Updated Analysis of Dissolved Oxygen Concentrations at Barton Springs
SR-14-11, May 2014

Abel Porras, P.E.

City of Austin
Watershed Protection Department
Environmental Resource Management Division

Abstract

The Barton Springs complex is one of the most valuable and unique resources in Central Texas. For this reason, close monitoring of conditions in the springs, and attention to changing water quality is an essential function of the Watershed Protection Department. Periodically, the available water quality record is re-examined for parameters of interest. Dissolved oxygen data have been updated and analyzed in this report and compared to previous analyses. Regression analysis confirms a continuing downward trend in DO concentrations at the springs over time under specific hydrological conditions. Forecasting this result into the future and evaluating the biological relevance of this trend are recommended as next steps. Further analysis using vector autoregression is also recommended to determine relevant low DO levels given some time series of flow. Analysis is included here to update the regression curve estimates on the DO concentrations to springflow . The additional data since the previous report supports the previous determination of a fairly strong relation between DO and spring flow. Analysis of Exponential III curve regression was also updated from previous work. Further analysis for estimates of DO levels outside of the range of flow data obtained is recommended using Extreme Value Theory.

Introduction

The concentration of dissolved oxygen (DO) is generally accepted to be one of the most an important water quality parameters measured, functioning as an indicator of impacts from many organic and inorganic pollutants (Chapra, 2008). Furthermore, adequate DO is an essential condition for the survival, growth, and reproduction of various aquatic biota. Of special importance to the City of Austin is the relationship between DO and the health of the endangered Barton Springs Salamander (*Eurycea sosorum*) and the Austin Blind Salamander (*Eurycea waterlooensis*), which reside in the Barton Springs area. The City of Austin has been monitoring DO in Barton Springs since 1975. Studies of temporal trends for various water quality parameters at Barton Springs have been documented by the City of Austin (Turner, 2000, Herrington et.al. 2005, Johns 2006, Turner 2007, Herrington and Hiers 2010). This report compiles information collected by the City of Austin regarding DO at Parthenia Spring, Eliza Spring, and Old Mill Spring and updates these studies with current data. This compilation of data will serve to assist the

Barton Springs/Edwards Aquifer Conservation District in developing its Habitation Conservation Plan for the Barton Springs and Austin Blind salamanders required by the Endangered Species Act.

Background

Evidence of a significant degrading trend in dissolved oxygen concentrations at Barton Springs was first published in a 2000 City of Austin report (Turner 2000). Prior to that, no statistically significant temporal trends were identified (COA, 1997; Barrett, 1996; TNRCC, 1995). A 2005 City of Austin report (Herrington, et.al. 2005) confirmed this decreasing trend in dissolved oxygen levels and correctly predicted that if combined discharge from the Barton Springs complex remained low (as defined as spring discharge below 40 ft^3/s), then the DO levels would go below 5 mg/L by 2011. A 2006 City of Austin report identified the impact on surface counts of salamanders from low flow and decreasing DO (Johns 2006), and a 2007 City of Austin data report calculated a relationship between the natural logarithm of flow to the DO concentrations at the Barton Springs complex (Turner 2007).

The 2007 report also examined a number of curves for fitting the DO/Springflow relationship settling on a logarithmic curve out of seven that had similar estimates of correlation coefficient and standard error. It was stated that this was an arbitrary selection in lieu of a physical or theoretical reason for selecting one curve over another. One of the seven curves was an Exponential Association III curve. This curve has advantages for its incorporation of a maximum and a non-zero minimum consistent with the physical representation of DO. Therefore, Appendix C is provided as another analysis of this curve using updated data.

Finally, a 2010 City of Austin report expanded the previous analyses to include data from Old Mill Spring and Eliza Spring collected by in-situ instruments at 15 minute intervals. Flow rates measured for each of these reports were based on the combined discharge from the Barton Springs complex at USGS gage 08155500.

Methods

This report continues with the methods used by the previous authors in analyzing the data, and expands the analyses to include DO and flow rate measurements until April 2014. All data from 1981 and 1982 were omitted from the analysis due to a documented sewer line break. The resulting data set was partitioned into three groups: a storm influenced data set, a recharge influenced data set, and a non-recharge influenced data set. The storm influenced group includes data from dates in which a sample yielded either a fecal coliform bacteria count greater than 100 colonies per 100 mL or a Total Suspended Solid (TSS) concentration greater than 10 mg/L. The recharge influenced data set was determined by including samples where the mean daily flow at the Barton Creek at Loop 360 USGS gage (USGS 08155300) was not zero. Thus, the non-recharge influenced data set was classified by included those samples where the mean daily flow at the USGS gage was zero. This partitioning serves as an effective heuristic for examining water quality under these three major flow conditions (COA, 1997), and water quality under non-recharge conditions is thought to reflect the underlying long term changes in the aquifer water quality (Turner, 2000).

Analysis

Given the partitioned data using the above methods, several analyses were performed. First, linear regression coefficients were determined using time and flow rate as independent variables. Then, another linear regression analysis was completed using solely flow rate as an independent variable. These analyses were performed on samples that were not correlated with one another, such as in data collected from grab samples. When the samples were correlated with each other, as in continuous monitoring data, then more advanced time series analysis is needed. This report will only look to update the previous analyses using the grab sample data (that is, only perform the linear regression on the independent grab samples). A short discussion on the analysis of the data as a time series will follow the linear regression models.

Linear Regression using Time and Flow

Previous analyses relied upon multiple linear regression models, which were represented by:

$$DO = \beta_o + \beta_1 \cdot Date + \beta_2 \cdot BSFlow$$

Where *Date* is the number of days and *BSFlow* is the corresponding amount of instantaneous flowrate discharged from combined Barton Springs complex. To determine whether there is a linear relationship between the response variable, *DO*, and the regressors, *Date* and *BSFlow*, the model was tested for significance of regression. The corresponding hypotheses are:

H_o: $\beta_o = \beta_1 = \beta_2 = 0$
H_1: $\beta_j \neq 0$ for at least one j

Rejection of the null hypothesis implies that at least one of the regressors contributes significantly to the model.

Using data up to 2000, the following regression model was fit to data (Turner, 2000):

$$DO_{\text{WITHOUT RECHARGE}} = -0.00015 \cdot Date + 0.03 \cdot BSFlow$$

This indicates, that under no recharge conditions, the DO at Parthenia Spring will decrease by approximately 0.054 mg/L every year (-0.00015 * 365 days/year = -0.054 mg/L) regardless of flow rate. Since there was no regression equation under recharge conditions, it is implied that there was no temporal trend found in DO during times of recharge. Further, although it was never specifically stated, it was mentioned that the overall median of the DO data up to 2000 was about 5.7 mg/L. If one can assume that, in this case, the median is an estimate of the average, then the intercept value in the 2000 regression equation can be approximated to be higher than 5.7.

By 2005, City of Austin staff updated the 2000 analysis with additional grab sample data (Herrington, 2005). The resulting regression equations were:

$$DO_{\text{WITHOUT RECHARGE}} = 5.6 - 0.0001 \cdot Date + 0.03 \cdot BSFlow + 0.00001 \cdot MethodDate$$
$$DO_{\text{WITH RECHARGE}} = 6.3 - 0.000097 \cdot Date + 0.022 \cdot BSFlow$$

Thus, by 2005, under recharge conditions at Barton Springs, a slight decreasing temporal trend in DO was found. Under no recharge conditions at Barton Springs, the decreasing DO trend had tapered to about 0.036 mg/L every year (as opposed to 0.054 mg/L every year in 2000). However, the intercept had decreased to 5.6 (as opposed to something above 5.7).

Nine years of additional grab sampled data have been collected since the publication of the 2005 study. This report will use the additional data to update the results from the previous study. Using the same statistical techniques used in the 2000 and 2005 study, the grab data were partitioned into the same two sets. The resulting equations were:

$$DO_{\text{WITHOUT RECHARGE}} = 4.514 - 0.000032 \cdot Date + 0.031 \cdot BSFlow$$
$$DO_{\text{WITH RECHARGE}} = 7.12 - 0.0054 \cdot BSFlow$$

At this time, under recharge conditions, the temporal trend is again statistically indistinguishable from zero. And, under non-recharge conditions, the decreasing DO trend continues to be reduced (0.012 mg/L every year) and the intercept is now at 4.5. Thus, as the temporal trends in each study continue to approach zero, it appears that Barton Springs is reaching its minimum DO value of somewhere below 4.5 mg/L. The statistical results from this analysis are displayed in Tables 1 and 2 below.

Table 1: Regression Model Summary under Non-Recharge Conditions

	Estimate	*Std. Error*	*t-value*	*Pr(x > |t|)*
(Intercept)	4.514	0.185	24.43	$< 2\cdot10^{-16}$
Date	$-3.2 \cdot 10^{-5}$	$1.3 \cdot 10^{-5}$	-2.48	0.014
BSFlow	0.0314	0.0014	22.9	$< 2\cdot10^{-16}$

Table 2: Regression Model Summary under Recharge Conditions

	Estimate	*Std. Error*	*t-value*	*Pr(x > |t|)*
(Intercept)	7.13	0.327	21.81	$< 2\cdot10^{-16}$
Date	$-7.34 \cdot 10^{-7}$	$2.29 \cdot 10^{-5}$	-0.032	0.975
BSFlow	-0.0054	0.0026	-2.07	0.040

The updated linear model without recharge had an adjusted R^2 value of 0.65 and a statistically significant p-value. This indicates that the model is a good fit to the data. However, the linear model with recharge had a poor R^2 value (0.012) and a statistically insignificant p-value. This may indicate a poor fit for regressing time with DO during recharge conditions and/or an inappropriate model.

Nevertheless, comparing the non-recharge model with the 2005 results indicates that a decreasing trend in DO can still be stated for Barton Springs during non-recharge conditions. Over a twenty year period, the predicted change in DO concentration without recharge is expected to be at least 0.24 mg/L (= -0.000032 * 7300 days) lower than current non-recharge DO levels.

<u>Linear Regression Using Flow</u>

In 2007, City of Austin staff estimated regression curves fitting site-specific DO data to discharge from Barton Springs complex combined (*BSFlow*) (Turner, 2007). The resulting regression equation for Parthenia Spring, Eliza Spring, and Old Mill Spring are:

$$DO_{\text{PARTHENIA SPRING}} = -0.162 + 1.55 \cdot \ln(BSFlow)$$
$$DO_{\text{ELIZA SPRING}} = 0.0237 + 1.46 \cdot \ln(BSFlow)$$
$$DO_{\text{OLD MILL SPRING}} = 0.877 + 1.22 \cdot \ln(BSFlow)$$

Using the additional data up to 2014, the DO-flow regression curves were updated as follows:

$$DO_{\text{PARTHENIA SPRING}} = 0.07 + 1.51 \cdot \ln(BSFlow)$$
$$DO_{\text{ELIZA SPRING}} = 0.36 + 1.38 \cdot \ln(BSFlow)$$
$$DO_{\text{OLD MILL SPRING}} = 1.49 + 1.0 \cdot \ln(BSFlow)$$

These updated curves can be shown to have similar parameters to those determined by the 2007 report. In fact, the parameters in the 2007 report fall within the confidence intervals of those from this study. Table 3 below shows the confidence intervals of the coefficient estimates for Parthenia Spring, Eliza Spring, and Old Mill Spring.

Table 3: Confidence Intervals for Regression Equation Coefficient Parameters

Site Name	Coefficient	2.5%	97.5%
Eliza Spring	Intercept	-0.16	0.88
	Ln(*BSFlow*)	1.25	1.51
Old Mill Spring	Intercept	0.11	2.86
	Ln(*BSFlow*)	0.72	1.37
Parthenia Spring	Intercept	-0.31	0.45
	Ln(*BSFlow*)	1.42	1.61

However, these parameter estimates are valid for only the range specified by the data. That is, estimating the DO concentration from flow below 14 ft^3/s or above 129 ft^3/s (which are the minimum and the maximum, respectively, of the data) would result in greater uncertainty. It remains to be seen whether the linear trend (in logarithmic scale) would continue at flow rates outside the minimum and maximum values. Instead, it is recommended that for such estimation, Extreme Value Theory be used which makes inferences on DO measurements based on the probabilities of exceeding the maximum flow value.

Time Series Analysis

The above updated analyses were performed in accordance with the methods used in previous analyses in order to compare any significant changes in the trends. A more extended analysis can be implemented for data that are autocorrelated, such as in the continuous time series found in the 2010 report. From this analysis, future values of DO or flow can be estimated based on previous values of DO or flow due to the potential autocorrelation between successive values. A time series was obtained from a data sonde, which recorded flow and DO at 15-minute intervals. Analysis showed that the two time series are cointegrated (that is, the information on one time series can be used to predict future time values of the other time series by examining correlations between the two series). However, that was the extent of the information that was collected from this analysis.

Conclusions

This report updates analyses performed on the continuing collection of DO and flowrate data obtained at Barton Springs. The analysis in this report confirms a continuing downward trend in DO concentrations at the springs over time under non-recharge conditions. Over the course of 20 years, DO concentrations during non-recharge conditions are likely to be at least 0.24 mg/L lower than under current non recharge conditions. However, DO concentrations during recharge conditions have remained constant over time. Further analysis should include vector autoregression to predict future DO concentrations given some time series of flow. The regression curve estimates on the DO concentrations to flowrates were updated, and the additional data supports the previous determinations of a fairly strong relation between the two. However, for estimates of DO levels outside of the range of flow obtained, Extreme Value Theory should be investigated in further analyses to examine the probabilities related to DO at values outside the flow measurements.

References

Barrett, M.E., and Charbeneau, R.J. 1996. A Parsimonious Model for Simulation of Flow and Transport in a Karst Aquifer: The University of Texas at Austin, Center for Research in Water Resources. Technical Report 269. 149 p.

Beirlant, J. et al. 2004. Statistics of Extremes: Theory and Applications. John Wiley and Sons, Inc. Hoboken, NJ

Chapra, S. C. 2008. Surface Water-Quality Modeling. Waveland Press, Inc. Long Grove, IL.

City of Austin. 1997. The Barton Creek report. City of Austin, Drainage Utility Department, Environmental Resources Management Division. Water Quality Report Series COAERM/ 1997. April 22, 1997. Austin, Texas.

Herrington, C. and S. Hiers 2010. Temporal Trend Analysis of Long Term Monitoring Data at Karst Springs, 2009. City of Austin Watershed Protection Department. SR-10-06.

Herrington, C. et. al. 2005. Update of Barton Springs Temporal Trend Analysis-2005. Water Resource Evaluation Section, Environmental Resource Management Division, Watershed Protection & Development Review Department, City of Austin. SR-05-09. August 2005.

Johns, D. 2006. Effects of low spring discharge on water quality at Barton, Eliza and Old Mill springs, Austin, Texas. City of Austin Watershed Protection Department. SR-06-05.

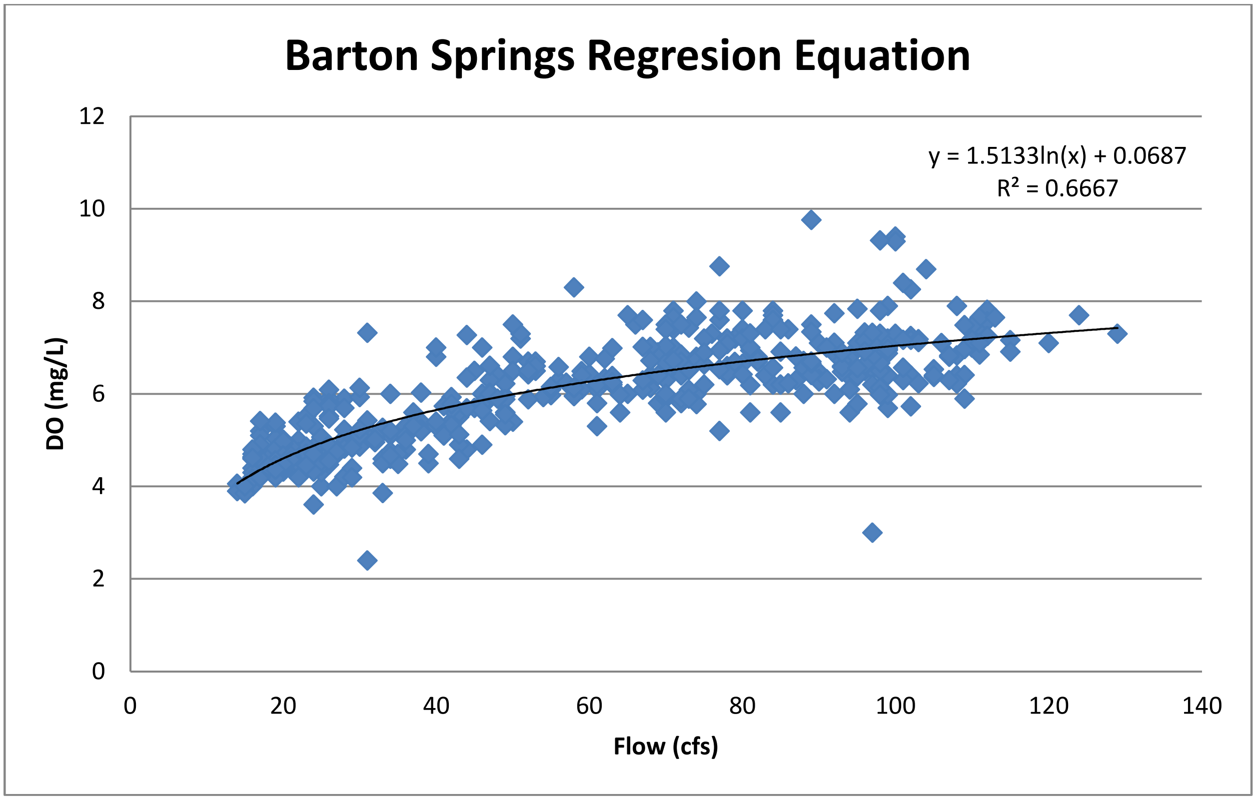
TNRCC 1995. Letter from Mark Jordan, Director Water Quality Policy and Regulation Division to Jana Grote, USFWS concerning proposed listing for the Barton Springs Salamander and absence of material decline in water quality of Barton Springs over 20-year period of nine studies between 1976 and 1994. May 17, 1995.

Turner, M., 2000. Update of Barton Springs Water Quality Data Analysis-Austin, Texas. City of Austin Watershed Protection Department, Environmental Resources Management Division. SR-00-03.

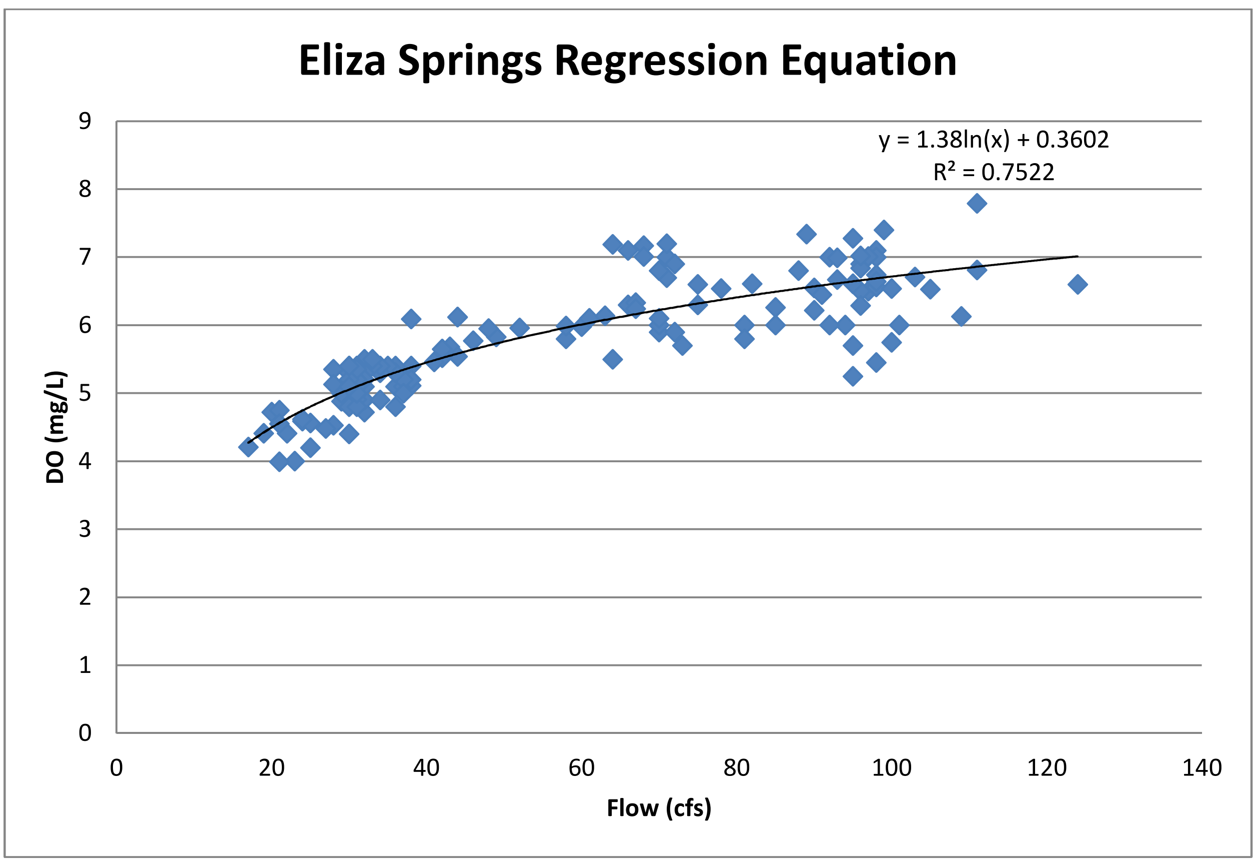
Turner, M. 2007. Data Report for the Barton Springs and Austin Blind Salamanders - 2006. City of Austin Watershed Protection Department. DR-07-01.

Appendix A

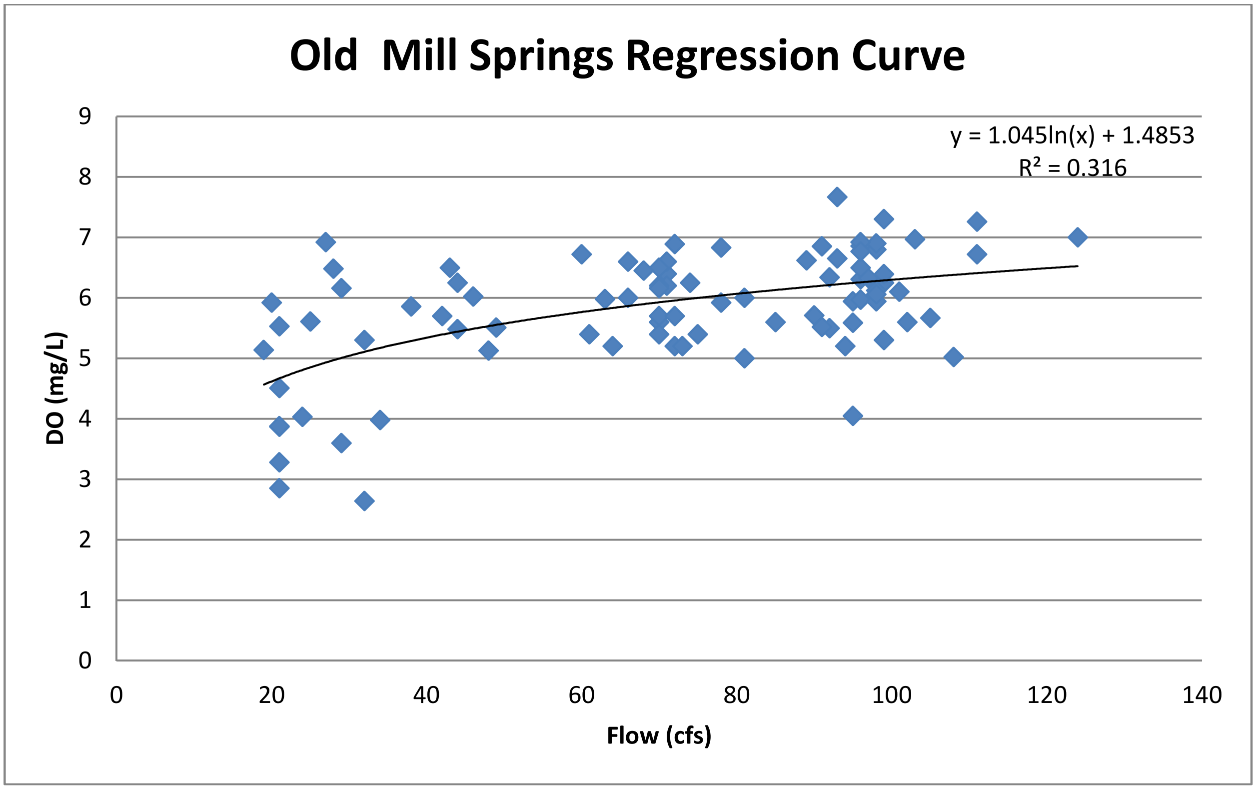
Plots of Regression Equations and Data
For
Barton Springs, Eliza Springs, and Old Mill Springs



Barton Springs Regresion Equation
y = 1.5133ln(x) + 0.0687
R² = 0.6667
DO (mg/L)
Flow (cfs)
0
2
4
6
8
10
12
20
40
60
80
100
120
140



Eliza Springs Regression Equation
y = 1.38ln(x) + 0.3602
R² = 0.7522
DO (mg/L)
Flow (cfs)
0
1
2
3
4
5
6
7
8
9
20
40
60
80
100
120
140


Old Mill Springs Regression Curve
y = 1.045ln(x) + 1.4853
R² = 0.316
DO (mg/L)
Flow (cfs)
0
1
2
3
4
5
6
7
8
9
0
20
40
60
80
100
120
140

Appendix C: Regression Equations under the Exponential Association III Curve

The City of Austin (Turner 2007) showed that the regression of DO to flow can be quantified by at least seven different, yet equally, valid family of curves. Each of the curves has their advantages and shortcomings; however, the Exponential Associate III Curve is notable among the curves for its incorporation of a top plateau (a maximum) and a non-zero bottom plateau (a minimum). These characteristics are useful because the physical representation of DO will consist of a maximum concentration and, potentially, a minimum concentration above 0 mg/L. The minimum DO is also the more biologically relevant given that it affects health of aquatic life. This appendix will look more closely at the parameters accompanying the Exponential Association III curve and its variability given the data; however it makes no preference for one family of curves over another.

The Exponential Association III curve is represented by the equation

$$y = a(b - e^{-c \cdot x}) \qquad \text{C.1}$$

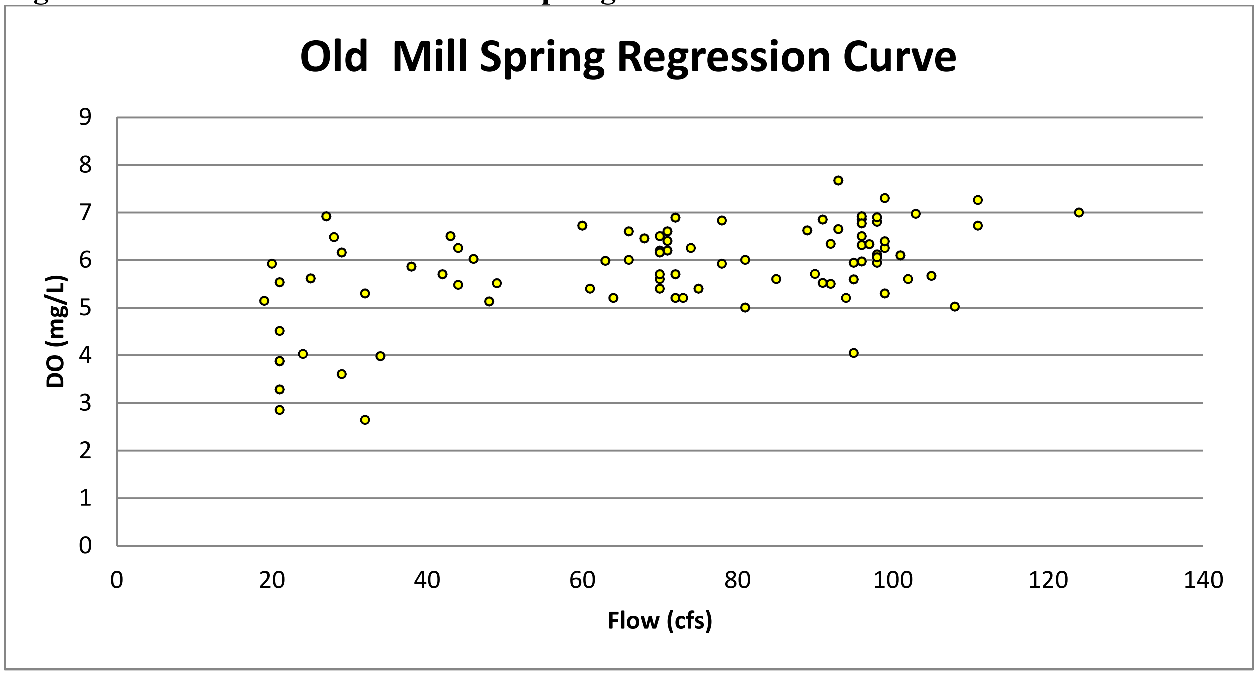
The x and y represent the flow in ft^3/s and DO concentration in mg/L, respectively. The parameters a, b, and c specify the curve that best fits the (x,y) data. Note that Equation C.1 is non-linear with respect to the unknown parameters, which makes the parameter estimation difficult to solve. These parameters are typically determined by iterating between various values of a, b, and c to find the smallest sum of squared errors between the model and the observed data. As a result of the non-linearity, uniqueness of the solution is not assured. To obtain uniqueness and examine the appropriateness of the parameter estimates, this appendix will first look at the possible qualitative solutions, which will then be followed by the quantitative solution derived by iteration.

Qualitative Approach

To look at the Exponential Association III curve qualitatively, we first deconstructed the equation in recognition of the physical representation of DO and biological significance as previously mentioned. First, at $c = 0$, $y = a \cdot b - a$, and at very large c, $y = a \cdot b$. Therefore, a represents the difference between the top and bottom plateau. Thus, for the DO-flow relationships given in Appendix B, a value of a less than 8 mg/L DO is to be expected with the product of a and b approximating the top plateau. Furthermore, the c parameter functions as both a weighting operator and as a decay rate. That is, as c approaches 0, $-c \cdot x$ also approaches 0, which gives little to no weight to the data set, x. The c parameter also fills the role of the specifying the rate of decay of a as flow increases so that as c gets large, $-c \cdot x$ also gets large, but $e^{-c \cdot x}$ approaches 0. Thus, the value of c must fall within some range that provides adequate weight to the data set without excessively decaying the data at large values of x.

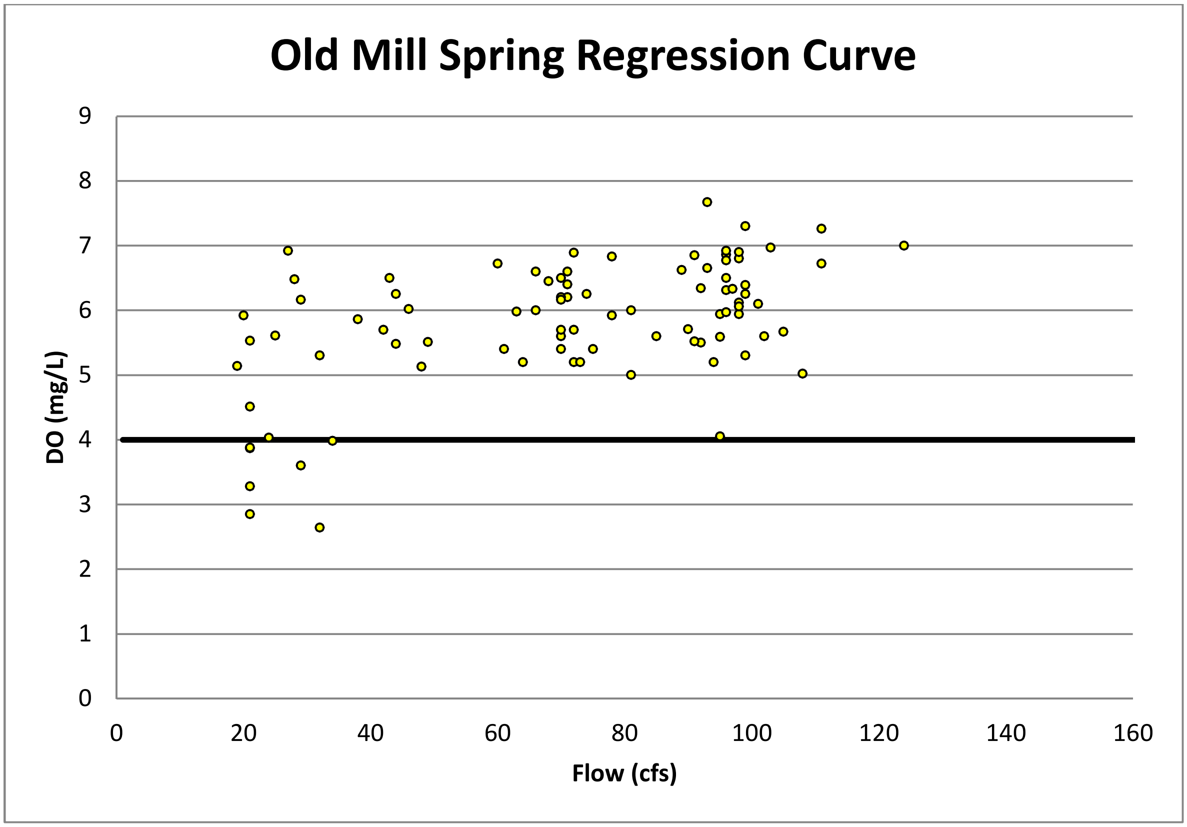
To see an example of this qualitative approach as applied to the data sets, consider the Old Mill Spring DO-flow relation (Figure C.1 below). It appears that the top plateau is approximately 8. Then suppose (for illustration purposes) that the bottom plateau is 4. Then by the approach given above, the starting parameters for iteration are $a = 4$ and $b = 2$.

Figure C.1 – Data Points for Old Mill Spring



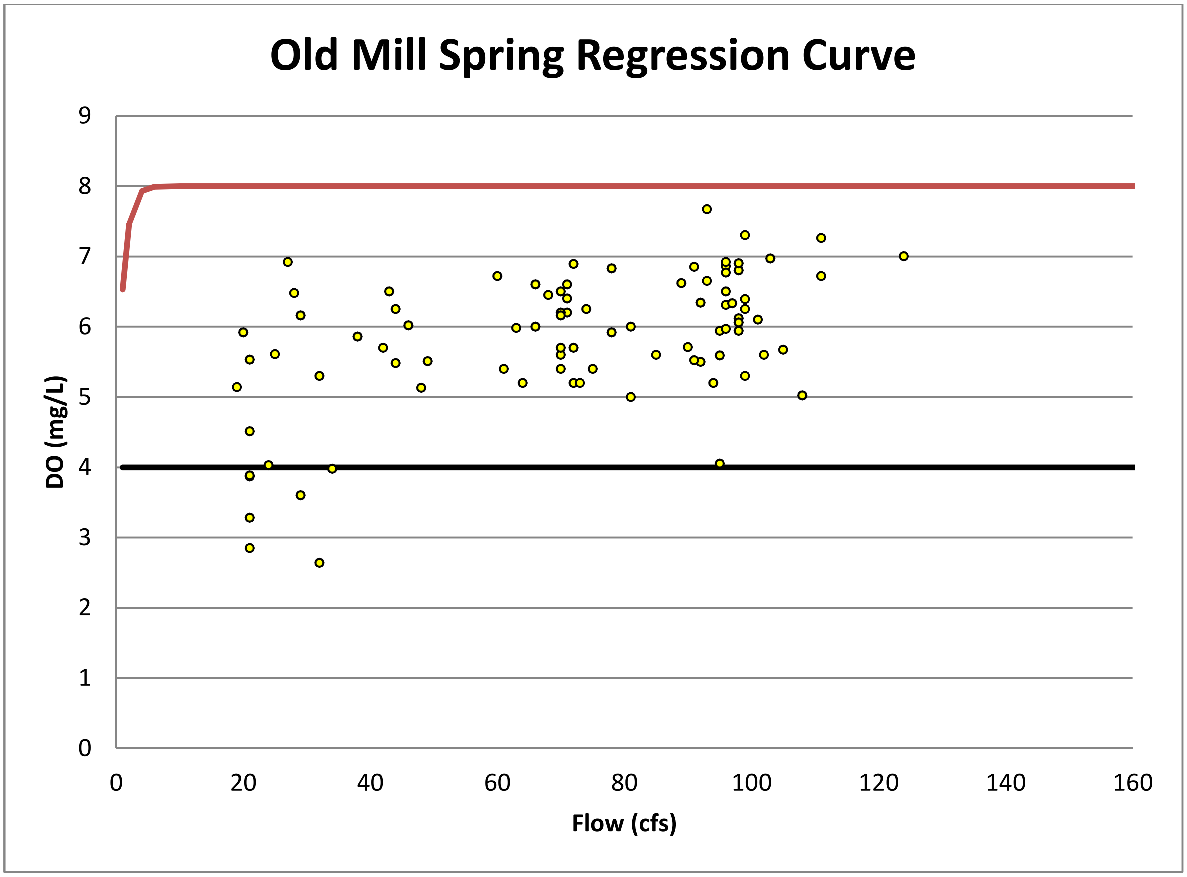
To establish an appropriate range of *c*, first consider the Exponential Association III curves at $c = 0$, shown in Figure C.2 below. The resulting curve is $y = 4$, regardless of the value of x.

Figure C.2 – Data Points with Exponential Association III curve at c = 0



Next, consider the Exponential Association III curves at $c = 1$, shown in Figure C.3 below. The resulting curve can be approximated by the equation $y = 8$, shown in red.

Figure C.3 - Data Points with Exponential Association III curve at c = 1



Note that even at a flow of 4 ft^3/s, the Exponential Association III equation gives a DO value of 7.9 mg/L and stays constant at values of flows greater than 4 ft^3/s. Thus, the model with a c value of less than 0 or greater than 1 is not useful in describing the data, and a range for *c* can be expected to fall between 0 and 1. In fact, each of the spring's regression curves all fall within a *c* value of between 0.01 and 0.03.

<u>Quantitative Approach</u>

With a good understanding of the bounds of the parameter estimates, a quick quantitative analysis of the non-linear regression model for Exponential Association III follows. Equation C.1 (from here on, *the model*) above provides the relationship between flow and DO, which can be compared to the observed data points. The difference between the model estimate of DO ($a(b - e^{-cx_i})$) and the observed data ($\hat{y}_i$) is called the error, and the sum of the squared errors is represented by the equation.

$$Q = \sum_{i=1}^{n} [\hat{y}_i - a(b - e^{-cx_i})]^2 \qquad \text{C.2}$$

The goal of regression (both linear and non-linear) is to estimate the regression parameters *(a,b,c)* that minimize the sum of the squared errors, *Q*. For linear regression, this estimation is *well-posed*. That is, the solution exists, is unique, and is *continuously dependent on the data*[1].

[1] The term "continuously dependent on the data" means that small variations in the parameter estimates will not result in large variations in the behavior of the solution.

Furthermore, the solution can be obtained with some algebra. For non-linear regression, estimation is not necessarily well-posed, and the solution cannot always be manipulated through algebra. However, there are two methods to obtain a solution. The first is by going through all possible combination of parameters and finding the smallest squared error through brute force. The second method, called the method of steepest descent, uses calculus to compute the gradient of Q with respect to each of the parameters to find the local minimum. The method of steepest descent was used, which came up with the following parameter estimates.

Table C.1 – Point Estimates for the Exponential Association III Regression Curve

Spring	a	B	c
Parthenia	4.6	1.60	0.025
Eliza	4.25	1.65	0.025
Old Mill	3.6	1.9	0.02

While these point estimates provide a concise summary of the data, they provide no information about their precision. That is, there is no indication of how "good" the estimates are. Thus, confidence intervals are needed.

The confidence intervals for each of the springs' regression curves were developed using a technique called *model comparison*. This technique compares two model fits, each of which results in a sum of squared error term. If the ratio of the two sum of squared errors is within the 95% confidence intervals of a specific F-distribution[2], then there is not enough data to indicate that the two fits are statistically different.

For regression curves with two or more parameters, confidence regions are used instead of confidence intervals. The results of the model comparison for the three spring regression curves are shown in Figures C.4 through C.6 below. These figures show that any values falling on the line shown on each graph (± 0.02 units) will result in model fits that cannot be shown to be statistically different from the one calculated by steepest descent. For the value of c, a range of 0.01 to 0.03 consistently provided model fits that consisted of equivalent confidence regions given in the figures below.

[2] This technique is based on the fact that the sum of the squared errors is distributed like a χ^2 distribution, and the ratio of two χ^2 distributions is distributed like an F-distribution. The degrees of freedom for a χ^2 distribution equal the number of squared errors while the degrees of freedom for an F-distribution are the degrees of freedom for each of the two χ^2 distributions.

Figure C.4 – Confidence Region for Parthenia Spring

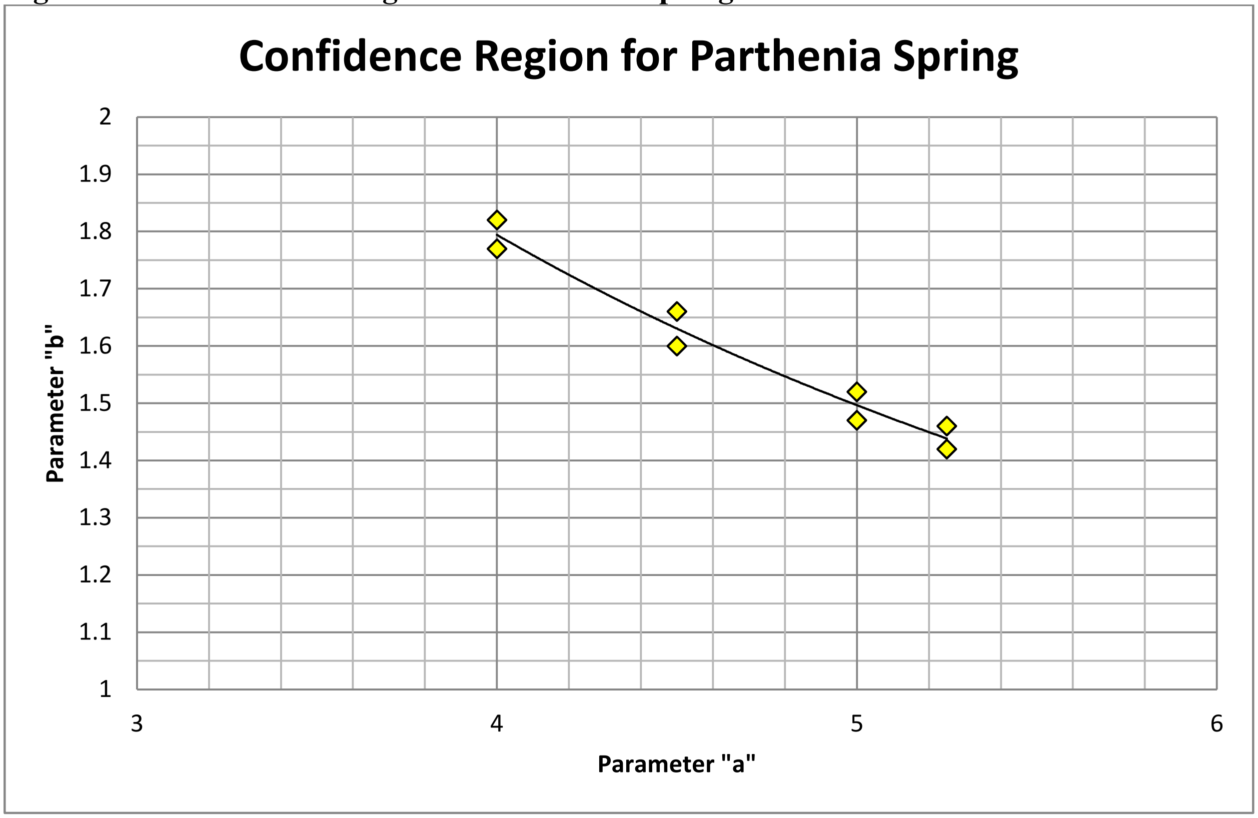


Figure C.5 – Confidence Region for Eliza Spring

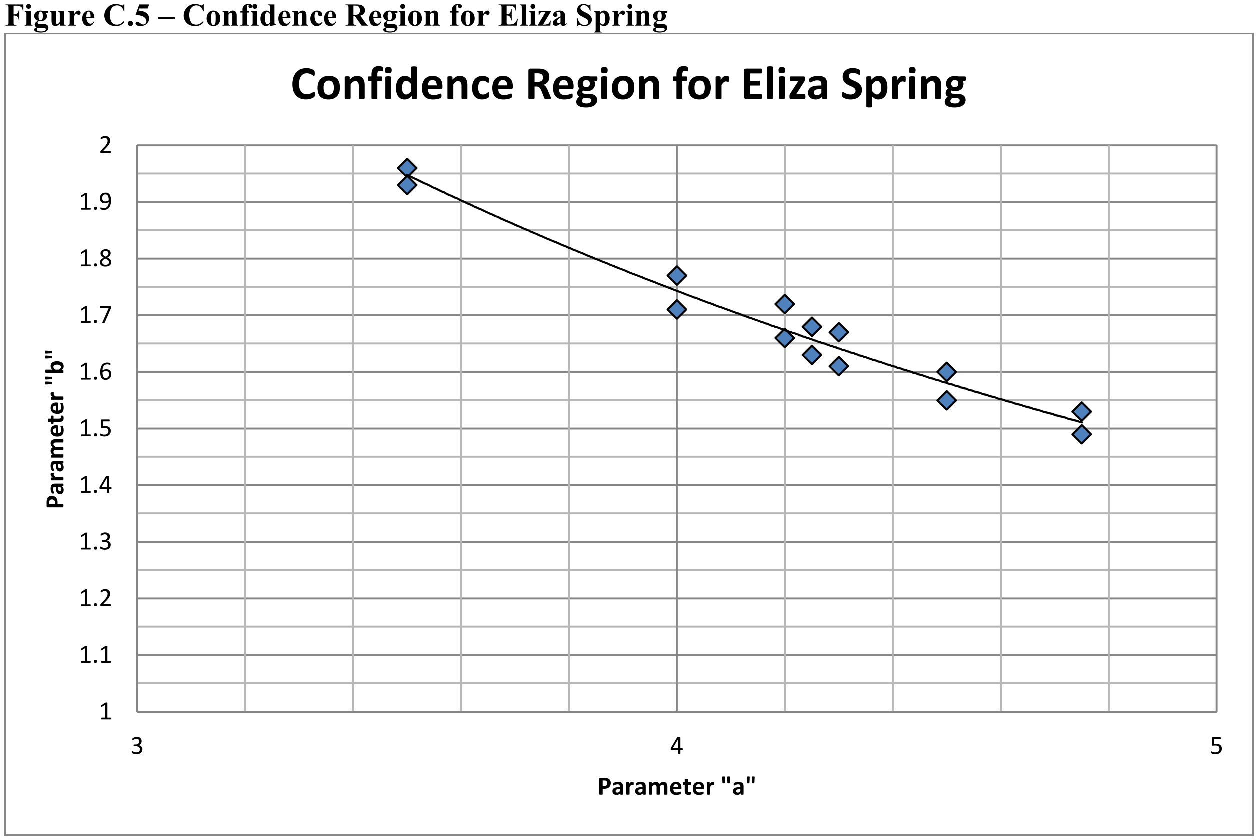
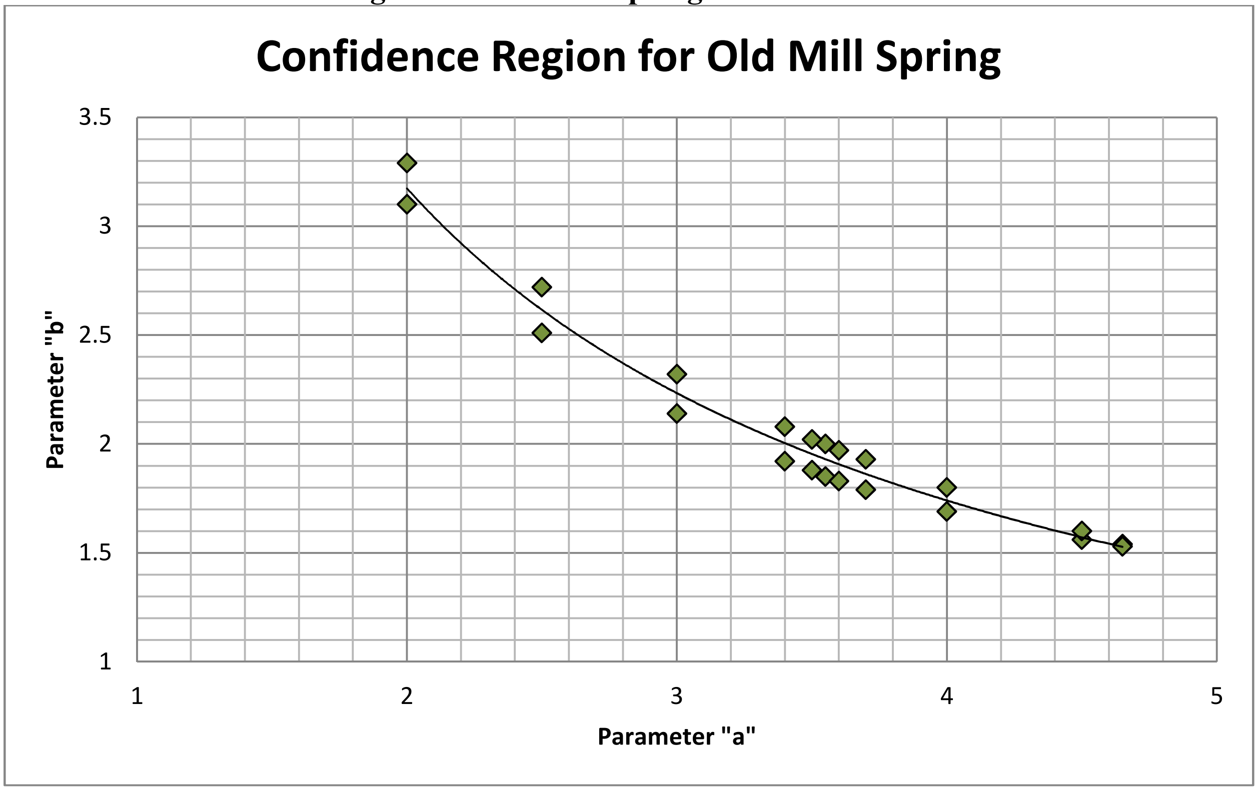


Table C.6 – Confidence Region for Old Mill Spring



Based on the data given and an Exponential Association III regression curve fit with parameters within a confidence region, six of the possible resulting curves are shown for each of the three springs in Figures C.7 through C.9 below. However, these curves are based on assumptions of maximum and minimum DO concentrations that have not been measured in the field. Whether these curves can be extrapolated to their assumed minimum and maximum is still an open question.

Figure C.7 – Exponential Association III curves for Parthenia Spring

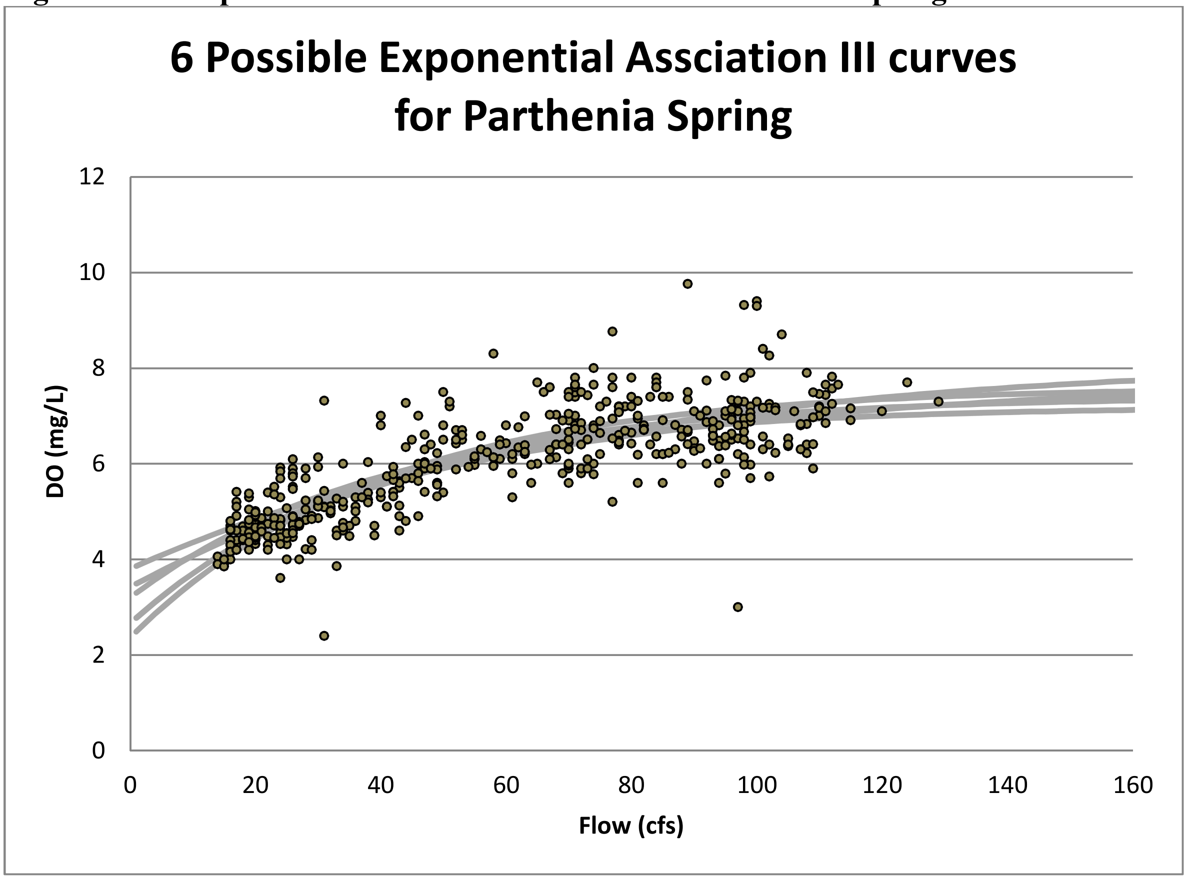


Figure C.8 – Exponential Association III curves for Eliza Spring

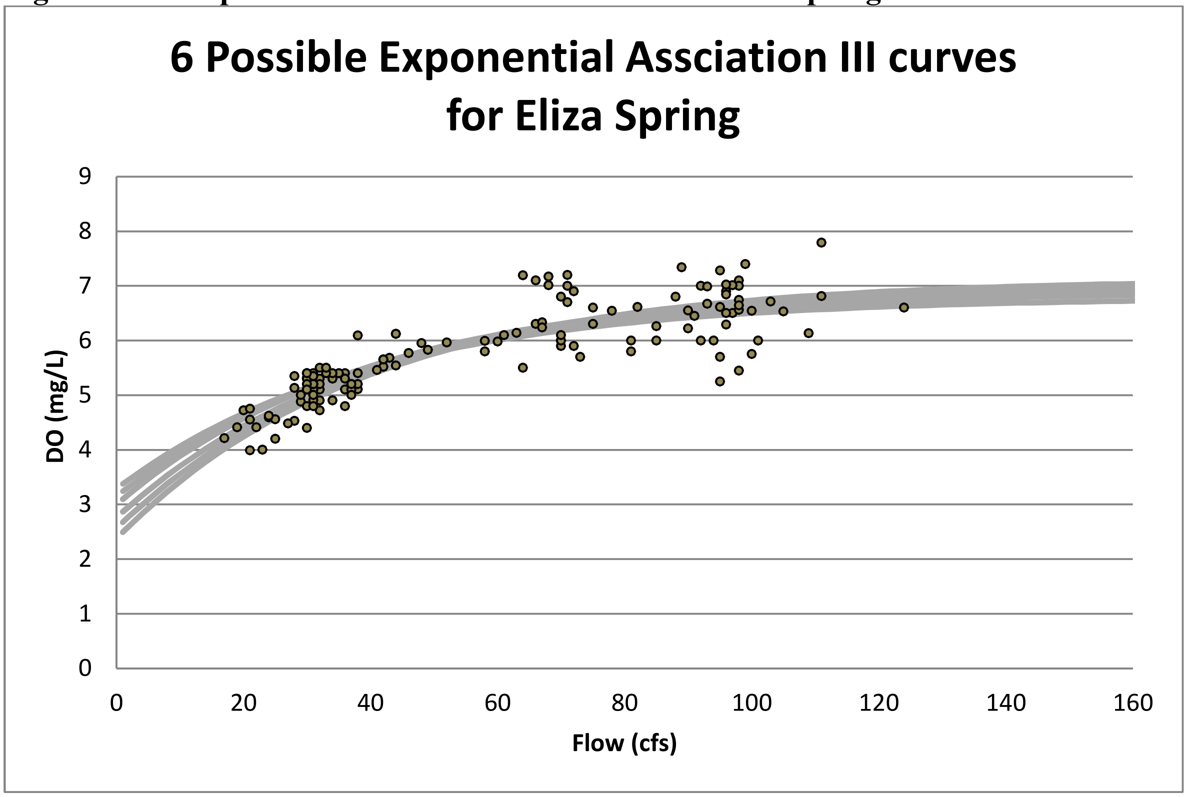


Figure C.9 – Exponential Association III curves for Old Mill Spring