Point Discharge Effluent Nutrient Concentrations in the Delaware River Basin: A Preliminary Review

DELAWARE RIVER BASIN COMMISSION

Delaware River Basin Commission

DELAWARE • NEW JERSEY

PENNSYLVANIA • NEW YORK

UNITED STATES OF AMERICA

WWW.DRBC.NET

May 2014

Acknowledgements

This report was prepared by the Delaware River Basin Commission staff: Richard C. Gore, Acting Executive Director. John Yagecic was the principal author of the report. Mr. Yagecic is the Supervisor of the Standards & Assessment Section in the Modeling, Monitoring, and Assessment Branch and a licensed professional engineer. Other contributing authors include Chad Pindar, Supervisor of the Watershed Planning and Compliance Section, and Dr. Erik Silldorff, Aquatic Biologist. David Sayers developed a computer program for identifying and copying files matching user criteria used in this effort. Eric Wentz performed the manual accuracy screening of the data. Technical recommendations and support were provided by Thomas Fikslin, David Kovach, Karen Reavy, Namsoo Suk, and Pamela J. V'Combe.

Suggested Citation

Yagecic, J.R., Pindar, C., Silldorff, E. 2014. Point Discharge Effluent Nutrient Concentrations in the Delaware River Basin: A Preliminary Review. Delaware River Basin Commission. West Trenton, NJ. May 2014.

Background and Introduction

One of the ongoing challenges in crafting nutrient management approaches is quantifying the relative point discharge contributions to the overall nutrient load. An understanding of typical effluent nutrient concentrations is necessary for establishing default loads for water quality models and for assessing the achievability of proposed effluent nutrient limits.

In its web page entitled *Action towards Limiting Total Nitrogen, Total Phosphorus, and Total Inorganic Nitrogen Loads from NPDES-Permitted Facilities*, EPA documented the relative sparseness of nutrient monitoring data from NPDES permitted facilities (EPA, 2014). When facilities do monitor, querying data management systems that vary from state to state in both structure and inclusiveness can be a challenge.

In 2007, the Delaware River Basin Commission (DRBC) performed a large scale data query from EPA's Permit Compliance System (PCS) in an attempt to better understand the range and distribution of effluent nutrient concentrations. Since PCS is oriented toward major dischargers, and toward facilities with some compliance action, it is unclear how well these values represented expected effluent concentration ranges overall. In addition, treatment performance in climatically different areas may yield results that differ from what would be seen in the Delaware River Basin.

DRBC issues dockets (similar to NPDES permits) to wastewater dischargers throughout the Delaware River Basin. In order to demonstrate compliance with applicable effluent limitations and/or monitoring requirements, most modern DRBC wastewater dockets require the submittal of an annual report. This report is known as the annual effluent monitoring report (AEMR). Most wastewater discharge dockets for facilities located in the geographic portion of the Delaware River Basin known as "Special Protection Waters", which is essentially everything upstream of Trenton, NJ, are required to monitor for nutrients. The nutrient data is a critical enabling component for the Commission to implement its Special Protection Waters Program, which is likely the largest continuous stretch of anti-degradation waters in the nation.

DRBC has developed procedures for consolidating and interpreting this compliance monitoring data. We believe that this information provides valuable insight into typical effluent nutrient concentrations in the Delaware River Basin across a range of plant sizes and categories. Our expectation is that over time we will be able to develop refined effluent nutrient ranges by plant size and treatment type, thus improving both our understanding of expected concentrations and our ability to manage overall nutrient loads. There are limitations to interpretation of the data, however, associated with reporting of only monthly mean concentrations and with uncertainties at the high and low concentration ends of the range.

This report demonstrates that more work is needed in terms of defining relevant facility and treatment characteristics and data acceptance ranges. This preliminary assessment is still valuable, however, for providing a sense of central tendency effluent nutrient concentrations in the Delaware River Basin, and for demonstrating a process for compiling and interpreting this information.

AEMR Data

The AEMR data set consists of monthly mean flows and concentrations and loads of nutrients and related parameters from facilities in the Delaware River Basin submitting effluent monitoring reports to the DRBC. The compiled data set described in this report includes nutrient monthly mean concentrations from 164 facilities, spanning a time frame from 2007 through 2013. The number of facilities submitting AEMRs has increased over time, resulting in a higher density of data toward the end of that period. Figure 1 below shows a typical AEMR report form. Note that monitoring requirements differ from facility to facility so that most facilities are not required to monitor every parameter.

Such data can be useful for a number of questions, but uncertainties about and inconsistencies with these data warrant a careful and forthright presentation. To begin with, although DRBC requests the monthly means of the reported parameters, the reported values could vary from single samples (for a parameter measured just once monthly) to averages from 10s to 100s of values (for parameters measured daily or many times each day). Thus, the raw data in the AEMR data set are a mixture of observations, some of which greatly smooth the hour-by-hour variation in the treatment plant effluent and others which capture that short-term variability and fully demonstrate the variance in the raw data. In addition, recorded observations were not verified against original laboratory reports and a number of transcription errors could exist in the observations. Such transcription errors could neither be measured nor could they be corrected in this data set. Finally, it is not clear how each facility itself handled non-detects and other extreme data values before computing monthly means.

As a result of these data uncertainties, the following analyses focus on the central part of the data distribution extracted through this program. Specifically, both graphical and tabular analyses present statistics only from the 25th percentile to the 75th percentile of the data distribution, i.e., the central 50% of the data distribution. More extreme statistics and individual data points are not presented, and mean values are likewise not presented because of both the unknown weighing implicit in such averaging and because of the undue influence of mis-reported data and outliers on such derived statistics.

The data analyses therefore present a rich and robust analyses of typical operating conditions for these wastewater treatment plants in the mid-Atlantic region of the United States. They do not represent the full range of conditions, but do provide insight into the more typical concentrations achieved by treatment plants operating over many years and across a range of sizes and types.

Year:	2011	DRBC Annual Effluent Monitoring Report		DRBC Mail Control Slip Number (DRBC USE ONLY):	email 8/27/2012
Docket Holder:	XXXXX XXXXXXXXXX				Date Form Edited: 5/30/2012
Docket Number:	X-XXXX-XXX-X				
NPDES Number:	XXXXXXXXX				
Outfall Number:	1				
Docket Approval Date:	XX/XX/XXXX	Docket Expiration Date:	XX/XX/XXXX		

Docket holders should begin by entering the calendar year being reported. The docket holder's name should be the same as the one on the actual docket. Docket and NPDES permit numbers must be entered on the form, or the Annual Effluent Monitoring Report will not be associated with the docket and may result in a compliance issue.

Applicability: BW = Basinwide, SPW = Special Protection Waters, SS = Special Study		BW, SPW		BW, SPW	BW, SPW		BW, SPW		BW		BW	BW		SPW
Enter Monitoring Requirements from Docket	Average Monthly Flow	Total Suspended Solids		Fecal Coliform	Ammonia - Nitrogen		CBOD (5-day) or BOD (5-day)		pH		Color	Total Dissolved Solids		Dissolved Oxygen
Units	mgd	mg/l	lbs/d	cfu / 100 ml	mg/l	lbs/d	mg/l	lbs/d				mg/l	lbs/d	mg/l
Monthly Average Effluent Limit (Winter-Spring)	5.7	30	1426	200	2.22	106	10	475	6 -- 9					7
Monthly Average Effluent Limit (Summer-Fall)	5.7	30	1426	200	0.74	35	5	238	6 -- 9					7
Monitor & Report Only														
Enter Monitoring Results 2011	Report Monthly Average	Report Monthly Average		Report Monthly Average	Report Monthly Average		Report Monthly Average; Indicate CBOD5 vs. BOD5 CBOD5		Report Monthly Minimum	Report Monthly Maximum	Report Monthly Average	Report as directed in docket		Report Monthly Average
January	2.849	6.1	147.8	8	0.19	4.2	4.1	96.2	7	7.5				9.8
February	4.843	5.6	227.2	19	0.16	6.8	4.5	180.9	7	7.2				10.3
March	5.154	4.9	230.5	14	0.38	15.9	4.2	185.1	7	7.5				10.3
April	4.641	4.8	198.4	6	0.26	9.3	3.1	123.4	7.1	7.5				10.1
May	3.484	4.4	126.6	10	0.28	7.5	3.5	102.3	7.1	7.6				9.5
June	2.843	6.2	142.4	66	0.3	6.9	3.4	78.2	7.2	7.6				8.7
July	2.683	3.8	85.3	44	0.12	2.6	3	68.1	7.2	7.7				8.3
August	3.521	3.8	133.7	68	0.11	3.6	2.9	91.2	7.2	7.9				8.5
September	4.797	3.4	163.9	26	0.11	4.8	2.4	101.4	7.1	7.6				8.9
October	3.491	2.6	77.6	27	0.1	2.9	2	59	7.2	7.7				9.2
November	3.863	4.3	152.6	12	0.1	3.2	2.4	78.9	7.3	7.7				9.4
December	4.102	6.2	217	10	0.1	3.5	3.1	106.4	7.1	7.5				10.1

Applicability: BW = Basinwide, SPW = Special Protection Waters, SS = Special Study	SPW		SPW		SPW		SPW		OTHER		SS		SS	SS	SS
Enter Monitoring Requirements from Docket	Total Phosphorous		Total Kjeldahl Nitrogen		Nitrate - Nitrite - Nitrogen		Total Nitrogen				CBOD20		Specific Conductivity	Acute Toxicity	Chronic Toxicity
Units	mg/l	lbs/d	mg/l	lbs/d	mg/l	lbs/d	mg/l	lbs/d			mg/l	lbs/d	µS/cm	LC_{50} %	IC_{25} %
Monthly Average Effluent Limit (Winter-Spring)															
Monthly Average Effluent Limit (Summer-Fall)															
Monitor & Report Only															
Enter Monitoring Results 2011	Report Monthly Average		Report Monthly Average		Report Monthly Average		Report Monthly Average				Report Monthly Average		Report as directed in docket	Report as directed in docket	Report as directed in docket
January					10.6	210.5									
February					12	417.8									
March					10.2	358.3									
April					9.6	253.5									
May					8.5	226.1									
June					13.8	332.2									
July					15.3	335.8									
August					16.6	373.4									
September					9.6	501.8									
October					4.8	176.3									
November					9	371.8									
December					14.9	392.6									

Docket Holder Comments:

Figure 1. Sample AEMR Report

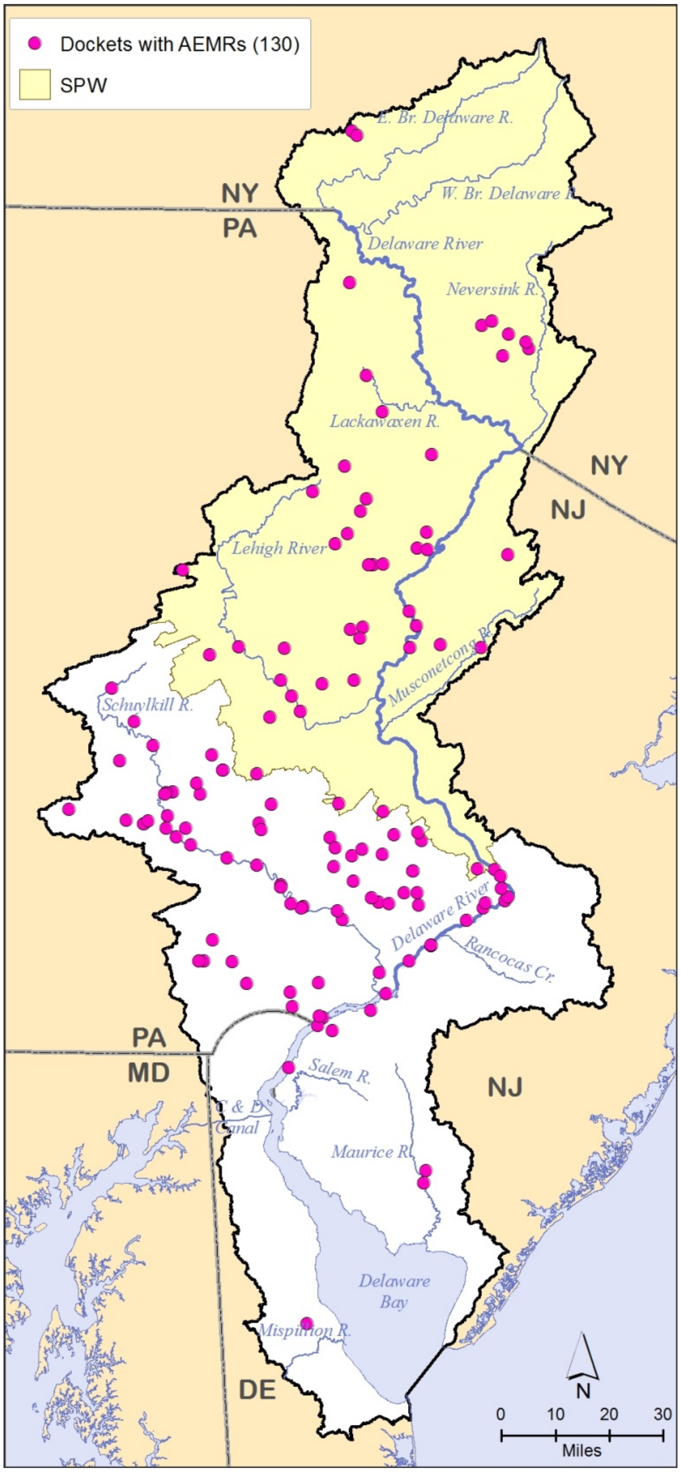

Figure 2. Map of Facilities in this AEMR Data Set

Data Caveats

The current AEMR data set may provide a sense of *typical* effluent values for wastewater treatment facilities by months of the year for the parameters reported. However, there are limitations to the data that users should consider:

- Monthly values provided by docket holders are monthly means, rather than discrete sample results. The discrete values used to derive the monthly means were not reported. In addition, the number of discrete samples contributing to each mean is likely to be different from facility to facility, and may be also be different by season. As such, it is probably infeasible to make inferences about the variability of discharge concentrations from this data;
- Defining the low concentration end of each parameter was hampered by the occasional reporting of non-detect values, and uncertainty regarding exactly what these non-detects represented. For example, if a facility was computing a monthly mean from four discrete observations, it is unclear how facilities would interpret a mix of non-detect and quantified observations for the same month. A detailed description of handling non-detect values is provided in the following section;
- Defining the high concentration end of each parameter was hampered by the occasional reporting of values that appeared to exceed likely upper limits based on best professional judgment.

As a consequence of the uncertainties at the high and low concentration ends of the range, we decided to focus on the interquartile range (25^{th} percentile, median, and 75^{th} percentile) to represent the central tendency of the data. The plots shown in Figures 3 through 10 are modified boxplots, excluding whiskers and outliers (outside the interquartile range).

Data Extraction and Processing

AEMRs are submitted by docket holders as formatted Microsoft Excel spreadsheets with data validation controls. Two internally developed computer programs were used to identify files matching the format of submitted AEMRs and copy the submitted data to a database. The first program crawled through all folders and subfolders potentially containing a submitted AEMR, and identified spreadsheets matching the AEMR format, regardless of file name. The second program copied the submitted effluent values into a centralized data spreadsheet.

Both programs were executed on March 17, 2014, capturing data submitted up that point.

Preliminary data cleaning was performed to facilitate subsequent interpretation. Data cleaning included:

- Blank or missing values were replaced with *NA*, the default nomenclature of the processing software;

- Obvious errors were replaced with *NA*. There were very few incidents of this type of replacement, usually resulting from a submitter having provided a note in the form cell, rather than a value;
- Non-numeric entries with no obvious interpretation (such as ***) were replaced with *NA*. Again, relatively few replacements of this type were required.
- Docket holders employed different formats to indicate monthly means below a reporting limit. In the data cleaning process, we addressed these edited values using the approaches below:
    - Entries reported as less than some value *X*, were set equal to that value *X*. For example, a monthly mean reported as <0.2 was replaced with 0.2;
    - Entries indicating an attempted measurement, for which no lower limit value could be inferred (such as "Not detected" or ND) were set equal to 0.
    - The need for replacement of edited data, as described above, occurred most frequently for Ammonia, with 220 incidences of "<" or "ND" replaced out of 3,296 reported monthly means (6.7%). Other parameters had fewer than 10 incidents of replacement each.

No upper end values were eliminated as outliers, although a handful of reported monthly means for each parameter appear to be unreasonably high. We reviewed the distribution of each parameter to determine if there was a threshold that would indicate obvious error for values above that threshold. This review was inconclusive, however, and we opted to retain all values and defer development of acceptable value ranges to future phases of work.

Data cleaning and initial processing was performed in Microsoft Excel. After cleaning, data interpretation and processing was performed using the R statistical programming software, as implemented in R Studio (Version 0.98.501 running R version 3.0.1 (2013-05-16)).

Manual Screening to Determine Accuracy of Automated Data Compilation

Since the data was harvested from individual reports using an automated process, we performed a manual screening of a subset of results to gain insight into the likely rate of error associated with the automated process. We randomly selected 5% of all compiled records for manual comparison to the original reports submitted. Of the 4,104 individual data points checked, 10 were identified as having been transcribed differently than what the docket holder had reported, for an error rate of 0.24%. In all instances, this error appeared to be the result of docket holder having submitted a different spreadsheet than the template provided.

Results and Discussion

Table 1 below shows quantiles and ranges of monthly submitted mean concentrations for Ammonia Nitrogen (mg/L), Nitrate-Nitrite Nitrogen (mg/L), Total Nitrogen (mg/L), CBOD (5-day) or BOD(5-day) (mg/L), Total Kjeldahl Nitrogen (TKN) (mg/L), Total Dissolved Solids (TDS), and Total Phosphorus (mg/L).

Although the ends of the each range indicate problematic values (0 at the low end and potentially erroneous values at the high end), we expect the interquartile range to be relatively immune from these influences, and to provide an empirically reasonable estimation of effluent concentrations for this region.

Table 1. Effluent Concentration Ranges from DRBC AEMR Reporting through March 2014

	Ammonia N (mg/L)	Nitrate + Nitrite (mg/L)	N Total (mg/L)	TKN (mg/L)	Phos. Total (mg/L)	TDS (mg/L)	BOD5 (mg/L)	CBOD5 (mg/L)
1st Quartile	0.11	0.96	1.27	0.6	0.2	200	2.44	2.0
Median	0.36	4.1	5.7	1.3	0.575	406.8	3.318	2.9
3rd Quartile	1	10.2	12.47	4.53	2.058	605	5.25	4.68
NA's	1264	3299	3656	3907	2690	2511	63	988
% facilities reporting this parameter	72.3%	27.7%	19.8%	14.3%	41.0%	44.9 %	76.4% (combined) Reporting either BOD5 or CBOD5	
% cleaned data	6.7%	0	0.7%	0.9%	0.3%	0	0	2.4%
Count of reported monthly mean values	3296	1261	904	653	1870	2049	333	3572

We developed box plots by month (excluding whiskers and outliers as described previously) for each of the parameters in Table 1, shown in Figures 3 through 10 below. In most cases, the range of the Y-axis was limited to facilitate inspection of the individual box and whisker diagrams. For ammonia, TKN, phosphorus, and BOD, some influence of season is apparent.

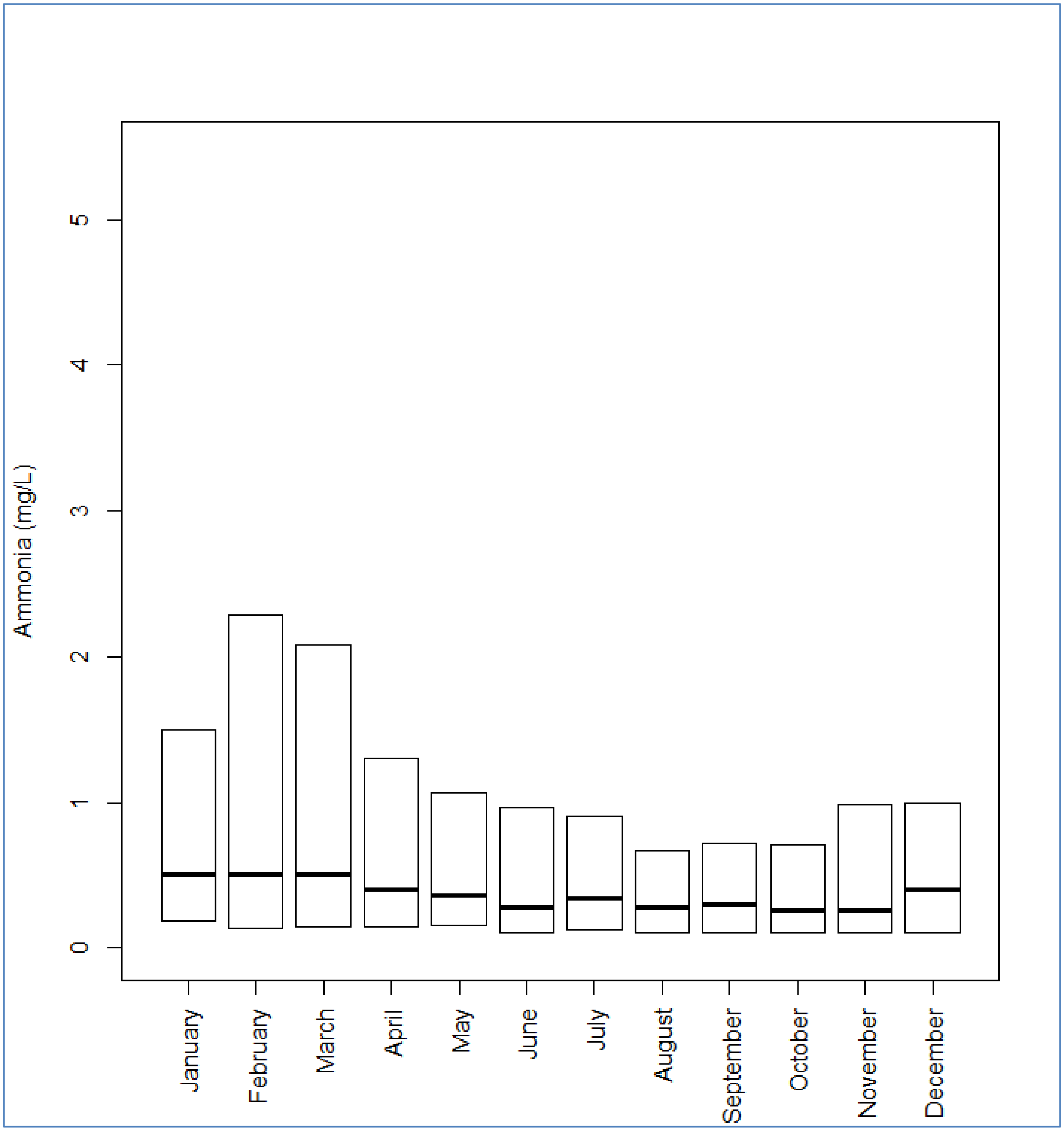

Figure 3. Box Plot of Ammonia by Month

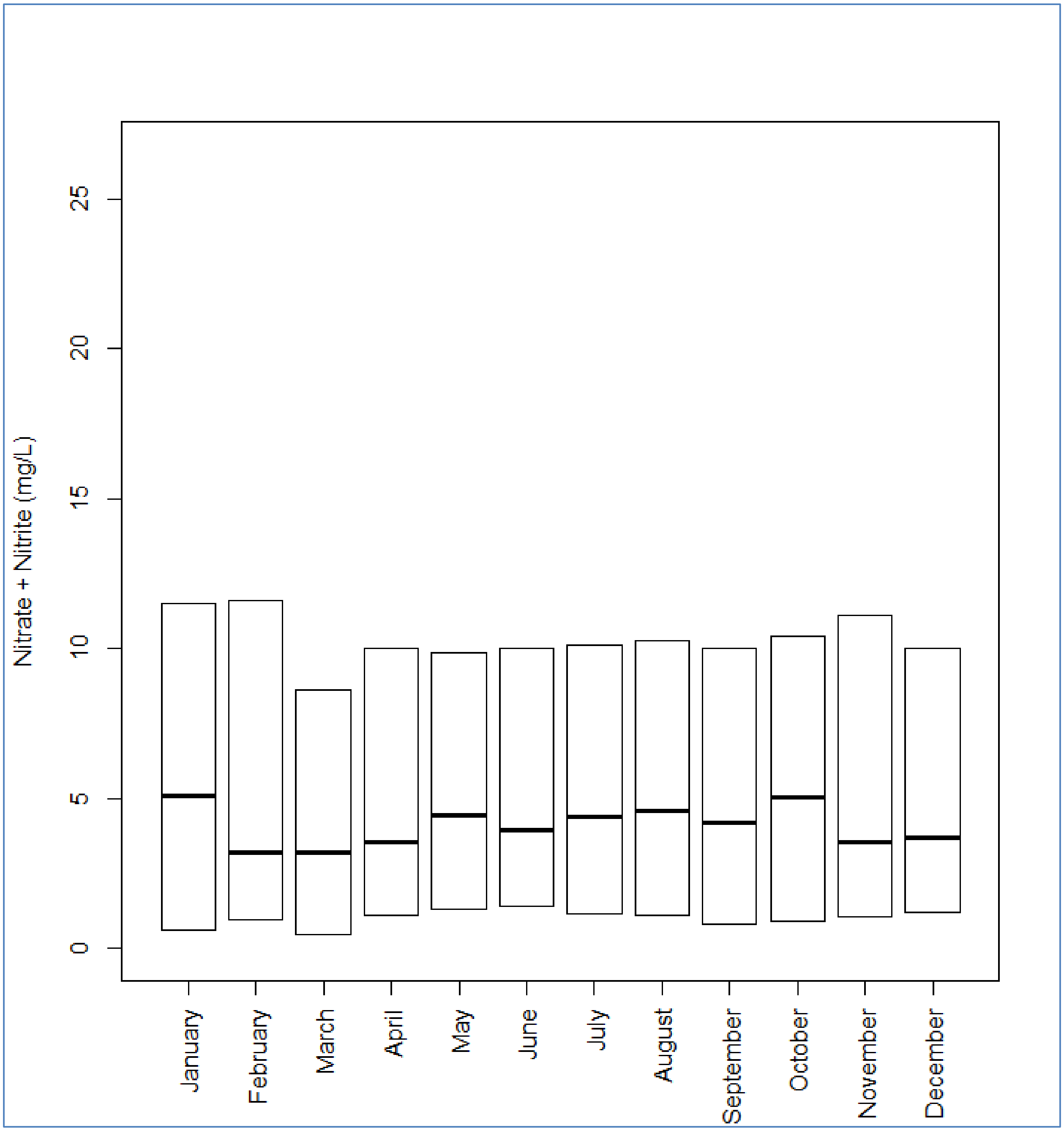

Figure 4. Box Plot of Nitrate + Nitrite by Month

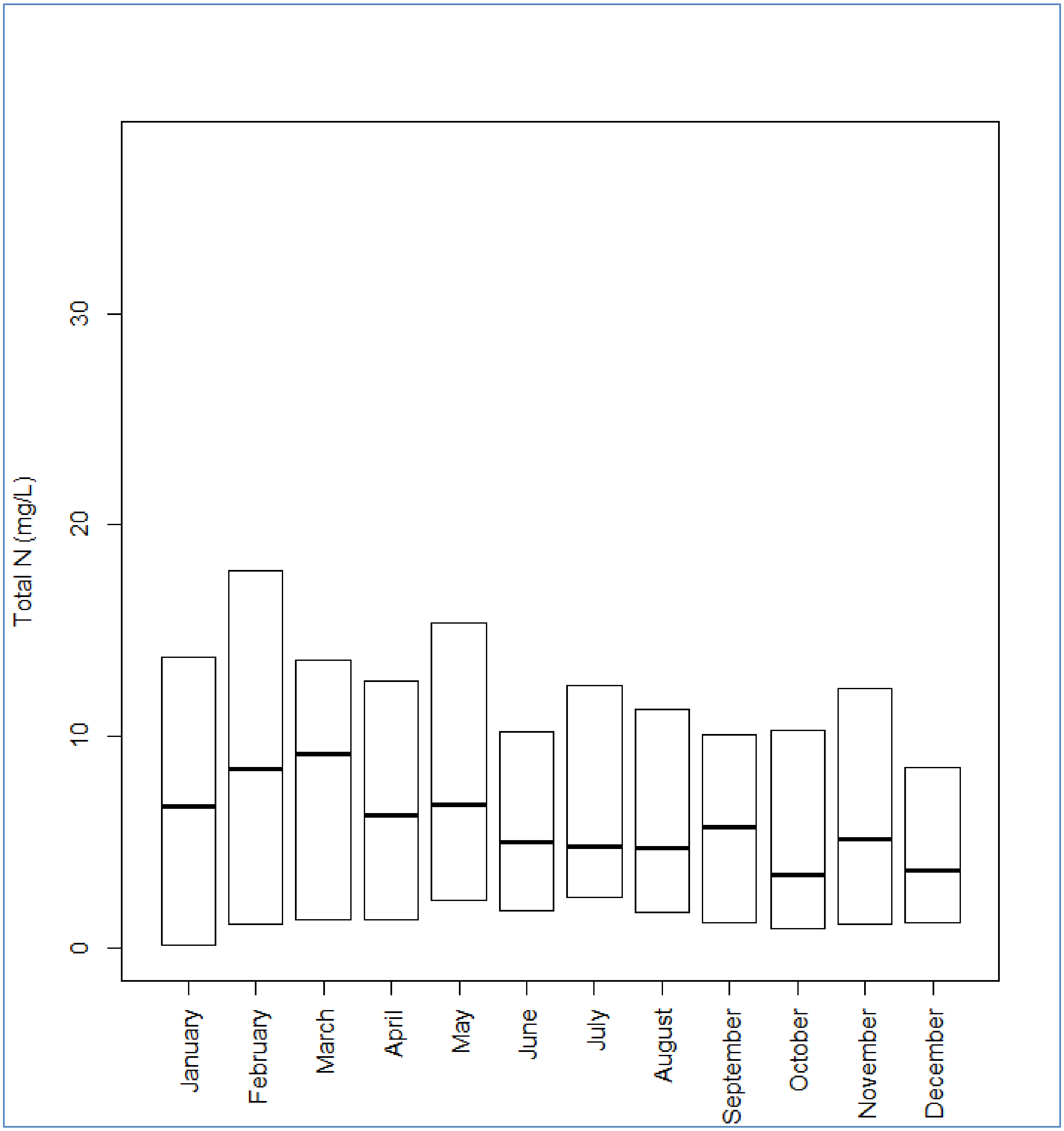

Figure 5. Box Plot of Total N by Month

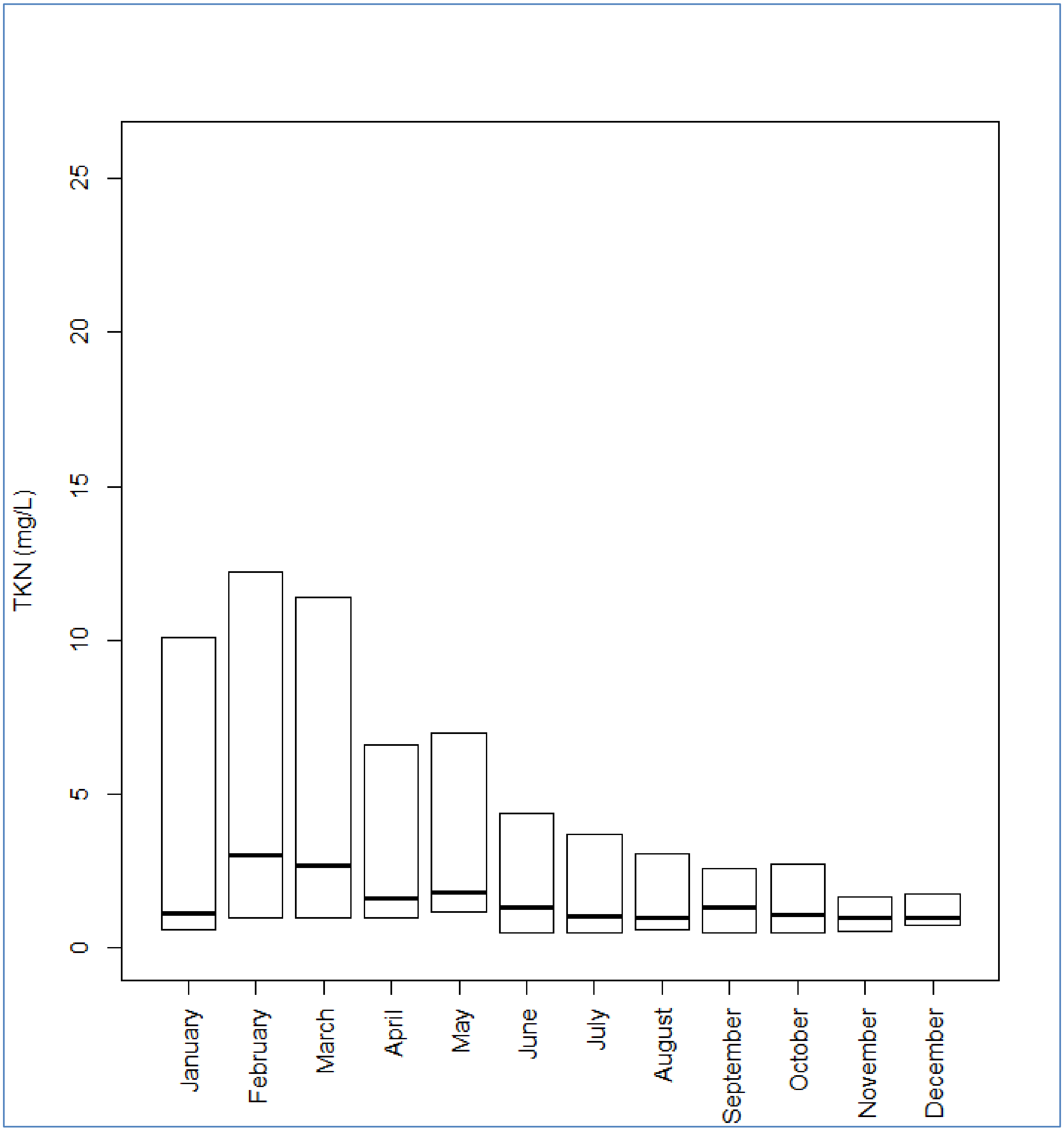

Figure 6. Box Plot of TKN by Month

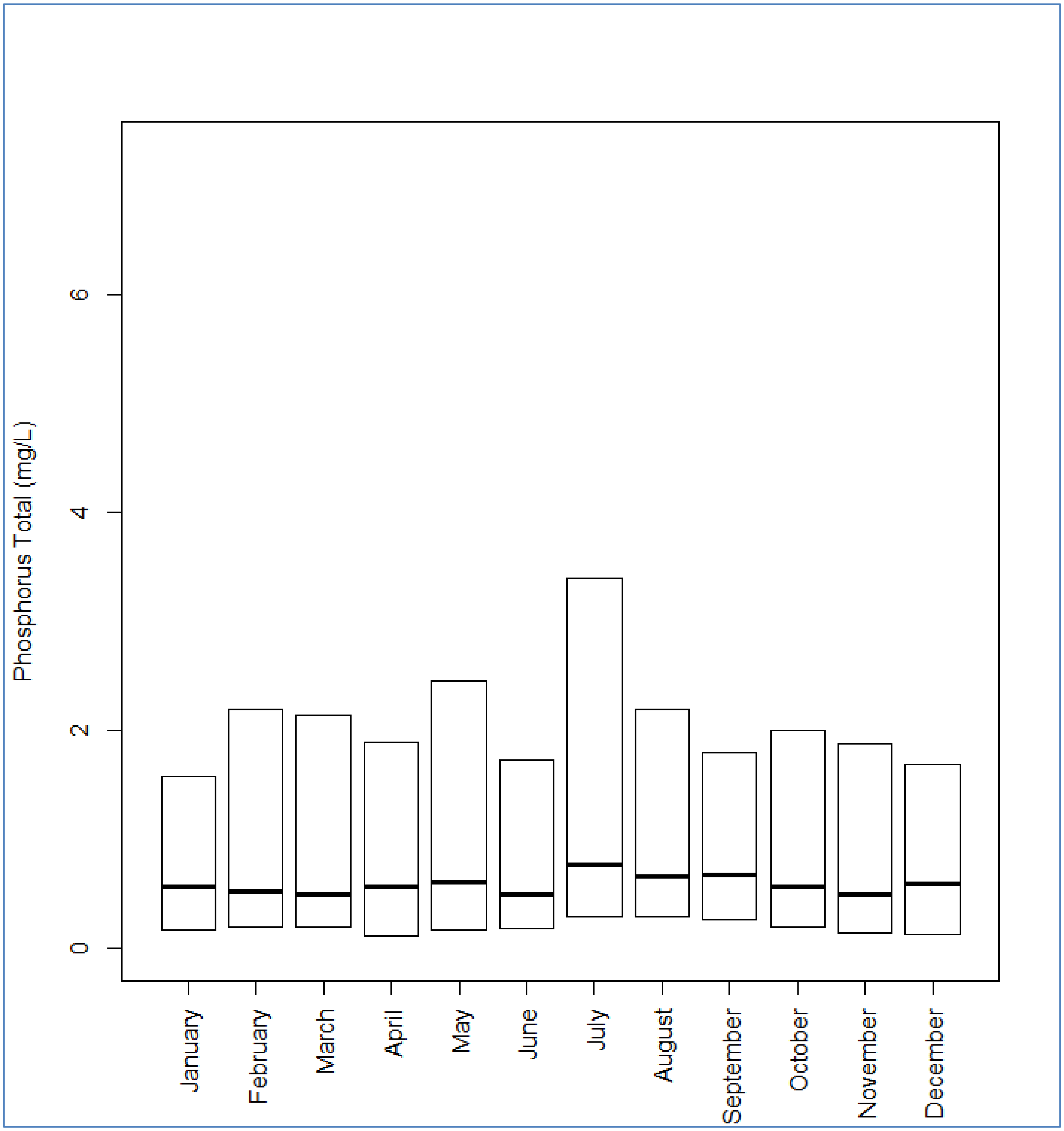

Figure 7. Box Plot of Total Phosphorus by Month

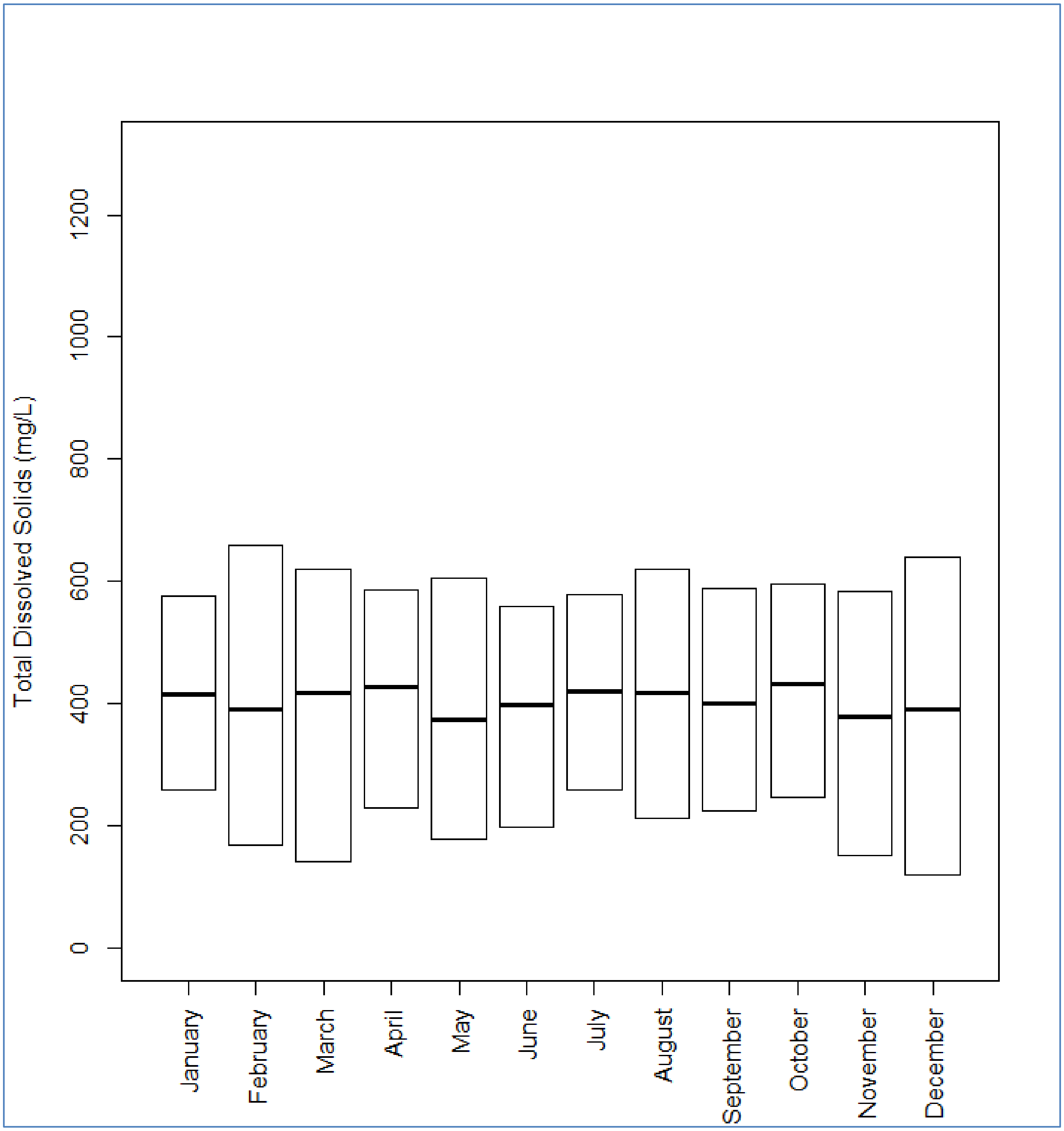

Figure 8. Box Plot of TDS by Month

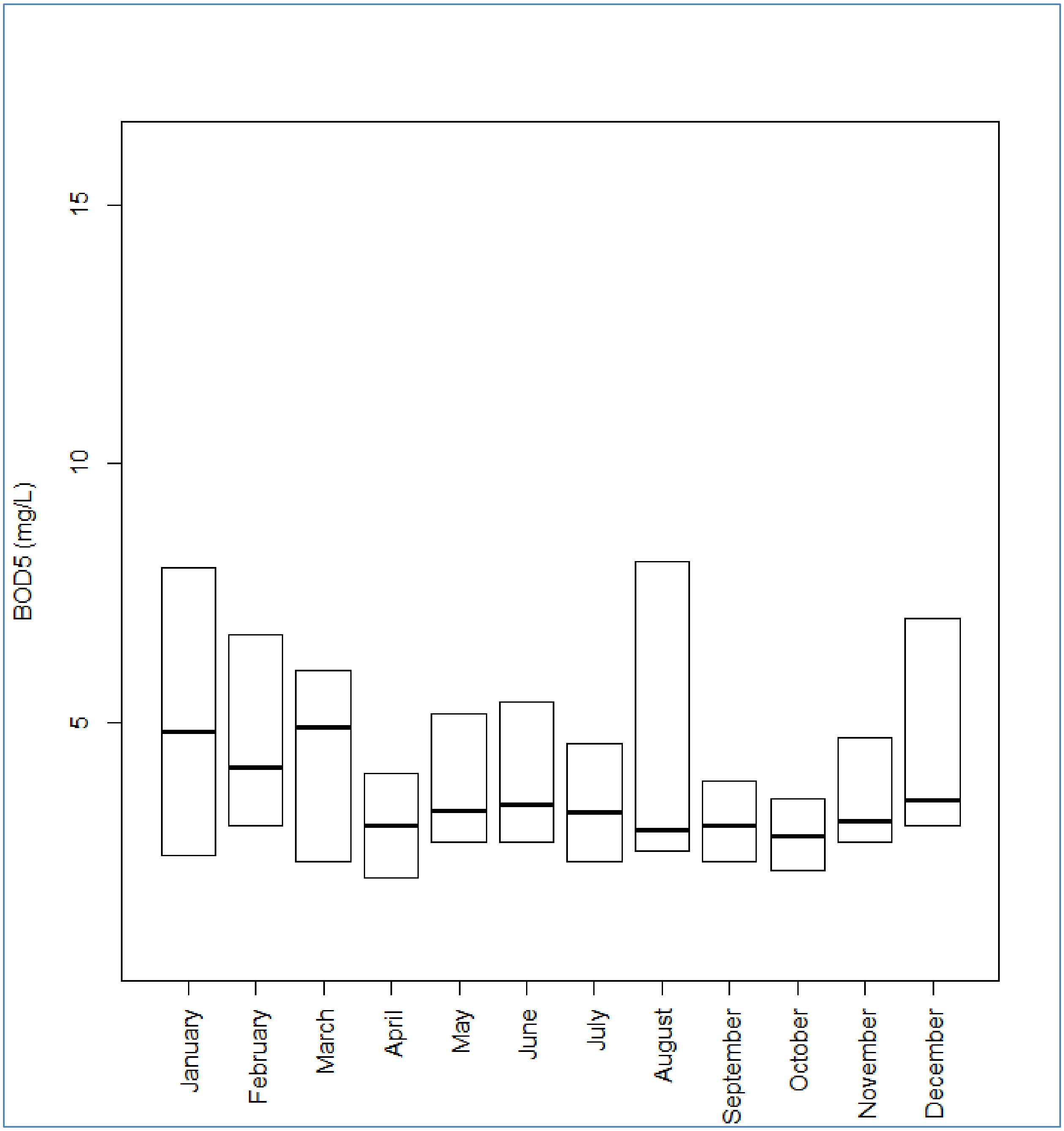

Figure 9. Box Plot of BOD5 by Month

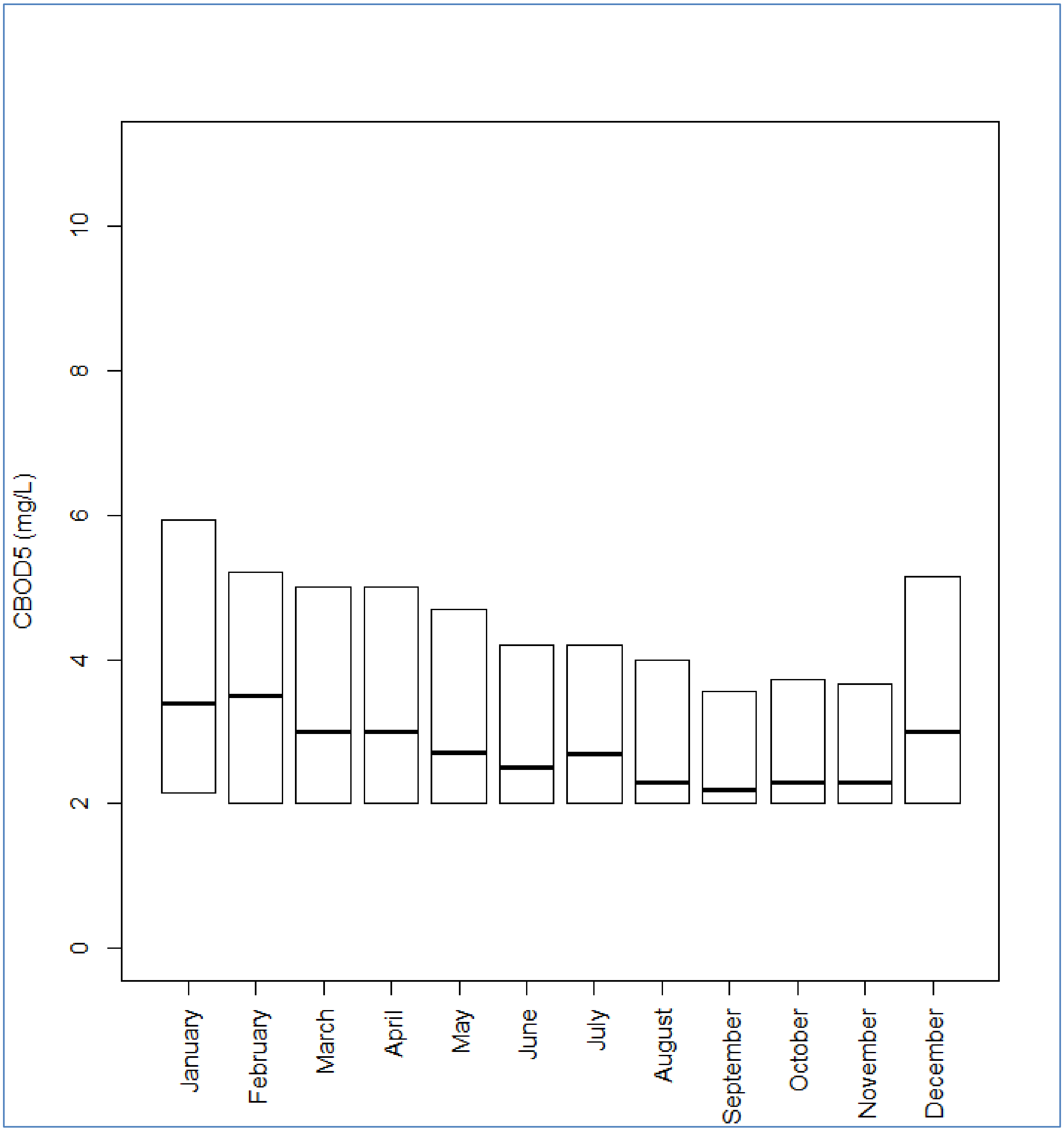

Figure 10. Box Plot of CBOD5 by Month

Recommendations for Future Work

Over the next year, DRBC staff propose to review Docket Files to extract information that will enhance the usefulness of the effluent concentration data. In particular, we expect to include in the data set information related to the following variables:

- Facility type, such as municipal or industrial waste water treatment plant;
- Secondary and tertiary treatment types; and
- Hydraulic design capacity and permitted flow for comparison to monthly mean flows.

In addition, we expect to investigate the following:

- Refined approaches for addressing edited data (below reporting limits);
- Refined approaches for flagging and investigating potentially erroneous data; and
- Development of completeness and QA checks to verify data compilation.

Our goal will be to produce a follow-up report approximately one year after the publication of this report to include the enhancements described above, as well as inclusion of new effluent data submitted during the interim period. Inclusion of facility descriptors and variables raises the possibility of more sophisticated analyses. Periodic follow up reports could be used to track the overall progress of managing effluent nutrient discharges in the Delaware River Basin.

References

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

RStudio (2012). RStudio: Integrated development environment for R (Version 0.98.501) [Computer software]. Boston, MA. Retrieved February 20, 2014.

USEPA (U.S. Environmental Protection Agency), 2014. Action towards Limiting Total Nitrogen, Total Phosphorus, and Total Inorganic Nitrogen Loads from NPDES-Permitted Facilities http://www2.epa.gov/nutrient-policy-data/action-towards-limiting-total-nitrogen-total-phosphorus-and-total-inorganic *Accessed* March 18, 2014.

Appendix A - R Script

setwd("~/AEMR Data Project") # This is the author's working directory. Other users' directories will be different

AEMR<-data.frame(read.csv("AEMRData03172014.csv"))

AEMR$Flow<-as.numeric(levels(AEMR$Flow))[AEMR$Flow]

AEMR$BOD5<-as.numeric(levels(AEMR$BOD5))[AEMR$BOD5]

AEMR$Month<-factor(AEMR$Month, levels=c("January", "February", "March", "April", "May", "June", "July", "August", "September", "October", "November", "December"))

summary(AEMR)

boxplot(AEMR$Ammonia ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3)

boxplot(AEMR$Ammonia ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3, ylab="Ammonia (mg/L)")

boxplot(AEMR$NitrateNitrite ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3, ylab="Nitrate + Nitrite (mg/L)")

boxplot(AEMR$Ntotal ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3, ylab="Total N (mg/L)")

boxplot(AEMR$TKN ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3, ylab="TKN (mg/L)")

boxplot(AEMR$PhosTotal ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3, ylab="Phosphorus Total (mg/L)")

boxplot(AEMR$TDS ~ AEMR$Month, whisklty=0, staplelty=0, outline=F, las=3, ylab="Total Dissolved Solids (mg/L)")

boxplot(BOD5 ~ Month, data=AEMR[AEMR$BODType=="BOD5",], whisklty=0, staplelty=0, outline=F, las=3, ylab="BOD5 (mg/L)")

boxplot(BOD5 ~ Month, data=AEMR[AEMR$BODType=="CBOD5",], whisklty=0, staplelty=0, outline=F, las=3, ylab="CBOD5 (mg/L)")