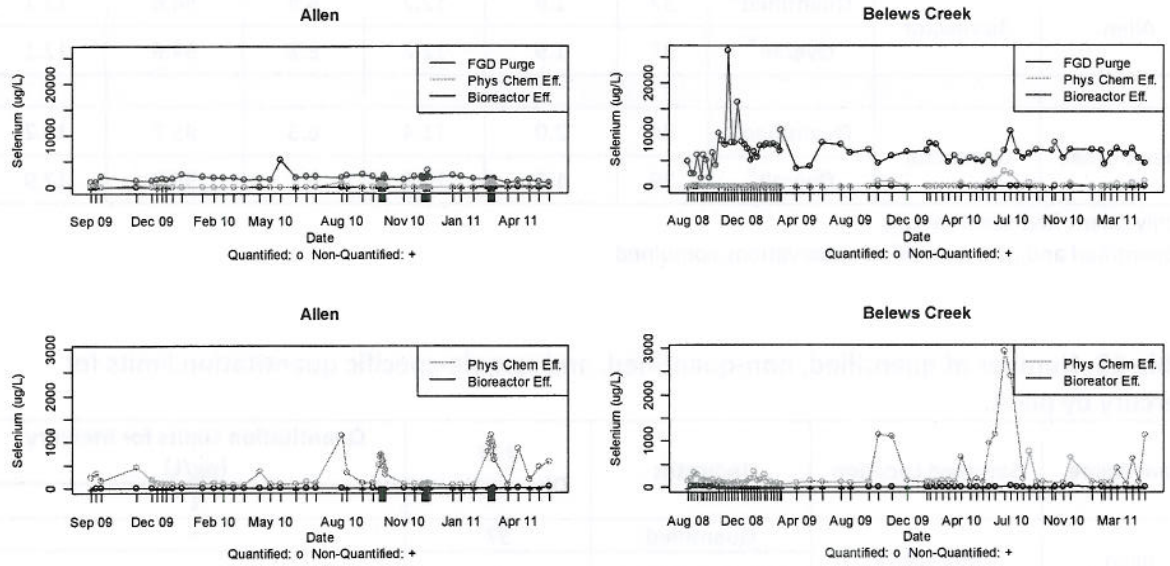

Figure 5. Plot of longitudinal data for selenium (ug/L) for the Allen and Belews Creek plants after all necessary exclusions.

b. Summary Statistics

Table 17. Summary statistics of selenium concentrations (ug/L) for quantified samples and all samples combined for FGD purge and physical-chemical effluent sampling locations by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Selenium (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	FGD Purge	Quantified [a]	64	849.0	1,731.6	1,655.0	5,350.0	642.5
		Overall [b]	64	849.0	1,731.6	1,655.0	5,350.0	642.5
	Phys Chem Eff.	Quantified [a]	64	57.7	322.7	115.5	1,140.0	319.9
		Overall [b]	64	57.7	322.7	115.5	1,140.0	319.9
Belews Creek	FGD Purge	Quantified [a]	70	1,660.0	6,850.9	6,665.0	26,200.0	3,242.6
		Overall [b]	70	1,660.0	6,850.9	6,665.0	26,200.0	3,242.6
	Phys Chem Eff.	Quantified [a]	69	58.0	305.2	121.0	2,940.0	497.5
		Overall [b]	69	58.0	305.2	121.0	2,940.0	497.5

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 18. Summary statistics of selenium concentrations (ug/L) for quantified samples and all samples combined for bioreactor effluent sampling location by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Selenium (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	Bioreactor	Quantified [a]	26	2.4	4.1	3.5	11.7	2.0
		Overall [b]	64	1.0	4.7	5.0	11.7	2.2
Belews Creek	Bioreactor	Quantified [a]	33	2.6	9.5	7.7	26.7	6.6
		Overall [b]	71	2.0	8.8	10.0	26.7	4.9

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 19. Number of quantified, non-quantified, and sample-specific quantitation limits for selenium by plant.

Plant Name	Sampled Location	Indicator	# of Observations	Quantitation Limits for Selenium (ug/L)				
				1	2	4	5	10
Allen	Bioreactor	Quantified	26	-	-		-	-
		Non-Quantified	38	1	6	0	26	5
Belews Creek	Bioreactor	Quantified	33	-	-	-	-	-
		Non-Quantified	38	0	2	1	9	26

c. Long Term Average, Daily Variability Factor, and Monthly Variability Factor

Table 20. Long term average and variability factors at bioreactor effluent sampling location for selenium (ug/L) by plant.

Plant Name	Sampled Location	N[a]	Long Term Average	Daily Variability Factor		Monthly Variability Factor
Allen	Bioreactor	64	4.68	2.14		1.36
Belews Creek		71	8.83	3.34		1.56

[a]: Quantified and non-quantified observations combined.

6. Zinc

a. Plots of Concentrations Over Time

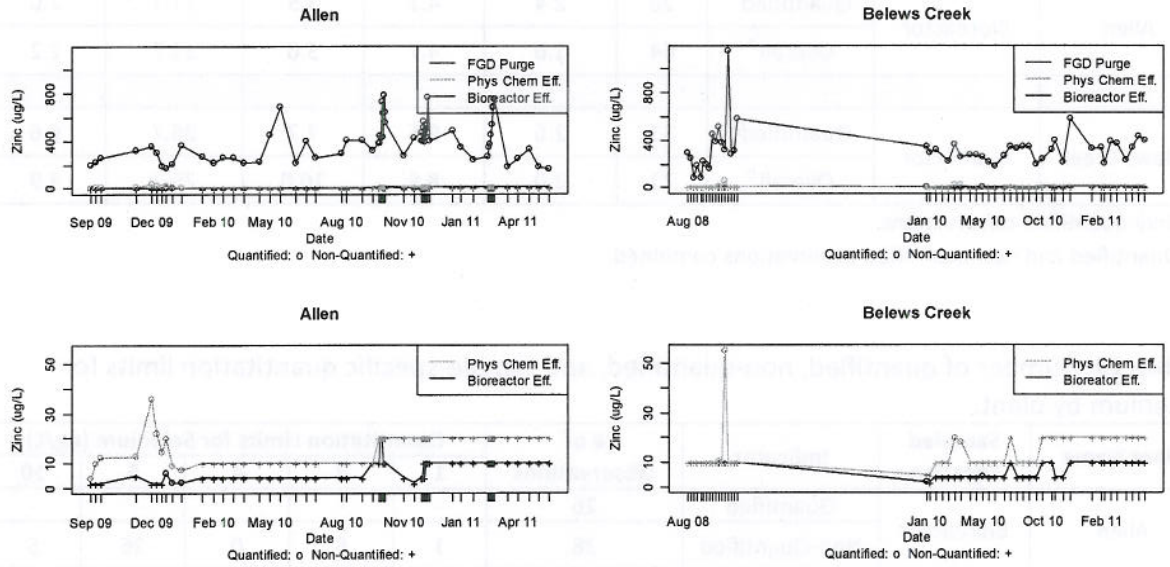

Figure 6. Plot of longitudinal data for zinc (ug/L) for the Allen and Belews Creek plants after all necessary exclusions.

b. Summary Statistics

Table 21. Summary statistics of zinc concentrations (ug/L) for quantified samples and all samples combined for FGD purge and physical-chemical effluent sampling locations by plant.

Plant Name	Sampled Location	Summary Statistics for Zinc (ug/L)						
		Indicator	N	Minimum	Mean	Median	Maximum	STD
Allen	FGD Purge	Quantified [a]	64	159.0	393.3	383.0	787.0	165.3
		Overall [b]	64	159.0	393.3	383.0	787.0	165.3
	Phys Chem Eff.	Quantified [a]	10	3.6	14.5	12.4	35.6	9.2
		Overall [b]	64	3.6	16.8	20.0	35.6	5.5
Belews Creek	FGD Purge	Quantified [a]	48	80.0	321.7	312.0	1,150.0	161.5
		Overall [b]	48	80.0	321.7	312.0	1,150.0	161.5
	Phys Chem Eff.	Quantified [a]	4	10.6	26.0	19.0	55.2	19.9
		Overall [b]	49	5.0	14.2	10.0	55.2	7.8

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 22. Summary statistics of zinc concentrations (ug/L) for quantified samples and all samples combined for bioreactor effluent sampling location by plant.

Plant Name	Sampled Location	Indicator	Summary Statistics for Zinc (ug/L)					
			N	Minimum	Mean	Median	Maximum	STD
Allen	Bioreactor	Quantified [a]	3	2.0	3.3	2.1	5.8	2.2
		Overall [b]	64	2.0	7.6	10.0	20.0	4.6
Belews Creek	Bioreactor	Quantified [a]	2	2.1	3.2	3.2	4.2	1.5
		Overall [b]	48	2.0	7.5	10.0	10.0	3.1

[a]: Only quantified observations.
[b]: Quantified and non-quantified observations combined.

Table 23. Number of quantified, non-quantified, and sample-specific quantitation limits for zinc by plant.

Plant Name	Sampled Location	Indicator	# of Observations	Quantitation Limits for Zinc (ug/L)				
				2	4	5	10	20
Allen	Bioreactor	Quantified [a]	3	-	-		-	-
		Non-Quantified [b]	61	7	19	1	30	4
Belews Creek	Bioreactor	Quantified [a]	2	-	-	-	-	-
		Non-Quantified [b]	46	1	16	0	29	0

c. Long Term Average, Daily Variability Factor, and Monthly Variability Factor

Table 24. Long term average and variability factors at bioreactor effluent sampling location for zinc (ug/L) by plant.

Plant Name	Sampled Location	N[a]	Long Term Average	Daily Variability Factor		Monthly Variability Factor
Allen	Bioreactor	64	7.57	2.64		1.00
Belews Creek		48	7.56	1.32		1.00

[a]: Quantified and non-quantified observations combined.

V. COMBINED LONG TERM AVERAGES, COMBINED VARIABILITY FACTORS, AND EFFLUENT LIMITS

As seen in previous sections, for each pollutant at Allen and Belews Creek plants the long term average, daily variability factor, and monthly variability factor were obtained. Now, the combined long term average can be found by taking the median of the long term averages of the two plants. The combined (daily or monthly) variability factor can be found by taking the mean of the variability factors of the two plants. The daily (monthly) limit is found by taking the product of the combined long term average and the combined daily (monthly) variability factor.

Table 25 below shows the long term average, 99th percentile of the distribution of the daily values, daily variability factor, 95th percentile of the distribution monthly average (assuming 4 samples per month), and monthly variability factor for each plant by pollutant.

Table 25. For each pollutant and plant, the long term average, percentiles, and variability factors.

Pollutant	Plant Name	Total # of Samples (N) [a]	Long Term Average	99th Percentile[b]	Daily Variability Factor	95th Percentile[c]	Monthly Variability Factor
Arsenic (ug/L)	Allen	57	6.05	18.63	3.08	9.05	1.50
	Belews Creek	65	6.28	9.99	1.59	6.66	1.06
Chromium (ug/L)	Allen	64	4.1[d]	-	-	-	-
	Belews Creek	32	3.3[d]	-	-	-	-
Copper (ug/L)	Allen	61	4.76	17.33	3.64	7.93	1.67
	Belews Creek	44	5.03	12.68	2.52	7.56	1.50
Mercury (ng/L)	Allen	37	12.15	58.56	4.82	23.41	1.93
	Belews Creek	36	9.46	50.33	5.32	19.32	2.04
Selenium (ug/L)	Allen	64	4.68	10.02	2.14	6.35	1.36
	Belews Creek	71	8.83	29.49	3.34	13.75	1.56
Zinc Ug/L)	Allen	64	7.58	20.01	2.64	1.00	1.00
	Belews Creek	48	7.56	9.98	1.32	1.00	1.00

[a]: Quantified and non-quantified observations combined.
[b]: 99th percentile of the distribution of daily values.
[c]: 95th percentile of the distribution of monthly averages (assuming 4 samples per month).
[d]: Arithmetic mean.

Table 26 below provides the combined long term averages, combined daily and monthly variability factors, and the resulting daily and monthly limits. The combined monthly variability factor for zinc is calculated as the mean of the monthly variability factors for the two plants. This would result in a combined monthly variability factor of 1.0 for zinc. EPA is concerned that

using a variability factor of 1.0 would not provide sufficient allowance for variability relative to the long term average. Therefore, EPA is transferring the combined monthly variability factor for copper (1.58) to substitute for the calculated combined monthly variability factor for zinc, because of similarities in the pollutant removal mechanisms for these pollutants. Further, EPA has decided to round the limits up to the next highest integer. For example, the daily maximum limits for arsenic and copper are calculated to be 14.8 and 15.1, respectively. After rounding, these limits have become 15 ug/L and 16 ug/L, respectively.

Table 26. Combined long term average, variability factors, and limits.

Pollutant	Combined Long Term Average	Combined Daily Variability	Combined Monthly Variability	Limits[d]	
				Daily Maximum Limit	Monthly Average Limit
Arsenic (ug/L)	6.17	2.34	1.28	15	8
Chromium (ug/L)	3.70[a]	-	-	10[c]	-
Copper (ug/L)	4.89	3.08	1.58	16	8
Mercury (ng/L)	10.80	5.07	1.98	55	22
Selenium (ug/L)	6.75	2.74	1.46	19	10
Zinc (ug/L)	7.57	1.98	1.58[b]	15	12

[a]: Arithmetic mean.
[b]: Value was transferred from the combined monthly variability factor for copper.
[c]: Limit set equal to the highest reported sample-specific quantitation limit.
[d]: Effluent limits have been rounded upward to the next highest integer.

References

Aitchison, J. and Brown, J.A.C. (1969). *The Lognormal Distribution*. Cambridge University Press.

Kahn, H.D. and Rubin, M.B. (1989). "Use of Statistical Methods in Industrial Water Pollution Control Regulations in the United States", *Environmental Monitoring and Assessment*, 12: 129-148.

APPENDIX 1: Plots of All Available Data (before any exclusions)

Appendix 1 contains the longitudinal plots of the monitoring data from the Allen and Belews Creek power plants provided by Duke Energy. These plots were based on all the available data, i.e., before excluding data for any reasons. This appendix contains plots for the following pollutants: arsenic, chromium, copper, mercury, selenium, and zinc. The red line on the plot represents the end of the commissioning period. For Allen, the commissioning period was determined to be from March 03, 2009 through August 26, 2009. For Belews Creek, the commissioning period was determined to be from February 6, 2008 through July 31, 2008. Note that for copper, all data were collected after the commissioning period for both plants; therefore, plots for copper do not have a red line to indicate the end of commissioning period.

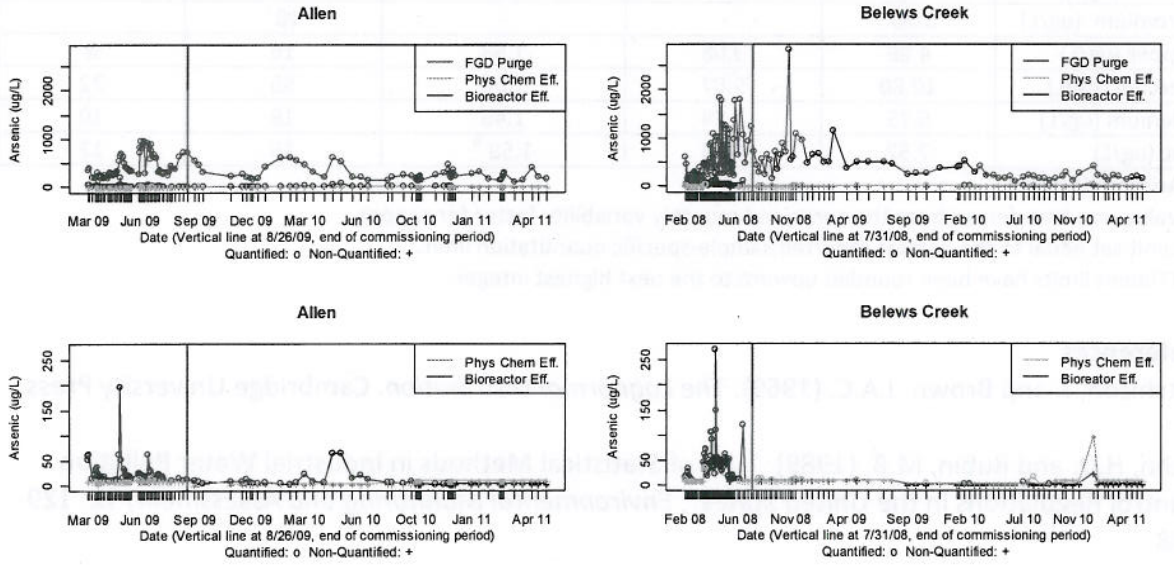

Figure 1.1 Plot of original data for arsenic (ug/L) for Allen and Belews Creek plants (before exclusions).

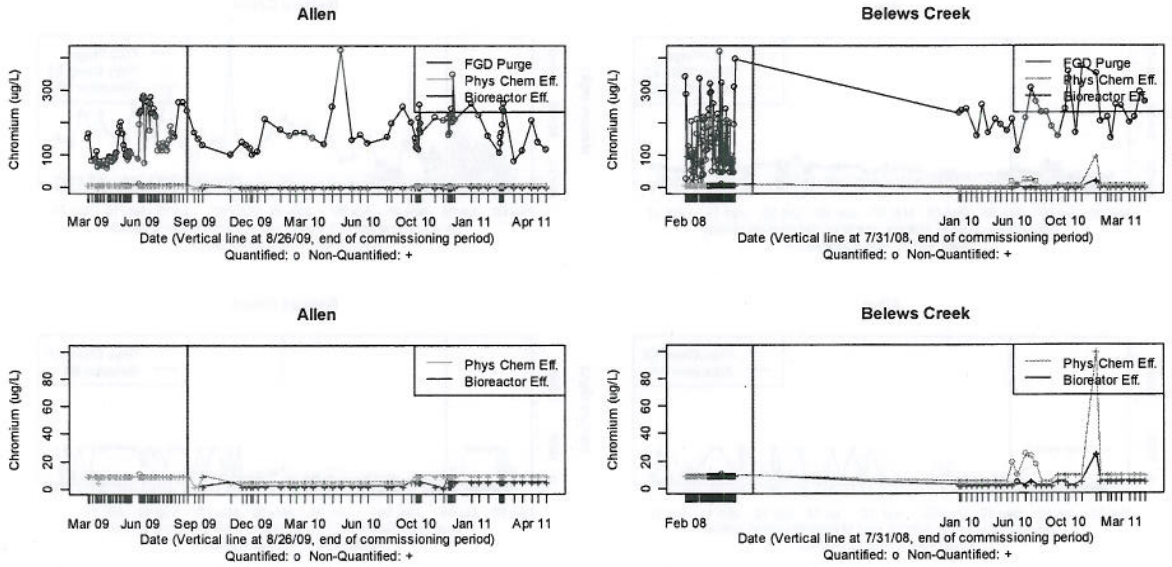

Figure 1.2. Plot of original data for chromium (ug/L) for Allen and Belews Creek plants (before exclusions).

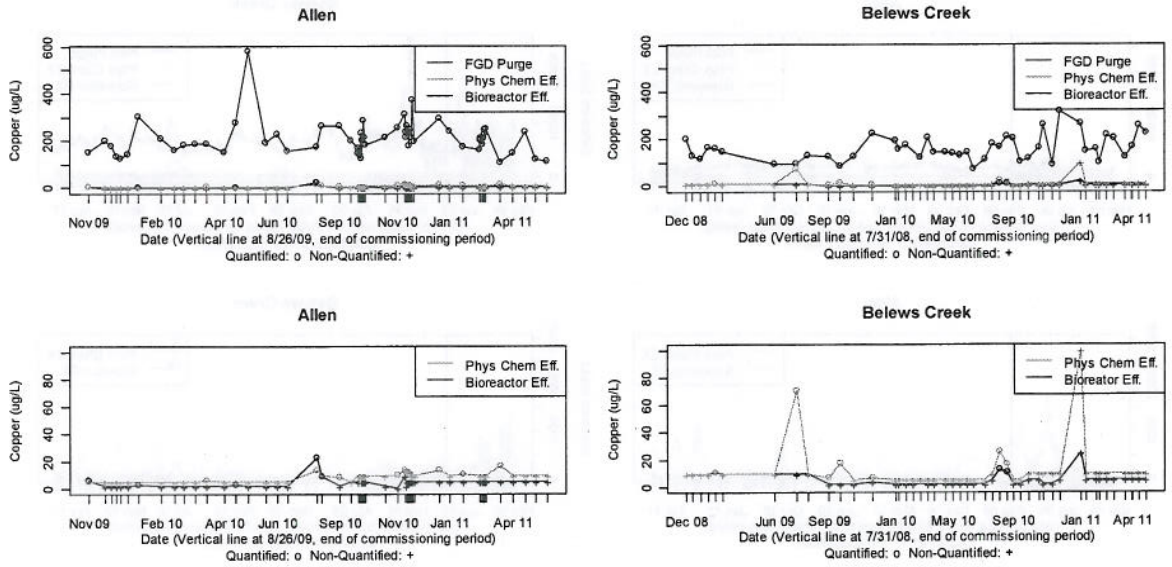

Figure 1.3. Plot of original data for copper (ug/L) for Allen and Belews Creek plants (before exclusions).

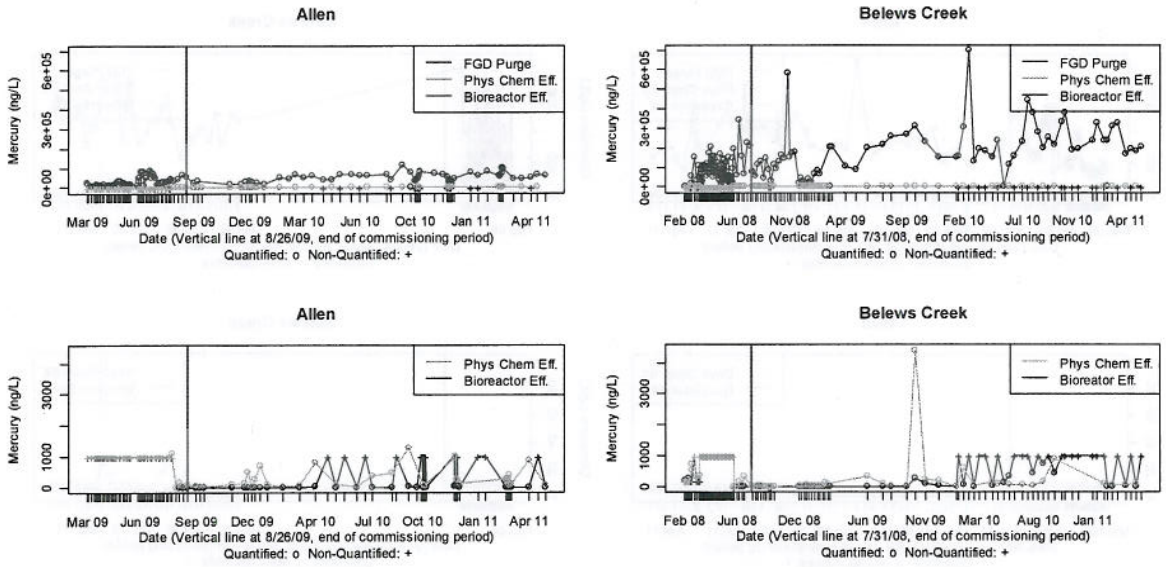

Figure 1.4. Plot of original data for mercury (ng/L) for Allen and Belews Creek plants (before exclusions).

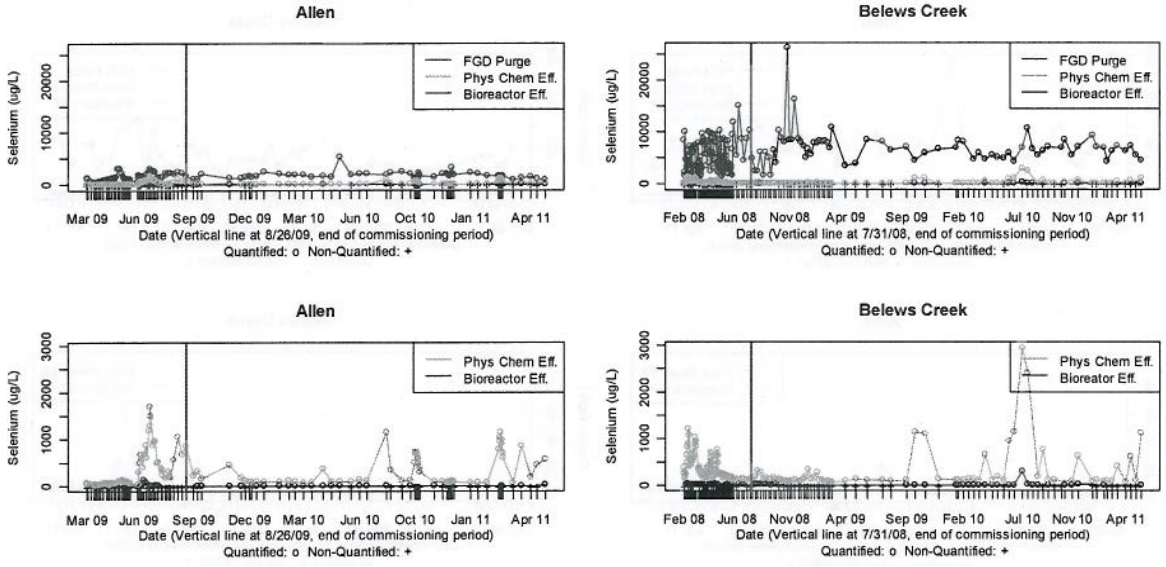

Figure 1.5. Plot of original data for selenium (ug/L) for Allen and Belews Creek plants (before exclusions).

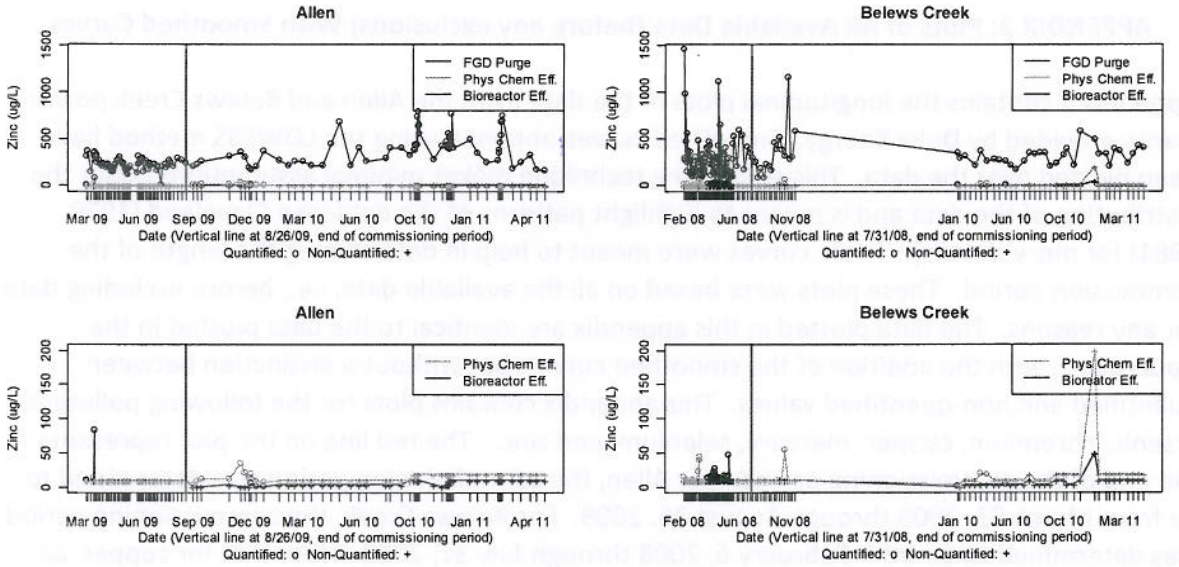

Figure 1.6. Plot of original data for zinc (ug/L) for Allen and Belews Creek plants (before exclusions).

APPENDIX 2: Plots of All Available Data (before any exclusions) With Smoothed Curves

Appendix 2 contains the longitudinal plots of the data from the Allen and Belews Creek power plants provided by Duke Energy. Smoothed curves obtained using the LOWESS method have been plotted over the data. This smoothing technique makes minimal assumptions about the distribution of the data and is meant to highlight patterns of the data (see Cleveland (1979, 1981) for more details). These curves were meant to help in determining the length of the commission period. These plots were based on all the available data, i.e., before excluding data for any reasons. The data plotted in this appendix are identical to the data plotted in the Appendix 1, with the addition of the smoothed curves, but without a distinction between quantified and non-quantified values. This appendix contains plots for the following pollutants: arsenic, chromium, copper, mercury, selenium, and zinc. The red line on the plot represents the end of the commissioning period. For Allen, the commissioning period was determined to be from March 03, 2009 through August 26, 2009. For Belews Creek, the commissioning period was determined to be from February 6, 2008 through July 31, 2008. Note that for copper, all data were collected after the commissioning period for both plants, therefore, plots for copper do not have a red line to indicate the end of commissioning period.

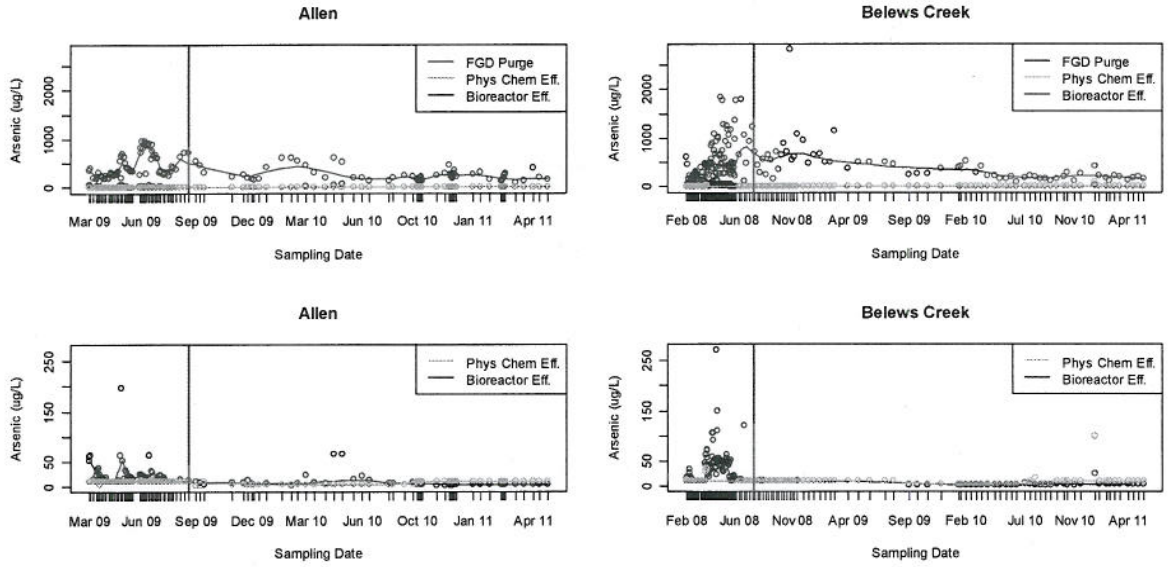

Figure 2.1. Smoothed plot (using LOWESS) of the original data for arsenic (ug/L) for Allen and Belews Creek plants (before exclusions).

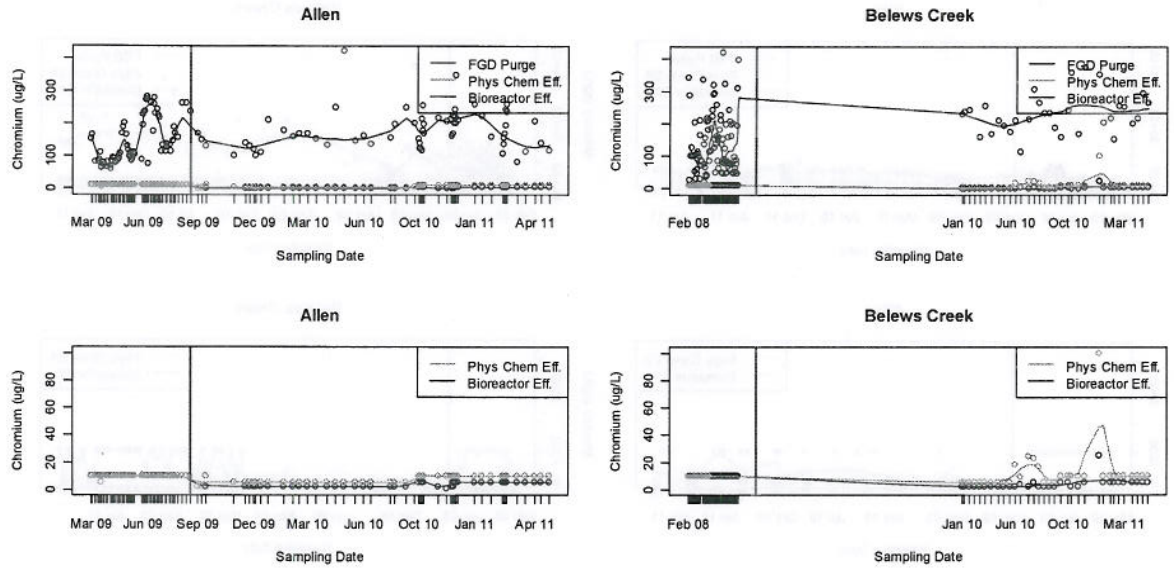

Figure 2.2. Smoothed plot (using LOWESS) of the original data for chromium (ug/L) for Allen and Belews Creek plants (before exclusions).

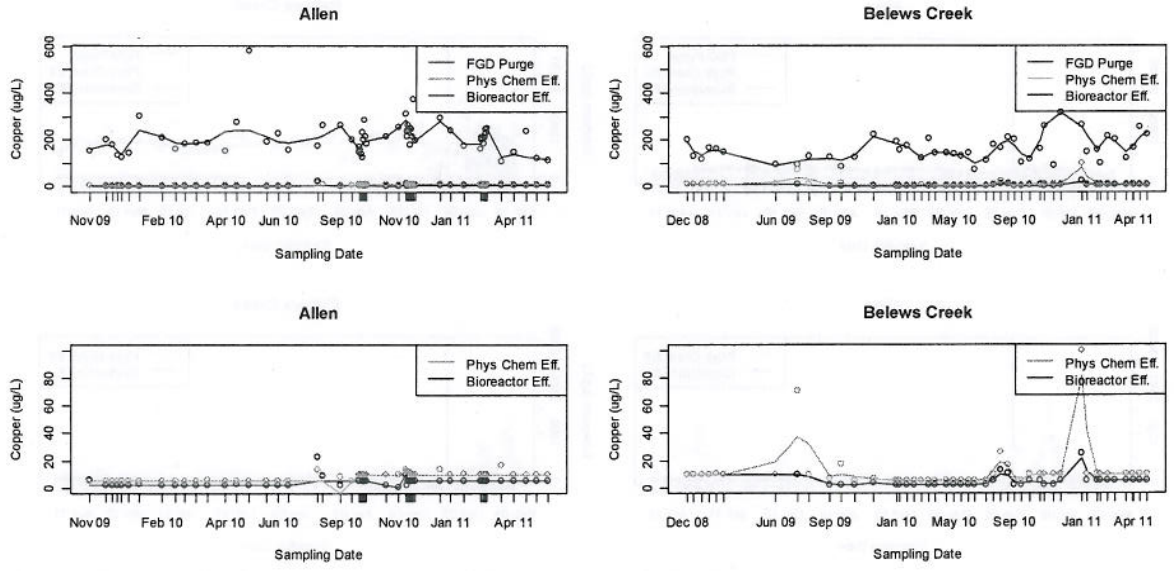

Figure 2.3. Smoothed plot (using LOWESS) of the original data for copper (ug/L) for Allen and Belews Creek plants (before exclusions).

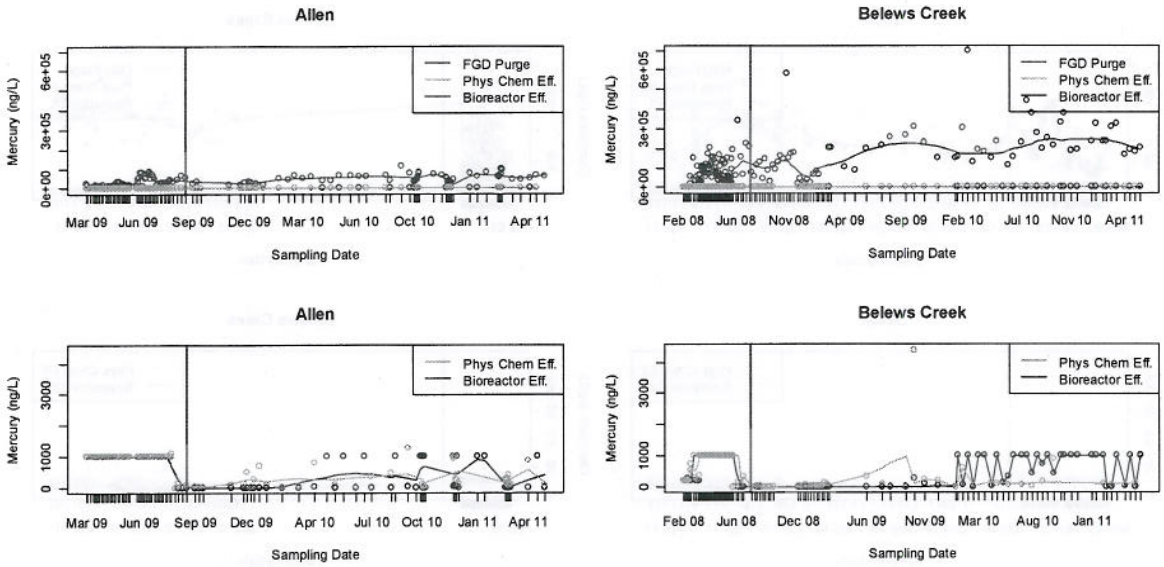

Figure 2.4. Smoothed plot (using LOWESS) of the original data for mercury (ng/L) for Allen and Belews Creek plants (before exclusions).

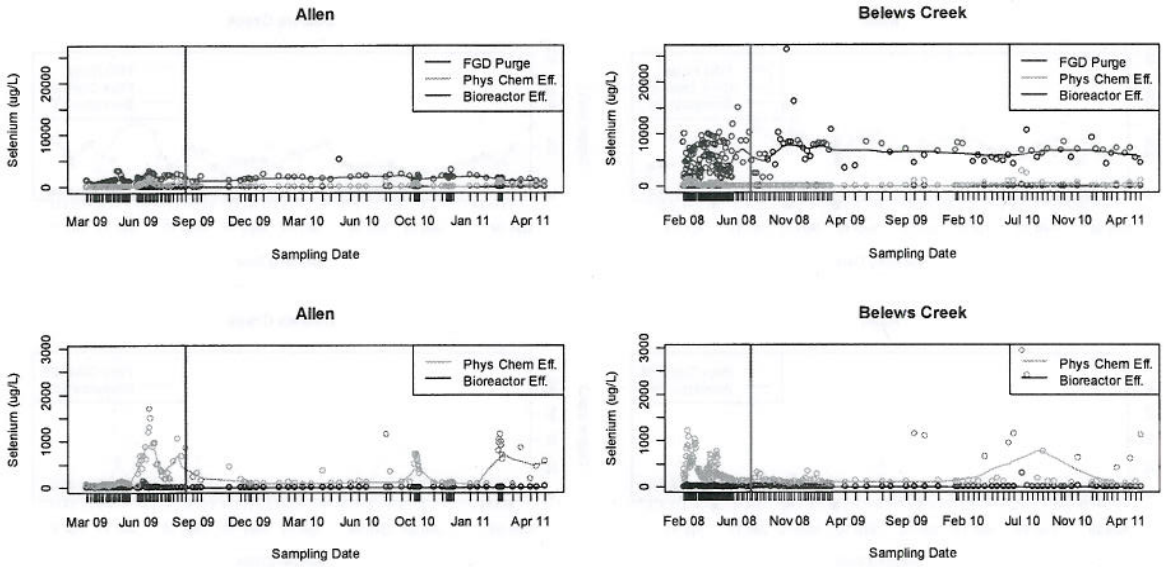

Figure 2.5. Smoothed plot (using LOWESS) of the original data for selenium (ug/L) for Allen and Belews Creek plants (before exclusions).

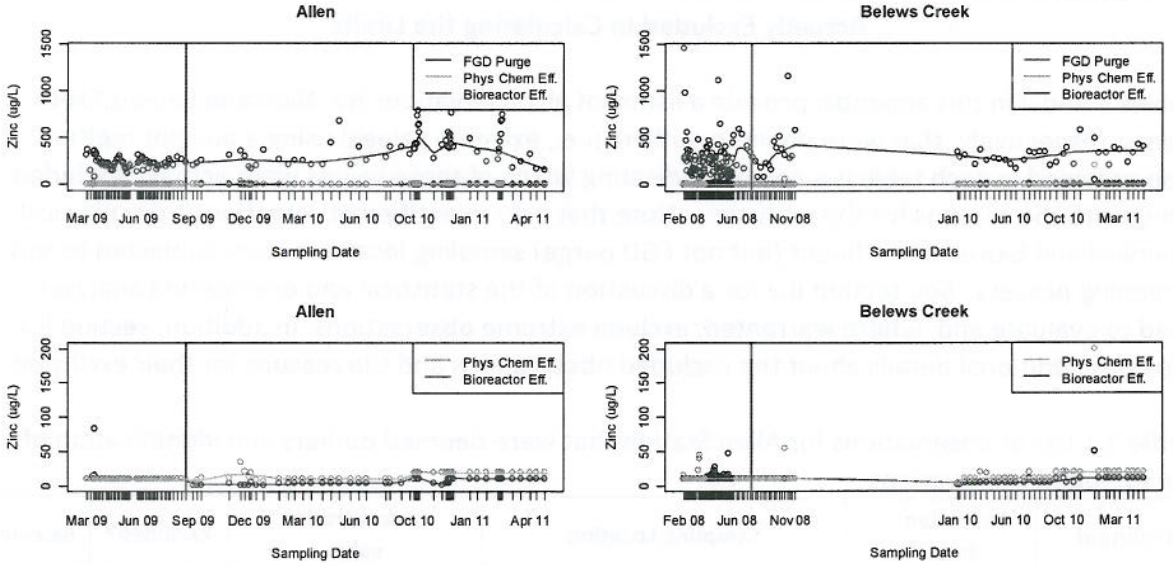

Figure 2.6. Smoothed plot (using LOWESS) of the original data for zinc (ug/L) for Allen and Belews Creek plants (before exclusions).

References

Cleveland, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829–836.

Cleveland, W. S. (1981) LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, **35**, 54.

APPENDIX 3: List of All Potential Extreme Observations and Identification of Extreme Values Actually Excluded In Calculating the Limits

Tables 1 and 2 in this appendix provide a listing of all observations for Allen and Belews Creek plants, respectively, that were deemed outliers (i.e., extreme values) using a boxplot method. Also provided in each table is a column indicating which of these values were actually excluded along with the reasons for the exclusion. Note that only quantified observations from physical-chemical and bioreactor effluent (but not FGD purge) sampling locations were subjected to this screening process. See section II.2 for a discussion of the statistical and engineering analyses used to evaluate and, where warranted, exclude extreme observations. In addition, section II.2 provides additional details about the excluded observations and the reasons for their exclusion.

Table 3.1 List of observations for Allen Station that were deemed outliers and identification of the excluded observations

Pollutant	Collection date	Sampling Location	Concentration value	Excluded?	Reason
Arsenic	29-Jun-10	Bioreactor Effluent	20.1 ug/L	No	-
Copper	11-Nov-09	Physical-chemical Effluent	5.6 ug/L	No	-
Copper	5-Apr-10	Physical-chemical Effluent	6 ug/L	No	-
Copper	16-Aug-10	Bioreactor Effluent	22.5 ug/L	No	-
Copper	28-Mar-11	Physical-chemical Effluent	16.4 ug/L	No	-
Mercury	13-Sep-10	Physical-chemical Effluent	1260 ng/L	No	-
Mercury	13-Sep-10	Bioreactor Effluent	38.4 ng/L	No	-
Mercury	2-Dec-10	Physical-chemical Effluent	929 ng/L	No	-
Mercury	2-Dec-10	Bioreactor Effluent	48.3 ng/L	No	-
Mercury	3-Dec-10	Physical-chemical Effluent	1020 ng/L	No	-
Mercury	3-Dec-10	Bioreactor Effluent	54.6 ng/L	No	-
Mercury	8-Dec-10	Bioreactor Effluent	25.5 ng/L	No	-
Selenium	23-May-11	Bioreactor Effluent	11.7 ug/L	No	-

Table 3.2 List of observations for Belews Creek Station that were deemed outliers and identification of the excluded observations

Pollutant	Collection date	Sampling Location	Value	Excluded?	Reason
Arsenic	2-Nov-09	Physical-chemical Effluent	5.1 ug/L	No	
Arsenic	11-Aug-10	Physical-chemical Effluent	17.1 ug/L	No	
Copper	13-Jul-09	Physical-chemical Effluent	70.9 ug/L	No	
Mercury	1-Jun-09	Physical-chemical Effluent	349 ng/L	No	
Mercury	**5-Oct-09**	**Physical-chemical Effluent**	**4400 ng/L**	**Yes**	**System upset**
Mercury	**5-Oct-09**	**Bioreactor Effluent**	**256 ng/L**	**Yes**	**System upset**
Mercury	2-Nov-09	Physical-chemical Effluent	247 ng/L	No	
Mercury	2-Nov-09	Bioreactor Effluent	95.7 ng/L	No	
Mercury	10-Feb-10	Physical-chemical Effluent	618 ng/L	No	
Mercury	10-Feb-10	Bioreactor Effluent	60.4 ng/L	No	
Mercury	**26-May-10**	**Bioreactor Effluent**	**136 ng/L**	**Yes**	**System upset**
Mercury	**9-Jun-10**	**Bioreactor Effluent**	**333 ng/L**	**Yes**	**System upset**
Mercury	**11-Aug-10**	**Bioreactor Effluent**	**423 ng/L**	**Yes**	**System upset**
Mercury	**8-Sep-10**	**Bioreactor Effluent**	**746 ng/L**	**Yes**	**System upset**
Mercury	**7-Oct-10**	**Physical-chemical Effluent**	**892 ng/L**	**Yes**	**System upset**
Mercury	**7-Oct-10**	**Bioreactor Effluent**	**442 ng/L**	**Yes**	**System upset**
Selenium	18-Aug-08	Physical-chemical Effluent	302 ug/L	No	
Selenium	29-Dec-08	Physical-chemical Effluent	346 ug/L	No	
Selenium	5-Oct-09	Physical-chemical Effluent	1130 ug/L	No	
Selenium	2-Nov-09	Physical-chemical Effluent	1090 ug/L	No	
Selenium	7-Apr-10	Physical-chemical Effluent	638 ug/L	No	
Selenium	9-Jun-10	Physical-chemical Effluent	941 ug/L	No	
Selenium	23-Jun-10	Physical-chemical Effluent	1130 ug/L	No	
Selenium	14-Jul-10	Physical-chemical Effluent	2940 ug/L	No	
Selenium	**14-Jul-10**	**Bioreactor Effluent**	**299 ug/L**	**Yes**	**Indeterminate extreme observation**
Selenium	28-Jul-10	Physical-chemical Effluent	2400 ug/L	No	
Selenium	8-Sep-10	Physical-chemical Effluent	757 ug/L	No	
Selenium	8-Dec-10	Physical-chemical Effluent	616 ug/L	No	
Selenium	23-Mar-11	Physical-chemical Effluent	399 ug/L	No	
Selenium	26-Apr-11	Physical-chemical Effluent	604 ug/L	No	
Selenium	25-May-11	Physical-chemical Effluent	1110 ug/L	No	

APPENDIX 4: Plots of Natural Logarithm of the Data Used in Calculating the Limits

Appendix 4 contains the plots of the natural logarithm of the data from the Allen and Belews Creek power plants used in setting the limits. These plots were based on only data that are used in setting the limits (after making all necessary exclusions). This appendix contains plots for the following pollutants: arsenic, chromium, copper, mercury, selenium, and zinc.

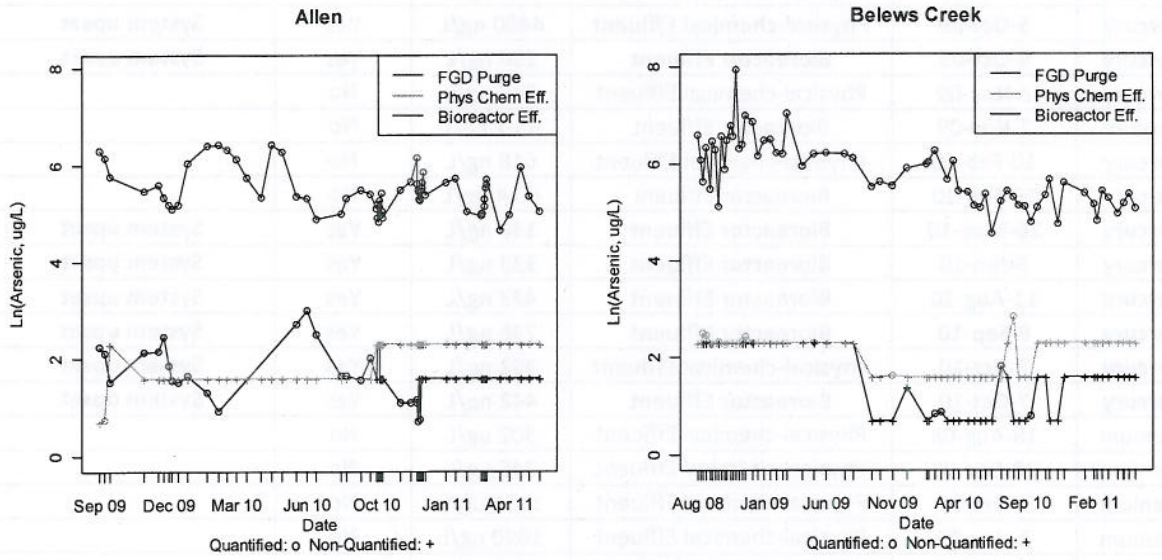

Figure 4.1. Plot of the natural logarithm of arsenic (ug/L) data for Allen and Belews Creek after all data exclusions.

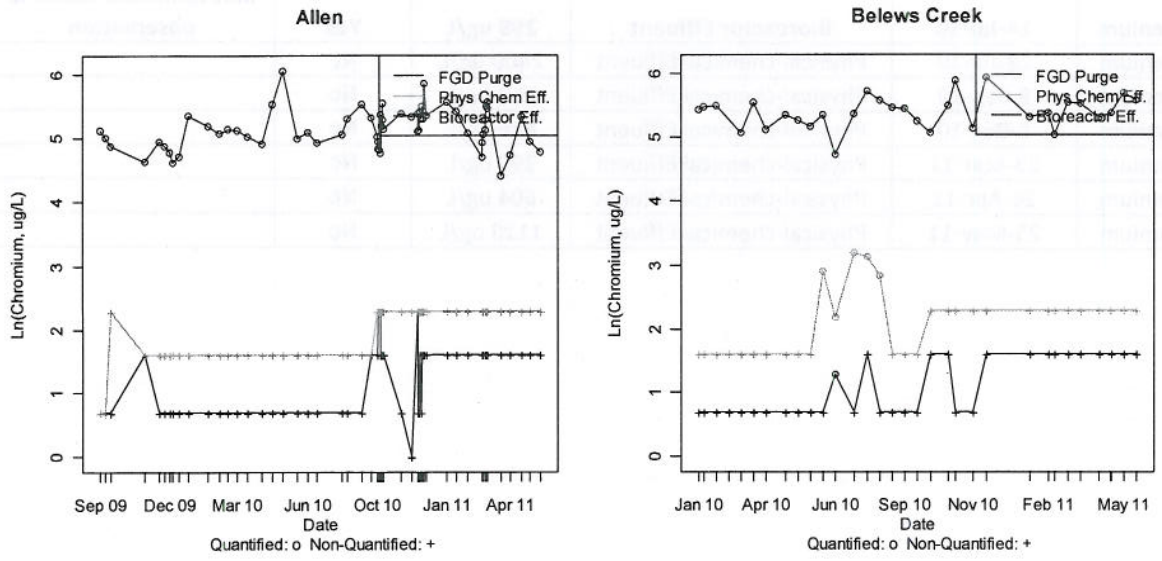

Figure 4.2. Plot of the natural logarithm of chromium (ug/L) data for Allen and Belews Creek after all data exclusions.

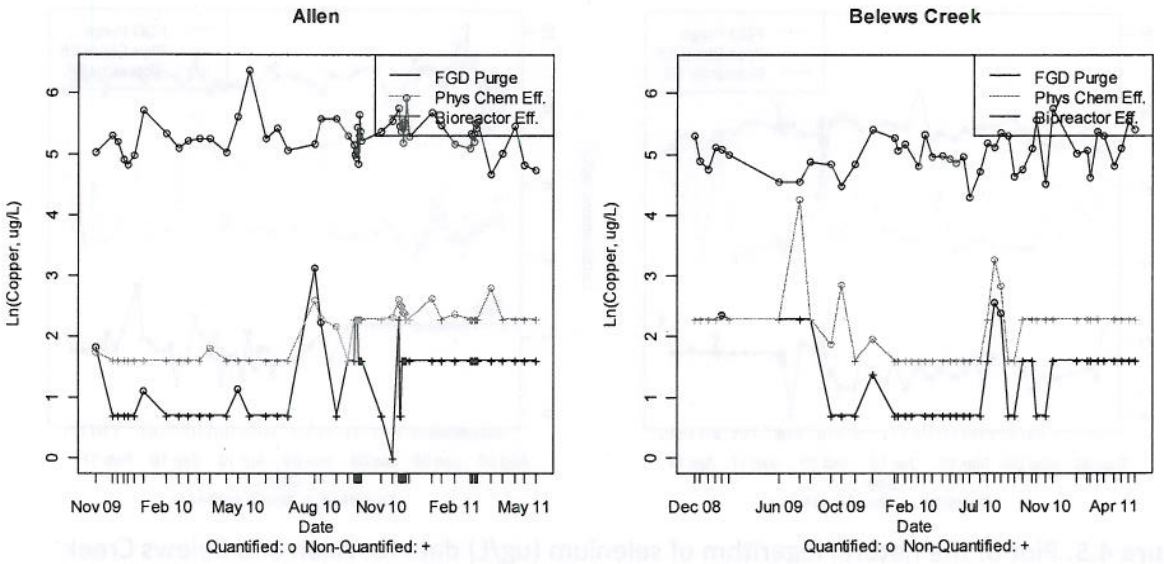

Figure 4.3. Plot of the natural logarithm of copper (ug/L) data for Allen and Belews Creek after all data exclusions.

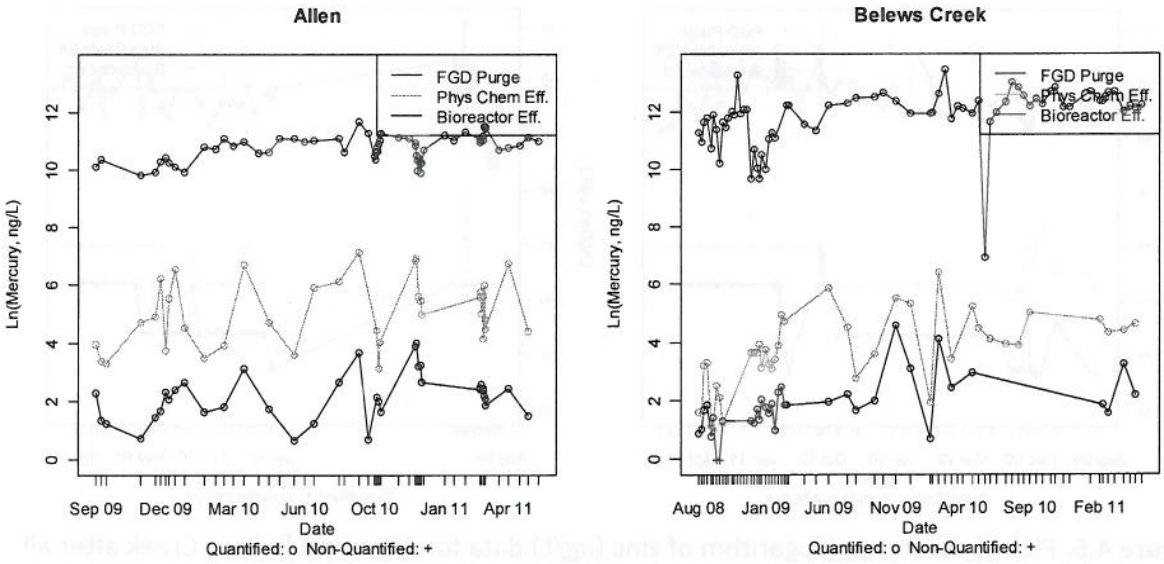

Figure 4.4. Plot of the natural logarithm of mercury (ng/L) data for Allen and Belews Creek after all data exclusions.

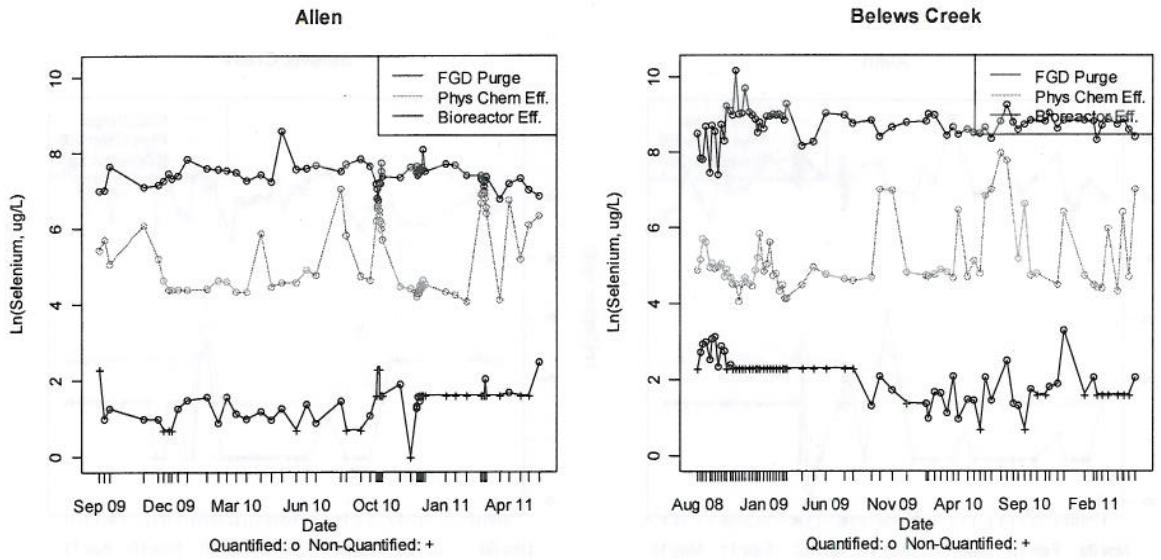

Figure 4.5. Plot of the natural logarithm of selenium (ug/L) data for Allen and Belews Creek after all data exclusions.

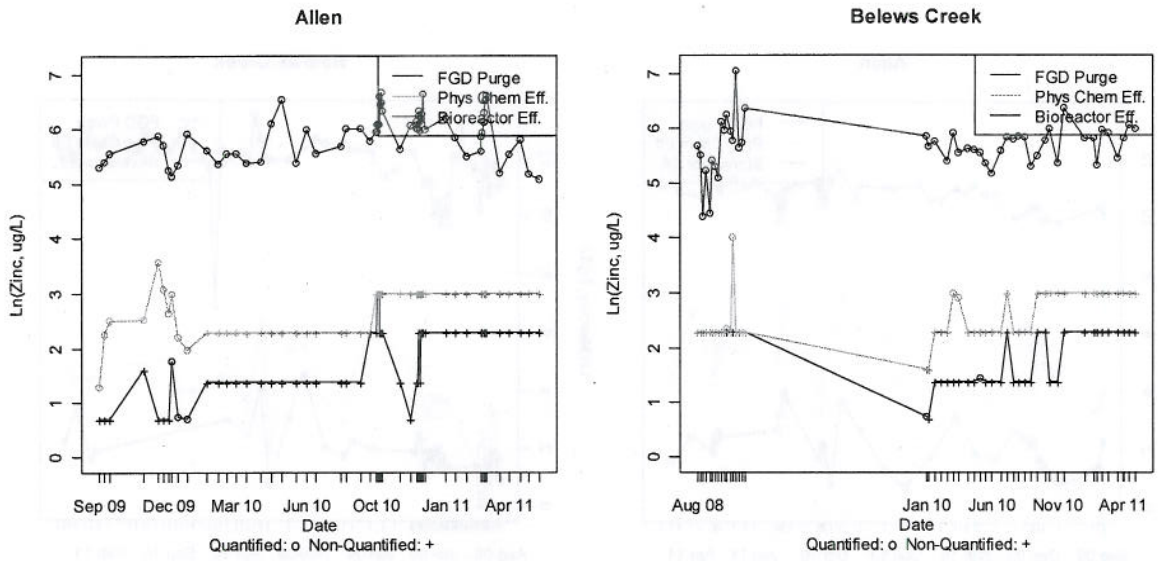

Figure 4.6. Plot of the natural logarithm of zinc (ug/L) data for Allen and Belews Creek after all data exclusions.

APPENDIX 5: Plots on a Logarithmic Scale of the Data Used in Calculating the Limits

Appendix 5 contains the plots on a logarithmic scale of the data from the Allen and Belews Creek power plants used in setting the limits. These plots were based on only data that are used in setting the limits (after making all necessary exclusions). This appendix contains plots for the following pollutants: arsenic, chromium, copper, mercury, selenium, and zinc.

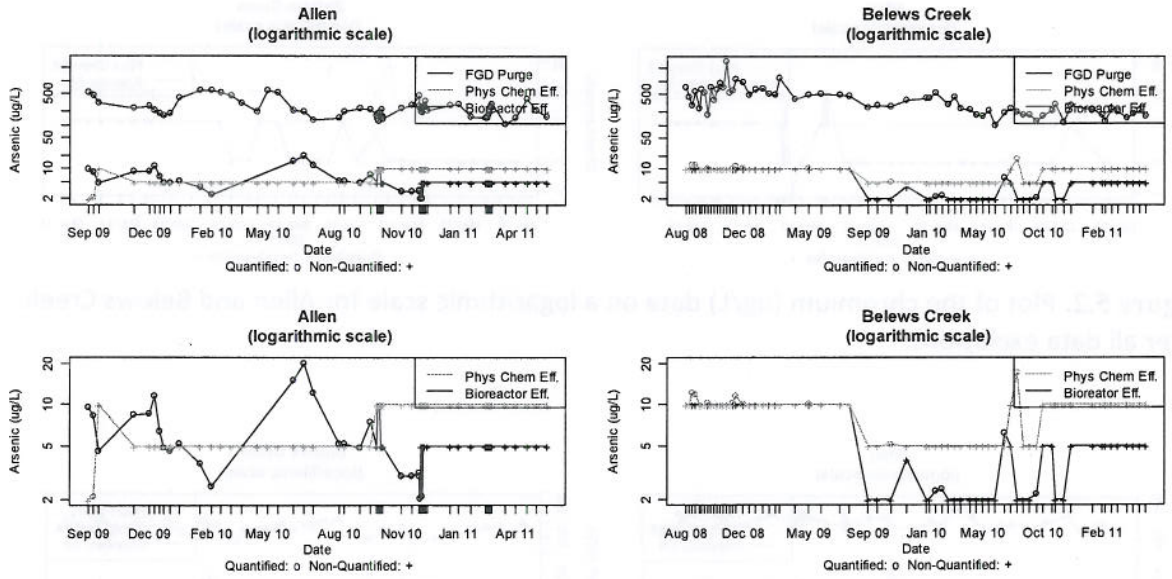

Figure 5.1. Plot of the arsenic (ug/L) data on a logarithmic scale for Allen and Belews Creek after all data exclusions.

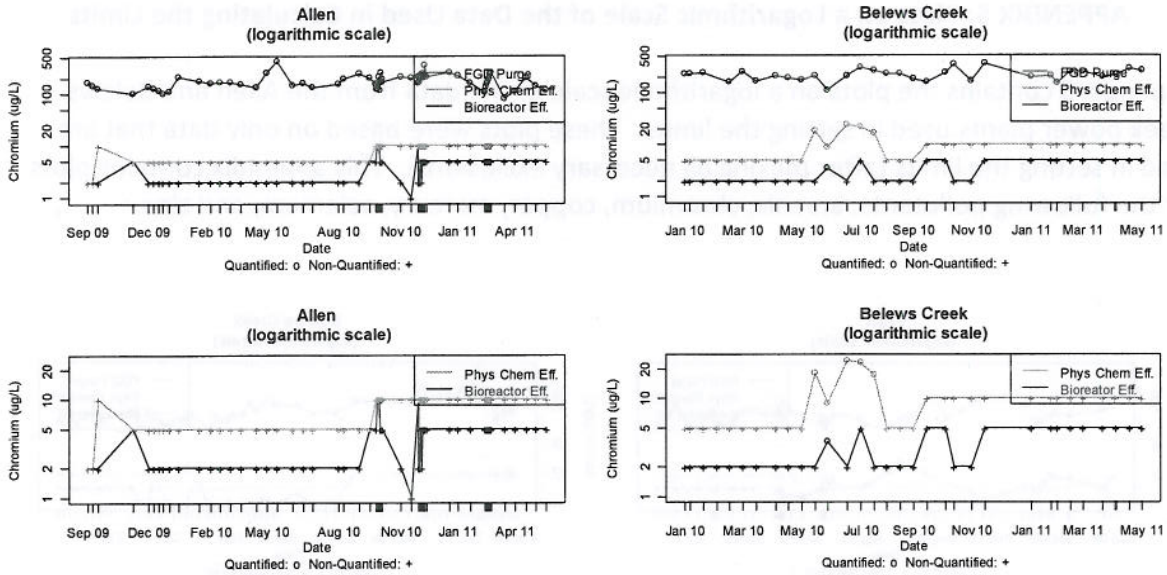

Figure 5.2. Plot of the chromium (ug/L) data on a logarithmic scale for Allen and Belews Creek after all data exclusions.

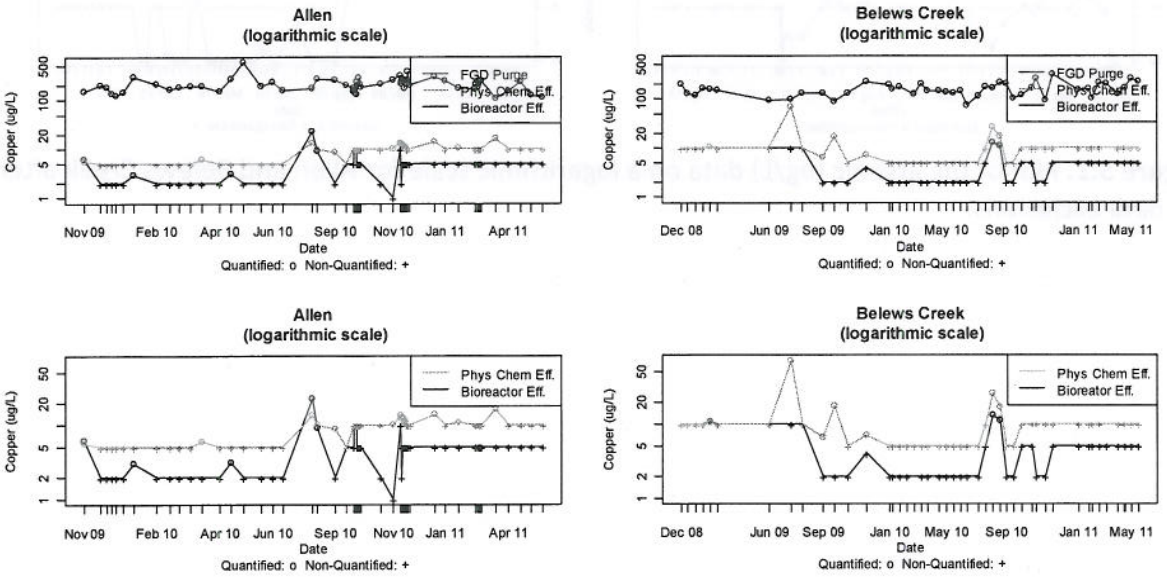

Figure 5.3. Plot of the copper (ug/L) data on a logarithmic scale for Allen and Belews Creek after all data exclusions.

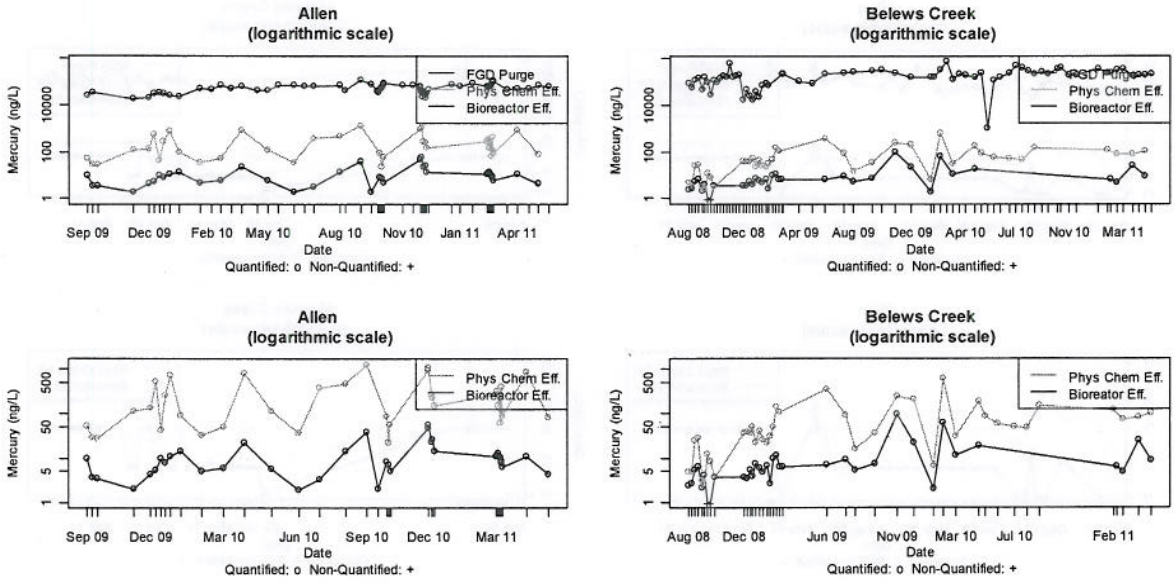

Figure 5.4. Plot of the mercury (ng/L) data on a logarithmic scale for Allen and Belews Creek after all data exclusions.

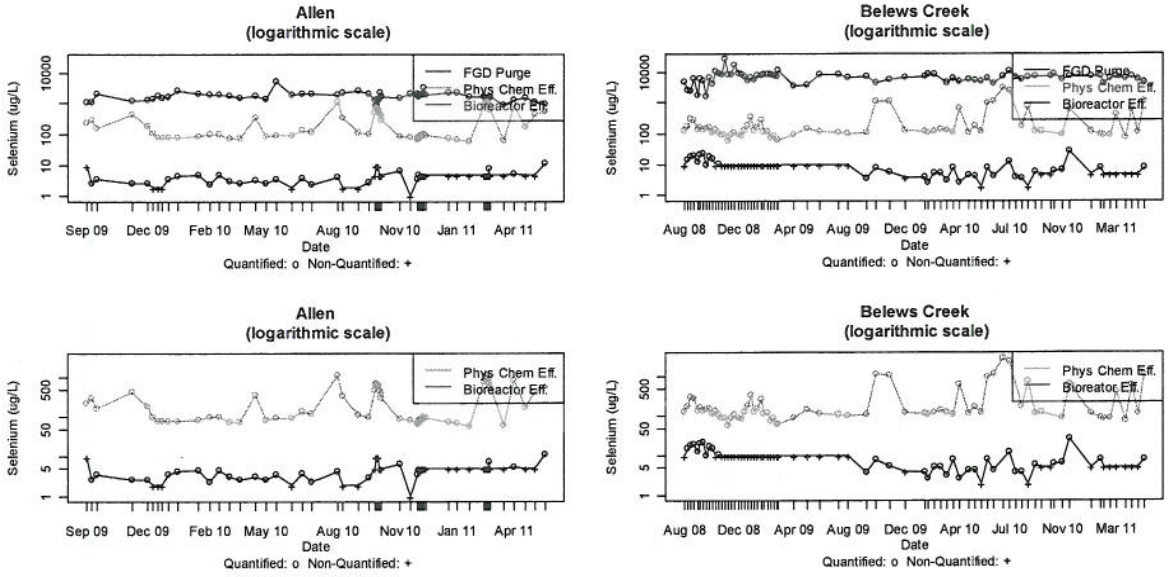

Figure 5.5. Plot of the selenium (ug/L) data on a logarithmic scale for Allen and Belews Creek after all data exclusions.

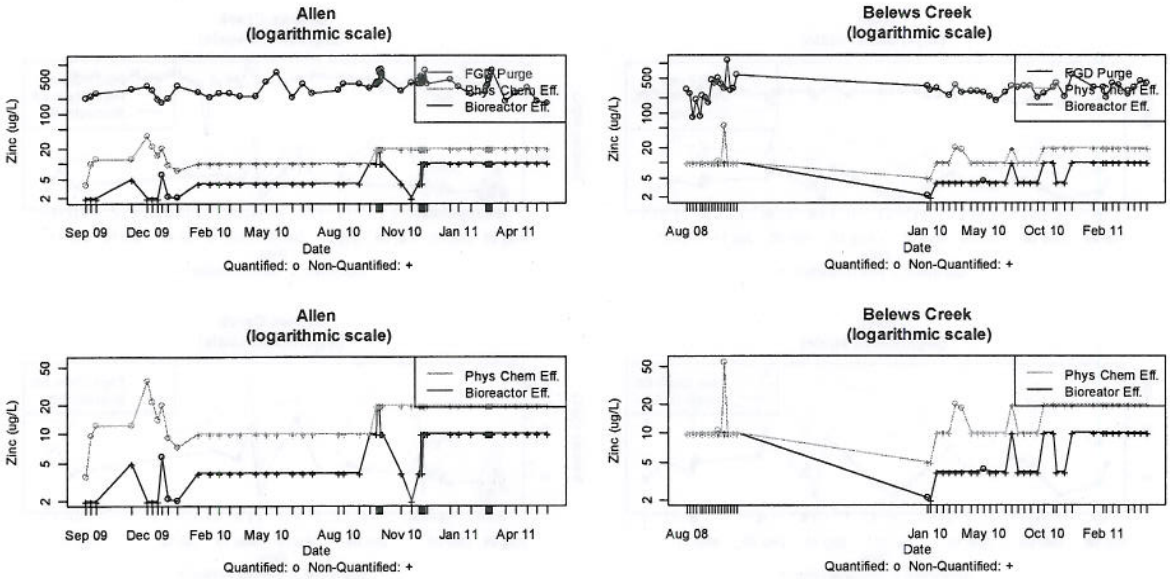

Figure 5.6. Plot of the zinc (ug/L) data on a logarithmic scale for Allen and Belews Creek after all data exclusions.

APPENDIX 6: Plots Evaluating Whether Autocorrelation is Present

This appendix contains plots of the autocorrelation and partial autocorrelation functions for the data from the five plant/analyte combinations that had at least 25 quantified observations. These plots do not suggest that an autocorrelation model is suitable for the data. See section III.4 for details on how to interpret the plots.

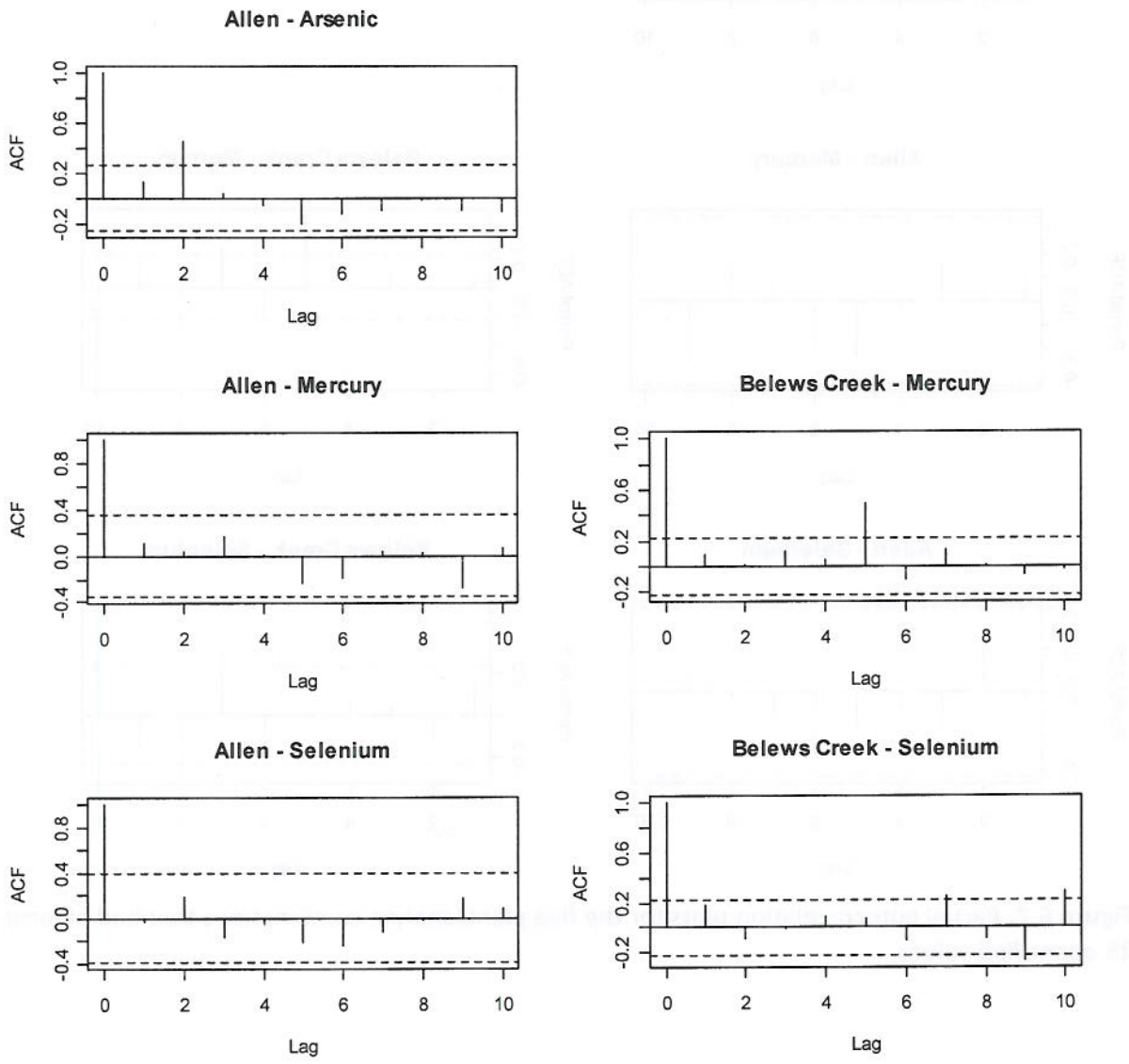

Figure 6.1. Autocorrelation plots for the five plant/analyte combinations that had at least 25 quantified values.

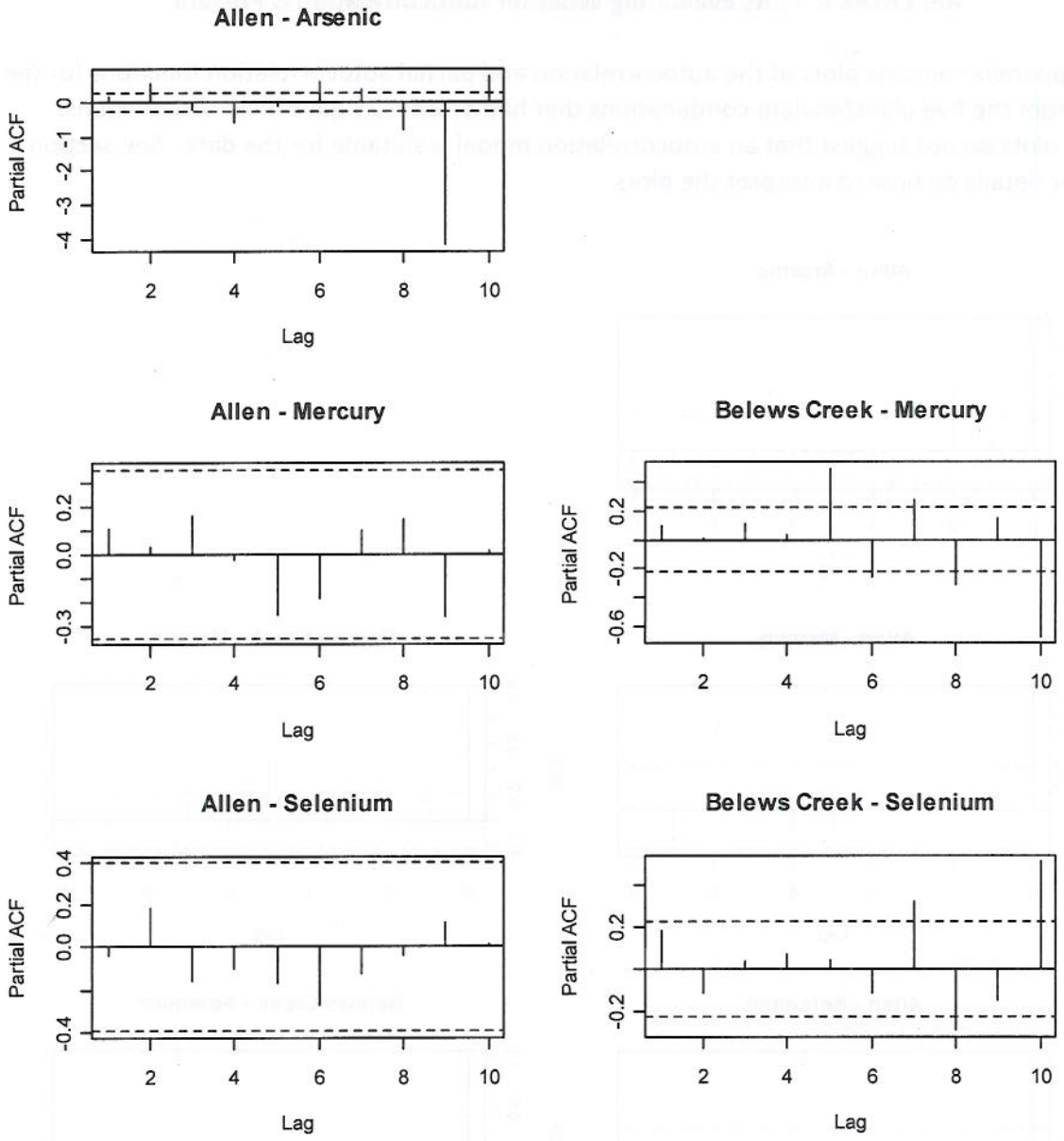

Figure 6.2. Partial autocorrelation plots for the five plant/analyte combinations that had at least 25 quantified values.