PAIRED WATERSHED ANALYSIS WITH A SIMULATED CONTROL SITE

A Thesis
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Master of Science

by
Dillon Michael Cowan
August 2008
This thesis provides a rigorous statistical analysis of whether agricultural best management practices (BMPs) have reduced total dissolved phosphorus (TDP) loads from dairy farms in the Cannonsville Reservoir basin in New York State, a drinking water supply for New York City. This paper introduces the use of a simulation model as a control site in a paired watershed analysis as a technique that can be applied in basins when a control site is not available or feasible. In our analysis, the simulated control site also plays a critical role in allowing the analysis to separate the effects of BMPs from those of other water quality interventions. A variety of univariate and multivariate statistical models are considered along with different model structures. In addition to paired watershed models, several single watershed multivariate regression models are considered. Statistical power and the minimum detectable treatment effect (MDTE) are computed. Multivariate paired watershed models were clearly superior to the univariate paired watershed models recommended by EPA in terms of adjusted $R^2$ and ability to detect statistically significant BMP treatment effects. Modifications to the traditional log-linear regression model structure allowed the analysis to control for wastewater treatment plant (WWTP) upgrades and changes in the dairy cow population, using a physically meaningful statistical model. The results document a 41.5% reduction in TDP loads from non-point sources associated with the implementation of BMPs in the Cannonsville basin. Overall, this study shows that paired watershed analysis with a simulated control site is an innovative statistical method which should have great value in many applications.
Dillon Michael Cowan was born on June 30, 1981 in Wheatridge, Colorado to parents Mike and Brooke and is the second son in a family of four boys. He was raised primarily in the suburbs of Denver, Colorado, though his family also lived in Roseville, California from 1984-1990. He graduated from Heritage High School in Littleton, Colorado, in May of 1999 and went on to study civil engineering at Colorado State University in Fort Collins. Starting in the summer of 2000, he took a two-year leave from his studies to serve a full-time mission for the Church of Jesus Christ of Latter Day Saints in northern Italy. Two summers later, on June 26, 2004 he was married to Angella Dawn Davis in Denver, Colorado. The following summer, he graduated cum laude with a B.S. in civil engineering from Colorado State University and began graduate studies in the School of Civil and Environmental Engineering at Cornell University. He and his wife, Angella, have two children: Miles Dillon Cowan, born July 7, 2006, and Carter Jack Cowan, born May 14, 2008. Following his completion of the M.S. degree, Dillon will begin working as a water resources engineer at CH2M Hill in Sacramento, California.
For my family.
ACKNOWLEDGMENTS

I would like to thank my advisor, Christine Shoemaker, for providing me with an opportunity to be a part of the Environmental and Water Resources Systems Engineering program at Cornell. I would also like to thank my minor advisor, Jery Stedinger, for his invaluable help and advice in completing the statistical analyses in this paper. Water quality data for this study were provided by Pat Bishop at the New York Department of Environmental Conservation. My research was supported in part by USDA CREES grant 48881/A001 to Professors Tammo Steenhuis and Christine Shoemaker and by NSF grants DMS 0434390 and BES 0229176 to Professor Shoemaker. I am grateful for wonderful family and friends who have provided support and encouragement throughout my period of study at Cornell.
# TABLE OF CONTENTS

BIOGRAPHICAL SKETCH.................................................................................. iii
DEDICATION...................................................................................................... iv
ACKNOWLEDGEMENTS.................................................................................. v
LIST OF FIGURES............................................................................................ vii
LIST OF TABLES............................................................................................... viii

1 INTRODUCTION .......................................................................................... 1

2 DATA AND METHODS.................................................................................. 6
   2.1 Study Watershed...................................................................................... 6
   2.2 SWAT2000 Simulation Model................................................................. 9
   2.3 Monitoring Data..................................................................................... 10
   2.4 Framework for Statistical Analysis....................................................... 12

3 RESULTS AND DISCUSSION................................................................. 19
   3.1 Observed Flows and Total Dissolved Phosphorus Loads..................... 19
   3.2 Model Performance.............................................................................. 23
   3.3 Disaggregated Statistical Models.......................................................... 26
   3.4 BMP Treatment Effects......................................................................... 28
   3.5 Power Analysis and Minimum Detectable Treatment Effects.............. 30
   3.6 Single Watershed Models..................................................................... 33

4 CONCLUSIONS.......................................................................................... 38

APPENDIX........................................................................................................ 40
   A.1 Computation of BMP Treatment Effects for Log-linear Paired Models with Interaction Term ......................................................... 40
   A.2 Regression Outputs for Statistical Models ......................................... 43
   A.3 Additional Tables and Figures............................................................. 46

REFERENCES............................................................................................... 50
LIST OF FIGURES

Figure 2-1: Map of Cannonsville Reservoir Basin, Delaware County, NY.............7
Figure 3-1: Box plots of flow rate during pre- and post-BMP periods...............19
Figure 3-2: Box plots of TDP load during pre- and post-BMP periods..............20
Figure 3-3: Plot of dissolved phosphorus concentration vs. average flow rate.....21
Figure 3-4: Plot of event flow volume and average TDP load.......................22
Figure 3-5: Plot of monthly average TDP loads vs. time.............................23
Figure 3-6: Statistical power for the DP model using three different values for $\alpha$ and the full set of biweekly data from the pre- and post-BMP periods..............32
Figure 3-7: Plot of MDTE vs. $\beta$ for the DP model using three different values for $\alpha$ and the full set of biweekly data from the pre- and post-BMP periods..............34
Figure A-1: Plot of residuals vs. lagged residuals for LLP-F model.................49
LIST OF TABLES

Table 2-1: BMPs implemented on a study watershed between June 1995 and October 1996…………………………………………………………………………………….8

Table 2-2: Number of events during the pre and post-BMP periods…………………..12

Table 2-3: Summary of log-linear paired models……………………………………….16

Table 3-1: Monthly average TDP loads from wastewater plants…………………..21

Table 3-2: Regression coefficients and adjusted R² values for log-linear paired models………………………………………………………………………………...25

Table 3-3: Summary of disaggregated paired models ………………………………..27

Table 3-4: Regression coefficients and adjusted R² values for disaggregated paired models………………………………………………………………………………...28

Table 3-5: Summary of single watershed regression models…………………………36

Table 3-6: Regression coefficients and adjusted R² values for single watershed models………………………………………………………………………………...36

Table 3-7: Comparison of load reductions for DP, DS-I, and LLS models…………36

Table A-1: Regression outputs for LLP model……………………………………….43

Table A-2: Regression outputs for LLP-I model……………………………………….44

Table A-3: Regression outputs for LLP-F model……………………………………….44

Table A-4: Regression outputs for LLP-F-I model……………………………………….44

Table A-5: Regression outputs for DP model …………………………………………..44

Table A-6: Regression outputs for DP-I model………………………………………..45

Table A-7: Regression outputs for LLS model………………………………………..45

Table A-8: Regression outputs for DS model………………………………………..45

Table A-9: Regression outputs for DS-I model………………………………………..45
Table A-10: Summary of all statistical models……………………………….46

Table A-11: Sampling protocol for WBDR at Beerston……………………………47

Table A-12: Observed biweekly average flow rates (cms) during winter, spring, and summer …………………………………………………………………………………..47

Table A-13: Observed biweekly average flow rates (cms) during fall and full year…48

Table A-14: Observed biweekly TDP loads (kg/day) during winter, spring, and Summer…………………………………………………………………………………………..48

Table A-15: Observed biweekly TDP loads (kg/day) during fall and full year………48
CHAPTER 1
INTRODUCTION

Eutrophication is often caused by non-point source pollution from urban and agricultural activities (Gianessi et al., 1986; Havens and Steinman, 1995). It is a widespread environmental problem (USEPA, 1998; USGS, 1999) and the most common impairment of surface waters in the United States (USEPA, 1990). Eutrophication results in the growth of algae and aquatic weeds that interfere with water use, and presents a public health problem (Carpenter et al., 1998). To reduce the occurrence of eutrophication, federal, state, and local governments have invested millions of dollars in watershed management programs. In the Catskills region of Upstate New York, the City of New York has implemented an extensive program to reduce eutrophication in the Cannonsville Reservoir, the city’s third-largest drinking water supply impoundment (NYC DEP, 1997, 2001). A goal of the watershed management program is to avoid the construction of a filtration plant for New York City, which would cost billions of dollars. Although New York City’s watershed management program addresses pollution from many sources, one thrust is on the implementation of agricultural best management practices (BMPs) on dairy farms, which have been identified as key contributors of phosphorus to the reservoir (Brown et al., 1989; Scott et al., 1998; NYC DEP, 2001; Delaware County Department of Watershed Affairs, 2002; Tolson and Shoemaker, 2004).

In the Cannonsville basin, funding for BMPs is administered via the Watershed Agricultural Program (WAP), an incentive-based program launched in 1993 by the New York City Department of Environmental Protection (NYC DEP) and the local Watershed Agricultural Council (WAC). BMPs for each individual farm participating in the WAP are selected based on a Whole Farm Plan (WFP) process, which takes into
consideration a variety of economic and environmental factors (Porter et al., 1997; Watershed Agricultural Council, 2003). Today, the participation level in the WAP by dairy farmers has surpassed 85%, and a wide variety of BMPs have been implemented across the basin. This study provides a rigorous statistical analysis that estimates the extent to which BMPs have actually reduced dissolved phosphorus loads in the West Branch Delaware River, which drains into Cannonsville Reservoir. In this analysis we use monitoring data and a watershed simulation model. The focus of the analysis is dissolved phosphorus because it is the most important nutrient in regulating algae growth in reservoir for most of the year (Doerr et al., 1998, Owens et al., 1998a).

The effectiveness of agricultural BMPs has been the topic of many research studies (Sharpley et al., 2006). Most research studies can be classified as either (1) modeling studies or (2) monitoring studies, though both approaches should be used together when possible (Easton et al., 2008c). Modeling studies vary in geographic scale from farms and small watersheds (Gitau et al., 2004, 2005) to large catchments such as the entire Cannonsville (Tolson, 2005; Tolson and Shoemaker, 2004, 2007) and Chesapeake Bay (Boesch et al., 2001) watersheds. The general approach is to use an environmental simulation model, such as the Soil and Water Assessment Tool (SWAT) (Tolson, 2005; Arabi et al., 2006, 2007; Bramcort et al., 2006), Annualized Agricultural Non-point Source Pollution model (AnnAGNPS) (Srivastava et al., 2002), or Generalized Watershed Loading Function (GWLF) (Rao et al., 2008), to generate forecasts of water quality constituents, such as sediment, phosphorus or nitrogen, under a suite of different management scenarios. The scenarios can be compared in terms of performance metrics such as load reduction and cost, and the best scenario(s) or suite(s) of BMPs identified. Precise estimates of BMP effectiveness may be difficult to obtain with simulation models due to uncertainty in climate inputs and calibration parameters, and the uncertainty in simple export
coefficients for representing BMPs. Simulated load reductions also do not address whether or not BMPs installed in the past have actually had a significant effect on water quality. Nevertheless, modeling studies are arguably one of the most useful tools that researchers and policy makers have for screening BMP scenarios at the watershed scale, and have been influential in shaping many watershed management programs.

Most monitoring studies consider a very small geographic area, such as a farm, field, or test plot, as a laboratory in order to carry out a physical test on the effectiveness of one particular practice. These small-scale analyses tend to focus on one specific BMP, such as fencing (James et al., 2007), buffer strips (Murray, 2001), precision feeding (Cerolasetti et al., 2004; Ghebremichal et al., 2007), or alternative cropping practices (Kleinman et al., 2007). Plot and farm-scale studies have been very effective at documenting the effects of BMPs, and the results from small-scale studies can be extrapolated to the watershed scale, though the extrapolated load reductions tend to be optimistic (Striffler, 1965; Wauchope, 1978).

In larger watersheds, monitoring studies tend to rely on data sampled from streams and rivers to perform trend analysis and other exploratory data analyses (Longabucco et al., 1998; Meals et al., 2001; Udawatta et al., 2002; Inamdar et al., 2002). Methods for the detection of changes or trends in water quality data are a topic of research (Esterby et al., 1996). There are several documented cases where widespread BMP implementation failed to yield noticeable improvements (Sharpley and Smith, 1994; Millard et al., 1997; Boesch et al., 2001), which Sharpley et al. (2006) blame on poor implementation and targeting. While implementation and targeting are important, monitoring studies must also take into consideration the fact that many years may pass before the impacts of BMPs are detectable at the basin scale (Boesch et al., 2001, Wang et al., 2002), even when BMPs are working as they should. In particular, a concern is that BMP effects may be masked by other factors such as
inter-annual variation (Longabucco and Rafferty, 1998) and long-term trends, which occur even in undisturbed watersheds (Esterby, 1996; Murtaugh, 2000; Moog and Whiting, 2002). In the Cannonsville Reservoir watershed, upgrades to wastewater treatment plants (WWTPs) and changes in the size of the dairy cow population should be considered.

At the scale of small catchments, paired watershed analysis has proven effective at controlling for hydrologic variability and overcoming many of the challenges associated with single-time series trend analyses over a shorter period of time (Spooner et al., 1985). Paired watershed analysis is a widely used statistical method that dates back over fifty years (Wilm, 1949; Kovner and Evans, 1954). As the name of the technique suggests, a paired watershed study requires two watersheds: a control site and a test site. Both sites are monitored before and after the implementation of BMPs on the test site, and a statistical analysis is used to detect changes in the relationship between loads at the control and test watersheds.

EPA has released several sets of guidelines recommending the use of a univariate log-linear regression model, which uses paired loads from the control site as the covariate (USEPA, 1993, 1997a, 1997b). Detection of BMP effects with a paired analysis remains an active area of research (Loftis et al., 2001; Bishop et al., 2005), though most recent studies focus more on the application of the EPA-recommended techniques (Clausen et al., 1996; Galeone et al., 1999; Brannan et al., 2000; Murtaugh, 2000; Udawatta et al., 2002; Ricker et al., 2008). Loftis et al. (2001) and Hewlett and Peinarr (1973) discuss the use of multiple covariates in paired watershed studies. Bishop et al. (2005) actually demonstrate the effectiveness of multivariate paired models, which significantly outperformed the EPA-recommended univariate models and were more likely to detect BMP effects.
In terms of geographic scale, paired watershed studies typically use small catchments on the order of 500 ha or less (Shirmohammadi et al., 1997; Galeone, 1999; Brannan et al., 2000; Bishop et al., 2005), with some study areas smaller than 2 ha (Clausen et al., 1996). There are also examples of paired studies using catchments larger than 1000 ha (Meals et al., 2001; Ricker et al., 2008). Locating an adequate control watershed for the Cannonsville basin, which has an area of nearly 120,000 ha, would be extremely difficult, if not impossible. In fact, a true paired study at the scale of the Cannonsville basin would require a combined land area equal in size to the entire State of Rhode Island.

This paper introduces the use of a simulation model as a control site in a paired watershed analysis as a technique that can be applied in basins where a control site is not available or feasible. In our analysis, the simulated control site also plays a critical role in separating the effects of BMPs from those of WWTP upgrades and changes in the dairy cow population. A variety of univariate and multivariate regression models are considered along with different model structures. In addition to paired models, several single watershed multivariate regression models are also tested. The results of the statistical models are used to compute estimates of dissolved phosphorus load reductions associated with the implementation of agricultural BMPs in the Cannonsville watershed.
CHAPTER 2
DATA AND METHODS

2.1 Study Watershed

Figure 2-1 shows the location of the Cannonsville Reservoir in the Catskills Region of Upstate New York. It is New York City’s third-largest drinking water reservoir. Eighty percent of the 1200 km$^2$ Cannonsville Basin is drained by the West Branch Delaware River (WBDR), which is 80 km long and flows to the southwest into the reservoir. Land use is dominated by forests (59%) and agriculture (36%), with dairy farms accounting for the vast majority of the agricultural land. There are four major villages located in the basin – Walton, Delhi, Hobart, and Stamford. The human population in the sparsely populated area is estimated at approximately 18,000 and less than 0.5% of the total land area is urban. The climate of the Cannonsville Basin is humid, with an average annual temperature of about 8ºC, and an average annual precipitation of about 1100 mm, one third of which falls as winter snow. Elevations range from 285 to 995 meters above mean sea level, and slopes range from zero to over 40%, with an average land-surface slope of 19%. Runoff production is driven primarily by saturation-excess overland flow (Walter et al., 2000).

The Cannonsville Reservoir has experienced eutrophication problems for decades (Schumacher and Wagner, 1973; USEPA, 1974; Wood, 1979; Brown et al., 1986; Bader, 1993; Effler and Bader, 1998). Studies have identified excessive phosphorus loads to the reservoir from point and non-point sources as the cause of eutrophication (Wood, 1979; Brown et al., 1980, 1983, 1986; Effler and Bader, 1998). Dissolved phosphorus, in particular, plays the most important role in regulating the growth of algae (Doerr et al., 1998, Owens et al., 1998a). The majority of point source loads come from wastewater treatment plant (WWTP) discharges, while the local dairy farms are responsible for most of the non-point source loads (Longabucco...
et al., 1998; Tolson and Shoemaker, 2004). Due to the reservoir’s important role as a New York City drinking water supply source, the Cannonsville basin has been the focus of many scientific studies relating to sediment and phosphorus transport, lake water quality, and BMPs (Auer et al., 1998; Auer and Forrer, 1998; Longabucco et al., 1998; Owens et al., 1998a, 1998b; Walter et al., 2001; Giasson et al., 2002; Miller et al., 2002; Schneiderman et al., 2002; Cerolasetti et al., 2004; Gitau et al., 2004, 2005; Benaman and Shoemaker, 2004, 2005; Tolson and Shoemaker, 2004, 2007; Tolson, 2005; Benaman et al., 2005a, 2005b; Bishop et al., 2005; Kleinman et al., 2005; Ghebremichael et al., 2007; James et al., 2007; Archibald et al., 2008; Easton et al., 2008a, 2008b, 2008c; Rao et al., 2008).

Figure 2-1: Map of Cannonsville Reservoir Basin, Delaware County, NY
New York City has invested millions of dollars in a watershed management program designed to reduce phosphorus pollution from point and non-point sources. Since the early 1990s all of the major WWTPs in the basin have been upgraded to release lower concentrations of phosphorus into the West Branch Delaware River, while dairy farms have been the focus of efforts to reduce pollution from non-point sources. Dairy farms in the basin began installing BMPs in 1993, and today the vast majority of the farms have or are currently implementing one or more BMPs. Data describing which BMPs were implemented at which locations in the basin are not available to the general public due to confidentiality concerns. However, Table 2-1 provides a list of BMPs that were tested on a dairy farm in the basin and used for a smaller scale statistical analysis described in Bishop et al. 2005. In addition to WWTP upgrades and the implementation of BMPs, a 43% decline in the cattle population between 1987 and 2002 likely contributed to lowering phosphorus loads during the period of this study.

Table 2-1: BMPs implemented on a study watershed between June 1995 and October 1996 (Bishop et al. (2005))

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near Barn</td>
<td>Installation of manure storage lagoon</td>
</tr>
<tr>
<td></td>
<td>Barnyard improvements including water management</td>
</tr>
<tr>
<td></td>
<td>Filter area established for barnyard runoff</td>
</tr>
<tr>
<td></td>
<td>Stream corridor relocated away from barnyard</td>
</tr>
<tr>
<td></td>
<td>Grazing cows excluded from stream and swale areas</td>
</tr>
<tr>
<td></td>
<td>Milkhouse wastewater diverted to manure storage</td>
</tr>
<tr>
<td></td>
<td>Relocation of silage storage bag away from stream</td>
</tr>
<tr>
<td></td>
<td>Improvement of stream crossings and roadways</td>
</tr>
<tr>
<td>Farm Scale</td>
<td>Access roads constructed to allow manure spreading on upper slopes</td>
</tr>
<tr>
<td></td>
<td>Distributed manure according to nutrient management plan</td>
</tr>
<tr>
<td></td>
<td>Fencing improvements to support rotational grazing</td>
</tr>
<tr>
<td></td>
<td>Spring development to supply drinking water away from stream</td>
</tr>
</tbody>
</table>
Table 2-1 continued

| Diversion ditches to improve field drainage |
| Subsurface drainage to reduce field saturation and runoff production |
| Contour stripping to reduce erosion |
| Crop rotation to reduce erosion |

2.2 SWAT2000 Simulation Model

The simulated control site for this study was created using a SWAT2000 model of the Cannonsville Basin. SWAT2000 (Neitsch et al., 2001a, 2001b) is a widely used hydrologic and water quality model that is specifically designed to simulate flow and nutrient transport in agricultural watersheds, and has been used in several studies of the Cannonsville Basin (Tolson and Shoemaker 2004, 2007; Beneman et al., 2005a, 2005b; Benaman and Shoemaker, 2004, 2005; Easton et al. 2007) and many other watersheds (Gassman et al., 2007). The procedures used to develop, calibrate, and validate the SWAT2000 model of the basin are described in Tolson and Shoemaker (2004, 2007).

Using SWAT to create a simulated control site in the statistical analysis offers many advantages. In a paired study, the ideal scenario is to have two watersheds that are in close proximity so that the climate is essentially the same in both basins. The SWAT model is driven by climate data that was collected in the Cannonsville basin, so both the test and control sites have the same daily weather inputs, to the accuracy level of the weather monitoring stations. Also, the SWAT model captures the land use distribution of the Cannonsville basin so that in this case, the test and control sites have the same landscapes, which is another ideal characteristic. Essentially, the simulated flow rates and TDP loads produced by the SWAT model look like flow and water quality data collected at a watershed with the same land use and climate as the Cannonsville, given the resolution of the environmental data.
The final advantage to using a SWAT model as the control site is the ability to control for factors other than BMPs that could affect water quality. The purpose of this study is to look at the effects of BMPs, and it is, therefore, necessary to control for other factors, such as upgrades to wastewater treatment plants and changes in the cattle population. SWAT is capable of incorporating point source loads, and cattle populations are reflected in the manure spreading rates in the management input files. Thus, using a SWAT model as the control site may allow us to specifically document BMP treatments effects, after accounting for other factors. Because the SWAT model represented a control site without BMPs, BMPs were not included in the simulation model. Thus, the ability of SWAT to correctly represent BMP inputs is not an issue.

To summarize, the SWAT simulated control site is designed to predict flow, sediment, and phosphorus under the scenario where no BMPs were implemented, but treatment plants were upgraded and the cattle population declined. This approach is particularly attractive because the relationship between the test and control sites should be stronger than in most paired studies, and because treatment effects due to factors besides BMPs can be controlled for in the simulation model.

2.3 Monitoring Data

Flow measurements are taken at the USGS station (Site ID 01423000) on the West Branch Delaware River at Walton (see Figure 2-1). The USGS reports daily average flows which are based on measurements taken every 15 minutes. Water quality measurements are taken at a sampling station maintained by the New York Department of Environmental Conservation (NY DEC). The station is located on the West Branch Delaware River (WBDR) at Beerston, which is about 8 km downstream from the USGS flow gage at Walton. The frequency at which stream water samples are taken depends on the stage of the river at the USGS gage upstream. As the stage
increases, so does the frequency of water quality sampling. The frequency varies from once per week (stage < 0.9 m) to once every hour (stage > 4.0 m) (Longabucco et al., 1998). Stream water samples are analyzed for TDP and loads are computed as the product of flow volume and concentration. The flow volume at Beerston is estimated at 1.06 times the flow volume measured at the USGS gage at Walton because of the 6% increase in drainage area (Longabucco et al., 1998). The majority of point source loads come from WWTPs at Stamford, Hobart, Delhi, and Walton (see Figure 2-1). The two largest plants are Walton and Delhi, followed by Stamford and Hobart. Data describing discharges from the four major WWTPs are also provided by NY DEC. Seven samples (on consecutive days) are taken per month at each site and used to generate an average daily load for that month. The motivation for sampling seven consecutive days is to capture an entire week, which will include diurnal and weekday/weekend effects. A more detailed discussion of WWTP sampling protocols is provided by Pacenka et al. (2002).

For the statistical analysis, events were defined using a biweekly averaging window. When water quality and hydrologic monitoring data are more closely spaced in time, serial correlation in the data will complicate the statistical analyses (Hirsch and Slack, 1984; Manly, 2001). Biweekly averages are more reliable than daily or instantaneous measurements, and will compensate for delays between when loads are measured in reality and predicted by simulation models. Larger time windows, such as months, would reduce the number of data points and would also introduce complicated load estimation issues (Cohn et al., 1992; Cohn, 1995). The pre- and post-BMP periods were defined as October 1991 – December 1994, and January 2000 – September 2004, respectively. As reported in Table 2-2, there were 84 events during the pre-BMP period and 123 events during the post-BMP period.
### Table 2-2: Number of events during the pre- and post-BMP periods

<table>
<thead>
<tr>
<th>Season</th>
<th>Pre-BMP</th>
<th>Post-BMP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter (December 18 - April 13)</td>
<td>28</td>
<td>42</td>
<td>70</td>
</tr>
<tr>
<td>Spring (April 14 - June 15)</td>
<td>12</td>
<td>21</td>
<td>33</td>
</tr>
<tr>
<td>Summer (June 16 - September 30)</td>
<td>24</td>
<td>40</td>
<td>64</td>
</tr>
<tr>
<td>Fall (October 1 - 15 December)</td>
<td>20</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>Full Year</td>
<td>84</td>
<td>123</td>
<td>207</td>
</tr>
</tbody>
</table>

#### 2.4 Framework for Statistical Analysis

The EPA has published several sets of recommendations on how to use paired watershed analysis to detect BMP effects (USEPA, 1993, 1997a, 1997b). All of the approaches recommended by EPA involve using a univariate log-linear regression model, where phosphorus, for example, at the control site is the explanatory variable for phosphorus at the test site. A binary $[0,1]$ indicator variable is used to model BMP effects. Though the univariate log-linear approach is widely used, Loftis et al. (2001) suggest that the addition of other covariates can improve model performance and the ability to detect BMP effects. Bishop et al. (2005) report that including multiple covariates increased the adjusted $R^2$ value from 0.65 for the EPA-recommended model to 0.89 for the multivariate model. Bishop et al. (2005) also report an increase in statistical power from 60 to 95%.

This study initially considers the multivariate log-linear paired model proposed by Bishop et al. (2005):

$$\ln(P_l^m) = a + b \ln(P_\delta^i) + c \ln(Q_l^m/Q_l^\delta) + d \ln(Q_{l,peak}^m) + e \ln(Q_{l,avg}^m) + f k_i + g k_i \ln(P_\delta^i - p) + \epsilon_i$$

[2-1]

or (written without the natural log transformation)
\[ p_i^m = (P_i^s)^b (Q_i^m / Q_i^s)^c (Q_i^m \text{peak})^d (Q_i^m \text{avg})^e \exp\{a + f k_i + g k_i [\ln(P_i^s) - p] + \epsilon_i\} \]  

where \( \ln \) is the natural logarithm, \( i \) is the biweekly period index, \( P_i^m \) is the phosphorus load measured at Beerston during period \( i \), \( P_i^s \) is the simulated phosphorus load produced by SWAT during period \( i \), \( Q_i^m \) is the flow volume measured at Walton for period \( i \), \( Q_i^s \) is the simulated flow volume for period \( i \) produced by SWAT, \( Q_i^m \text{peak} \) and \( Q_i^m \text{avg} \) are the peak and average daily flow rates measured at Walton during period \( i \), \( k_i \) is the \([0,1]\) BMP indicator variable, \( p \) is the average of \( \ln(P_i^s) \) during the post-BMP period, and \( \epsilon_i \) is the residual error.

The natural log transform (from Eq. [2-2] to Eq. [2-1]) was employed, as is common practice, to normalize the flows and phosphorus loads and to avoid the problem of heteroscedasticity or nonconstant variance in the residuals of the regression models (Cohn et al., 1989; USEPA, 1997b). This study will consider which of the covariates from the full model improve model performance, and the best model will be used to estimate the impact of BMPs on dissolved phosphorus loads to the Cannonsville Reservoir.

### 2.4.1 Interpretation of Model Terms

In a traditional paired study, the intercept, \( a \), accounts for size differences between the two watersheds as well as systematic differences in the magnitude of TDP loads. For this study the control site is a numerical model of the test site, so both sites have the same size. However, many simulation models have systematic biases, and the SWAT2000 model has a tendency to under-predict larger phosphorus loads. Thus, we expected the intercept \( a \) to be greater than zero.

The phosphorus loads from the simulated control site, \( \ln(P_i^s) \), should account for hydrologic and environmental variability. Loads from the control site also account
for factors besides BMPs, such as changes in the basin-wide cattle population and improvements to wastewater treatment plants.

When the control site is a physical watershed, the flow ratio term, \( \ln(Q_i^m/Q_i^s) \), can account for imbalances between the precipitation at the two sites. For this study, the flow ratio term may account for flow forecasting errors and the resulting error in SWAT simulated loads.

Hydrology plays a significant role in determining phosphorus loads. Thus, including measured peak and average flow, \( \ln(Q_i^{m\ peak}) \) and \( \ln(Q_i^{m\ avg}) \), respectively, as covariates could significantly improve model performance. These terms are very important for this study because SWAT is a daily time step model, and often does not capture the effects of short, intense precipitation events. SWAT also tends to underestimate large phosphorus loads that occur over longer durations.

Here \( k \) is a binary indicator variable that is set to zero during the pre-BMP period and one during the post-BMP period. The coefficient \( f \) describes the log change in phosphorus loading between the pre- and post-BMP periods. Lastly, the BMP-phosphorus interaction term \( g k_i[\ln(P_i^s) - p] \), where \( p \) is the sample mean of post-BMP simulated phosphorus loads, can describe a change in the regression slope in the post-BMP period. One would expect to see that BMPs perform better as magnitude increases, suggesting a negative value for \( g \).

### 2.4.2 Screening of Covariates

In order to avoid multicollinearity, which causes numerical instability when solving for the regression coefficients, the covariates were examined to determine if there were strong linear relationships between any of the explanatory variables. At the level of biweekly averages, peak and average flow, \( \ln(Q_i^{m\ peak}) \) and \( \ln(Q_i^{m\ avg}) \), are very strongly correlated \( (R^2 > 0.95) \). Thus, average flow was removed from the
model. Peak flow was chosen over average flow because models with only peak flow performed slightly better than those with only average flow. With average flow removed, there were no strong linear relationships between the remaining covariates. The flow ratio term, \( \ln(\frac{Q_i^m}{Q_i^s}) \), was usually close to one because SWAT is quite effective at predicting flow at the biweekly level. The coefficient on the flow ratio was never statistically significant. Removing the flow ratio and average flow terms from Eq. [2-1] results in the following simplified statistical model:

\[
\ln(P_i^m) = a + b \ln(P_i^s) + c \ln(Q_i^{m,\text{peak}}) + f k_i + g k_i[\ln(P_i^s) - p] + \varepsilon_i
\]

or (written without the natural log transformation)

\[
P_i^m = (P_i^s)^b(Q_i^{m,\text{peak}})^c \exp\{a + f k_i + g k_i[\ln(P_i^s) - p] + \varepsilon_i\}
\]

### 2.4.3 Variations of Statistical Models

Four different log-linear paired watershed models were explored. The first model is the univariate model recommended by EPA, which uses the phosphorus load at the control site and the BMP indicator as covariates.

\[
\ln(P_i^m) = a + b \ln(P_i^s) + f k_i + \varepsilon_i
\]

or (written without the natural log transformation)

\[
P_i^m = (P_i^s)^b \exp\{a + f k_i + \varepsilon_i\}
\]

The second model is the EPA model with a BMP-phosphorus interaction term:

\[
\ln(P_i^m) = a + b \ln(P_i^s) + f k_i + g k_i[\ln(P_i^s) - p] + \varepsilon_i
\]
or (written without the natural log transformation)

\[ p_i^m = (P_i^s)^b \exp\{a + f \cdot k_i + g \cdot k_i \ln(P_i^s) - p + \varepsilon_i\} \]  

[2-8]

The third model is the EPA paired watershed model with peak flow included:

\[ \ln(p_i^m) = a + b \ln(p_i^s) + c \ln(Q_{i\text{peak}}^m) + f \cdot k_i + \varepsilon_i \]  

[2-9]

or (written without the natural log transformation)

\[ p_i^m = (P_i^s)^b (Q_{i\text{peak}}^m)^c \exp\{a + f \cdot k_i + \varepsilon_i\} \]  

[2-10]

The final model includes all of the covariates in the simplified multivariate paired model (Eq. [2-3]). For the purposes of abbreviation throughout the paper, the four log-linear paired models will be referred to as log-linear paired (LLP) in Eq. [2-5], log-linear paired with BMP-phosphorus interaction (LLP-I) in Eq. [2-7], log-linear paired with peak flow (LLP-F) in Eq. [2-9], and log-linear paired with peak flow and BMP-phosphorus interaction (LLP-F-I) in Eq. [2-3]. See Table 2-3.

Table 2-3: Summary of log-linear paired statistical models

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Eq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLP</td>
<td>( \ln(p_i^m) = a + b \ln(p_i^s) + f \cdot k_i + \varepsilon_i )</td>
<td>[2-5]</td>
</tr>
<tr>
<td>LLP-I</td>
<td>( \ln(p_i^m) = a + b \ln(p_i^s) + f \cdot k_i + g \cdot k_i \ln(p_i^s) - p + \varepsilon_i )</td>
<td>[2-7]</td>
</tr>
<tr>
<td>LLP-F</td>
<td>( \ln(p_i^m) = a + b \ln(p_i^s) + c \ln(Q_{i\text{peak}}^m) + f \cdot k_i + \varepsilon_i )</td>
<td>[2-9]</td>
</tr>
<tr>
<td>LLP-F-I</td>
<td>( \ln(p_i^m) = a + b \ln(p_i^s) + c \ln(Q_{i\text{peak}}^m) + f \cdot k_i + g \cdot k_i \ln(p_i^s) - p + \varepsilon_i )</td>
<td>[2-3]</td>
</tr>
</tbody>
</table>

2.4.4 Computation of BMP Treatment Effects

The results of the regression analyses were used to compute BMP treatment effects when the coefficients relating to BMPs were statistically significant at the 5%
level. The approach used to compute BMP effects varied, depending on whether or not the BMP-phosphorus interaction term was included in the model.

Consider the LLP and LLP-F models in Table 2-3, which do not include the BMP-phosphorus interaction term. For these two models, a constant multiplicative load adjustment factor of \(\exp(f)\) is applied to all loads in the post-BMP period (when \(k_i = 1\)). The percent reduction (when \(f < 0\)) in loads associated with the implementation of BMPs is given by:

\[
\%\text{Reduction} = 100(1 - \exp(f))
\]  

[2-11]

The computation of BMP treatment effects becomes more complicated for the LLP-I and LLP-F-I models because the multiplicative load adjustment factor \(\exp\{f k_i + g k_i [\ln(P_{si}) - p]\}\) varies with each biweekly period. For these models, an average load adjustment factor was computed by comparing loads predicted in a post-BMP period, with loads that one would have expected to see in a pre-BMP period if the BMPs had never been installed.

Consider the full LLP-F-I model, which includes peak flow and the BMP-phosphorus interaction term. Loads for a post-BMP period can be estimated using Eq. [2-4]. In order to compute estimates of loads for the same period but without BMPs, one must use the same equation, but with \(f\) and \(g\) fixed at zero. The average load reduction factor can then be computed as the ratio of expected loads:

\[
\frac{E[P^m]}{E[P^m|f = g = 0]} = \frac{E[(P^s)^b (Q^{mpeak})^c \exp\{a + f + g[\ln(P^s) - p] + \varepsilon\}]}{E[(P^s)^b (Q^{mpeak})^c \exp\{a + \varepsilon\}]}
\]  

[2-12]
The average percent load reduction factor is:

\[
\% \text{Reduction} = 1 - \frac{E[P^m]}{E[P^m | f = g = 0]} \tag{2-13}
\]

Expectations can be taken analytically, assuming \(P^s\) and \(Q^{m,\text{peak}}\) have lognormal distributions, which results in the corresponding percent reduction factors for the LLP-I and LLP-F-I models in Equations [2-14] and [2-15], respectively.

\[
\% \text{Reduction} = 100 \left[ 1 - \frac{\exp\{f + g(\mu_P - p) + 2(b + g)c\sigma_{PQ} + 1/2 ((b + g)^2 \sigma_P^2)\}}{\exp\{2bc\sigma_{PQ} + 1/2 (b^2 \sigma_P^2)\}} \right] \tag{2-14}
\]

\[
\% \text{Reduction} = 100 \left[ 1 - \frac{\exp\{f + g(\mu_P - p) + 1/2 ((b + g)^2 \sigma_P^2)\}}{\exp\{1/2 (b^2 \sigma_P^2)\}} \right] \tag{2-15}
\]

where the subscripts \(P\) and \(Q\) represent \(\ln(P^s)\) and \(\ln(Q^{m,\text{peak}})\), respectively. The mean, \(\mu_P\), variance, \(\sigma_P^2\), and covariance, \(\sigma_{PQ}\), can be estimated using sample statistics. Sample statistics should be computed using the entire period of record in order to provide more precise estimates of \(\mu_P\), \(\sigma_P^2\), and \(\sigma_{PQ}\). A full derivation of the expectations is provided in section A.1 of the Appendix. An alternative approach would be to replace the expectation operators in Eq. [2-13] with sample averages. The use of sample averages would be attractive when the assumption of log-normality is violated. Expectations and sample averages should be taken using measured flows and simulated loads from the entire study period in order to generate a more precise estimate of the true load reduction factor.
CHAPTER 3
RESULTS AND DISCUSSION

3.1 Observed Flows and Total Dissolved Phosphorus Loads

Figures 3-1 and 3-2 contain box plots of observed flow rates and TDP loads. With the exception of winter, average biweekly average flows were higher during all seasons and the full year in the post-BMP period. This might have resulted in higher average TDP loads in the post-BMP period, but instead, average TDP loads decreased across all seasons and the full year. Summary statistics are provided in Tables A-12, A-13, A-14, and A-15, in Section A.3 of the Appendix.

Figure 3-1: Box plots of flow rate during pre- and post-BMP period across all seasons (W: Winter, Sp: Spring, Su: Summer, F: Fall, FY: Full Year)
The most reasonable explanation for the overall reduction in TDP loads despite the increase in flow rates is that the combination of WWTP upgrades, cattle population declines, and BMPs resulted in a decrease in TDP concentrations. The log-log plot of TDP concentration vs. average flow, located in Figure 3-3, shows that, in fact, concentrations were generally lower across all flow rates.

During lower flow periods, the reduction in concentration in the post-BMP period was likely driven by upgrades to the WWTPs in the basin because very little phosphorus loading from non-point sources occurs during low-flow conditions (Pionke et al., 1997). Table 3-1 contains summary statistics for the combined TDP load from the four largest plants in the basin during the pre- and post-BMP periods, as well as the duration of the entire study. The table reveals that TDP loads from

---

**Figure 3-2: Box plots of TDP load during pre- and post-BMP period across all seasons (W: Winter, Sp: Spring, Su: Summer, F: Fall, FY: Full Year)**
WWTPs in the basin were dramatically lower during the post-BMP period, which would explain the decrease in TDP concentration during low-flow conditions.

![Figure 3-3: Plot of dissolved phosphorus concentration vs. average flow rate](image)

**Table 3-1: Monthly average TDP loads (kg/day) from wastewater plants**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Min</strong></td>
<td>0.09</td>
<td>15.32</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Q1</strong></td>
<td>6.43</td>
<td>16.37</td>
<td>0.23</td>
</tr>
<tr>
<td><strong>Median</strong></td>
<td>7.76</td>
<td>20.40</td>
<td>6.35</td>
</tr>
<tr>
<td><strong>Q3</strong></td>
<td>15.43</td>
<td>27.74</td>
<td>7.96</td>
</tr>
<tr>
<td><strong>Max</strong></td>
<td>45.92</td>
<td>45.92</td>
<td>28.46</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>10.92</td>
<td>23.07</td>
<td>5.99</td>
</tr>
<tr>
<td><strong>St Dev</strong></td>
<td>8.97</td>
<td>8.33</td>
<td>5.85</td>
</tr>
</tbody>
</table>

The impact of the WWTP upgrades also would have affected TDP concentrations during periods of moderate or high flow, though likely to a lesser
extent. Figure 3-4 contains a plot of biweekly flow volume and average TDP load. The figure illustrates that medium to large TDP loads correspond with periods of moderate to high flow. In such loading periods, the relative contribution to the total load from WWTPs is much smaller than it is during lower flow conditions.

Occasionally during low-flow conditions, the sum of the measured WWTP loads is greater than the load measured in the river at Beerston (see Figure 3-5). This could be due to sampling error, as the time series were obtained using different sampling frequencies, or to the fact that three of the four major WWTPs are located more than 30 km upstream from the gage at Beerston so that some portion of the TDP load from the WWTPs is taken up by algae and other plants or organisms.

![Figure 3-4: Plot of biweekly flow volume (million cubic meters) and average TDP load (kg/day)](image)

Figure 3-4: Plot of biweekly flow volume (million cubic meters) and average TDP load (kg/day)
3.2 Model Performance

All four variations of the log-linear regression model (LLP, LLP-I, LLP-F, and LLP-F-I, Table 2-3) were analyzed using ordinary least squares (OLS), which assumes that the residuals, \( e_i \), are independent and identically distributed (i.i.d). Tests for heteroscedasticity, or non-constant variance in the residuals, did not show any strong relationships between the residuals and the observed data or covariates. There was, however, significant correlation in the regression residuals for all four statistical models. The correlation coefficient was estimated using:

\[
\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=2}^{n} e_t^2}
\]

[3-1]

A hypothesis test using the normally distributed test statistic \( W \) (Eq. [3-2]) was used to assess the statistical significance of \( \hat{\rho} \) (White, 1992).
The correlation coefficient \( \hat{\rho} \) highly significant for all models considered (p < 0.0001). While the OLS estimator of the vector of regression coefficients, \( \beta \), is unbiased under all conditions, strong autocorrelation in the residual errors, \( \varepsilon_i \), will result in less efficient parameter estimation and an incorrect computation of the covariance matrix, \( \Sigma \). This incorrect computation will often lead to artificially low standard errors (Greene, 2003). Standard errors play an important role in determining statistical significance and in performing statistical inference, both of which are particularly important for this study. In order to correct for the presence of autocorrelation, a feasible generalized least squares (FGLS) approach for AR(1) residuals was employed. A generalized least squares (GLS) analysis weights the model residuals appropriately to reflect the adopted model for the correlation structure. The subsequent analysis and discussion in this study are based on the results of the FGLS as opposed to OLS regressions.

For generalized least squares (GLS) models, there is no precise counterpart to the \( R^2 \) commonly reported for OLS regressions. While several have been proposed, they all suffer from one or more shortcomings and cannot reliably be used to compare models. The drawback to using the traditional \( R^2 \) metric with an FGLS model for autocorrelation is that it might incorrectly suggest that the fit was degraded by the FGLS correction (Greene, 2003). This behavior was not noticeable in the results of this study, and thus the traditional adjusted \( R^2 \) metric, which is based on \( R^2 \) and penalizes the addition of unnecessary covariates, was used to assess model performance.

\[
W = \frac{\hat{\rho}}{\sqrt{n}}
\]
The results of the regression analyses are provided in Table 3-2. Based on an FGLS analysis, the LLP model recommended by EPA (Eq. [2-5]) explained 69% of the variability in measured TDP loads (using log units), but failed to detect statistically significant BMP treatment effects. The LLP-I model (Eq. [2-7]), which is the LLP model with a BMP-phosphorus interaction term, explained 76% of the variability in observed loads (using log units) and detected statistically significant treatment effects. The two log-linear models with flow, LLP-F (Eq. [2-9]) and LLP-F-I (Eq. [2-3]), both explained 87% of the variability in measured TDP loads (using log units) and both detected significant treatment effects, though the BMP-phosphorus interaction term was not significant in the LLP-F-I model.

Table 3-2: Regression coefficients and adjusted $R^2$ values for log-linear paired watershed models (Table 2-3)

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>f</th>
<th>g</th>
<th>Adj. $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLP</td>
<td>1.12</td>
<td>0.59</td>
<td>-</td>
<td>0.07 (ns)</td>
<td>-</td>
<td>0.69</td>
</tr>
<tr>
<td>LLP-I</td>
<td>-0.80</td>
<td>1.10</td>
<td>-</td>
<td>0.66</td>
<td>-0.56</td>
<td>0.76</td>
</tr>
<tr>
<td>LLP-F</td>
<td>1.02</td>
<td>0.09</td>
<td>0.71</td>
<td>-0.80</td>
<td>-</td>
<td>0.87</td>
</tr>
<tr>
<td>LLP-F-I</td>
<td>0.66</td>
<td>0.20</td>
<td>0.72</td>
<td>-0.66</td>
<td>-0.11 (ns)</td>
<td>0.87</td>
</tr>
</tbody>
</table>

(ns) Coefficient not statistically significant (p > 0.05)

From the results of the log-linear models it is clear that the addition of covariates to the traditional LLP model results in improved model performance and ability to detect BMP treatment effects. The peak flow covariate in particular provided the largest increase in performance, increasing the adjusted $R^2$ from 0.69 to 0.87. The results of the regression analysis are similar to those reported by Bishop et al. (2005), who also showed that the LLP model recommended by EPA performed poorly in comparison with the multivariate models and failed to detect significant BMP effects.
The regression outputs also show that adding the peak flow covariate caused the coefficient, $b$, on the SWAT load covariate to decrease significantly. Although $b$ and $c$ were significant at the 1% level in the LLP-F and LLP-F-I models, the smaller value of $b$ could result in an inadequate representation of important changes that occurred in the watershed, such as the dairy cow population decline and WWTP upgrades. While it is unclear whether the cow population declines would have significantly affected TDP loads during the study period, the monitoring data presented earlier show that WWTPs were a significant source of TDP during the pre-BMP period. Thus, failure to properly account for WWTP upgrades in the statistical model will likely to lead to an overestimate of BMP effectiveness.

3.3 Disaggregated Statistical Models

In order to ensure that the effects of treatment plant upgrades were adequately represented in the statistical model, a separate variable, $wwtp_i$, was added for WWTP loads. Also, Equation [2-4] shows that the regression models with flow (LLP-F and LLP-F-I) estimate $P_i^m$ by multiplying $Q_i^{m,peak}$ and $P_i^s$, when in fact, loads are computed by multiplying flow and concentration. Thus, the SWAT concentration, $C_i^s$, was substituted for the SWAT load, $P_i^s$, and average flow was substituted for peak flow to achieve a more physically meaningful model. Because WWTPs were added as a separate covariate, they were removed from the SWAT simulation model for the alternative models described here. Consider the following disaggregated paired model with a BMP-phosphorus interaction term:

$$\ln(P_i^m) = \ln[a + b (wwtp_i) + (Q_i^{m,avg})^c (C_{si})^{d+gk_i}\exp(h + f k_i)] + \epsilon_i$$  

[3-3]

or (written without the natural log transformation)
The term disaggregated refers to the separation of loads from point and non-point sources in the regression equation. For abbreviation and notation purposes, the disaggregated paired watershed model with the BMP-phosphorus interaction term (Eq. [3-3]) will be referred to as the DP-I model. When the BMP-phosphorus interaction term is left out of the model (i.e. $g=0$), the DP-I model becomes the DP model:

$$ln(P_i^{m}) = ln[a + b (wwtp_i) + (Q_i^{m \; avg})^c (C_i^{s})^d \exp{[h + f \; k_i]}] + \epsilon_i$$ \hspace{1cm} [3-5]$$

or (written without the natural log transformation)

$$P_i^{m} = [a + b (wwtp_i) + (Q_i^{m \; avg})^c (C_i^{s})^d \exp{[h + f \; k_i]}] \exp(\epsilon_i)$$ \hspace{1cm} [3-6]$$

For the DP and DP-I models, OLS cannot be used because Equations [3-3] and [3-5] contain products of unknown regression coefficients. Instead, a non-linear feasible generalized least squares analysis with a correction for AR(1) residuals was employed. Table 3-3 provides a summary of the two disaggregated paired watershed models considered. The results appear in Table 3-4.

**Table 3-3: Summary of disaggregated paired (DP) models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Eq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>$ln(P_i^{m}) = ln[a + b (wwtp_i) + (Q_i^{m ; avg})^c (C_i^{s})^d \exp{[h + f ; k_i]}] + \epsilon_i$</td>
<td>[3-5]</td>
</tr>
<tr>
<td>DP-I</td>
<td>$ln(P_i^{m}) = ln[a + b (wwtp_i) + (Q_i^{m ; avg})^c (C_i^{s})^d + g ; k_i \exp{[h + f ; k_i]}] + \epsilon_i$</td>
<td>[3-3]</td>
</tr>
</tbody>
</table>
Both the DP and DP-I models explained 88% of the variability in observed loads (using log units), although three of the coefficients in the DP-I model were not statistically significant \((d, f, g)\). Thus, of the two models, the simpler DP model without the BMP-phosphorus interaction term is the preferred choice. In addition to performing slightly better than the best log-linear model, and having statistically significant parameters, the DP model also explicitly accounts for WWTP loads and provides a more physically reasonable model structure, thus, enhancing the credibility of the results.

### 3.4 BMP Treatment Effects

A variety of statistical models were developed and tested, with the goal of explaining variability in the observed data and separating the effects of BMPs from those of WWTP upgrades and declines in the dairy cow population. The DP model (Eq. [3-5]) best met the objectives of model performance and ability to separate BMPs effects from WWTP upgrades and cow population declines. Thus, the DP model was selected over the other paired watershed models for the computation of BMP treatment effects. A load reduction factor was computed using Equation [2-13] in Section 2.4.5, by substituting the analytical expectations with sample averages from the entire period of record. Expectations were not taken analytically for this model as

### Table 3-4: Regression coefficients and adjusted R\(^2\) values for DP models

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>Adj. R(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>-1.87</td>
<td>0.35</td>
<td>1.02</td>
<td>0.13</td>
<td>-0.59</td>
<td>-</td>
<td>2.29</td>
<td>0.88</td>
</tr>
<tr>
<td>DP-I</td>
<td>-1.95</td>
<td>0.36</td>
<td>1.03</td>
<td>0.12</td>
<td>-0.93</td>
<td>-0.03</td>
<td>2.11</td>
<td>0.88</td>
</tr>
</tbody>
</table>

(ns) Coefficient not statistically significant (p > 0.05)
the expectations become much more complicated with the loss of the mathematically
convenient log-linear model structure.

Based on the results of the non-linear regression with the selected DP model,
TDP loads were reduced on average by 41.5%. Confidence intervals were obtained
using a simple Monte Carlo uncertainty analysis of Eq. [2-13] with period-of-record
flows and dissolved phosphorus loads. One thousand parameter coefficient vectors
were sampled from a multivariate normal (MVN) distribution, where the mean and
covariance were defined using the outputs from the DP regression model. The
purpose of randomly varying all of the coefficients together was to accurately capture
relationships between all of the coefficients in computing the distribution of the load
reduction factor. Based on the Monte Carlo analysis, the 95% confidence bounds for
the load reduction factor were 33.6 to 49.2%. Based on the original regression and its
error analysis, the Monte Carlo simulation indicates that if the experiment were
repeated many times, 95% of such confidence intervals would contain the true load
reduction factor.

The computed load reductions are similar to those reported by Bishop et al.
(2005), whose study was performed on a small farm inside the Cannonsville basin.
Those authors reported that TDP event loads were reduced by 43% in the post-BMP
period, with 95% confidence bounds of 36 and 49%. Although one might have
expected to see smaller reductions at the large-basin scale, the Cannonsville basin and
the farm watershed have essentially the same land use distribution and should respond
similarly to BMPs. Bishop et al. (2005) also point out that the herd size on the farm
they studied experienced a gradual increase in size of about 30% during the study
period (Hively, 2004), which was not reflected in the statistical analysis. Had the herd
size remained constant instead of increasing, the authors would have likely reported
larger reductions in event TDP loads.
In the farm-level study, Bishop et al. (2005) point out that TDP loads were reduced despite the fact that none of the BMPs reduced the amount of phosphorus imported onto the farm as feed or fertilizer or the amount exported as products. Therefore, if the mass balance on the farm did not change, TDP concentrations in the stream were reduced by retaining a greater portion of phosphorus on the farm. If the mass balance of phosphorus is not changed, accumulation of soil phosphorus will continue. As a result, soil phosphorus levels could accumulate until they reach a saturation point, above which TDP concentrations in surface runoff may increase (Beuachemin and Simard, 1999; McDowell and Sharples, 2001).

The extent to which certain individual BMPs have been implemented in the Cannonsville basin is publicly available, as detailed records are private due to confidentiality concerns. Thus, it is not possible to determine what percentages of BMPs were focused on limiting phosphorus transport or altering the phosphorus mass balance. While the results of this study provide encouraging evidence of BMP effectiveness at the large-basin scale, Tolson and Shoemaker argue that a long-term emphasis should be placed on achieving a more sustainable phosphorus mass balance in order to avoid the accumulation of soil phosphorus (Tolson and Shoemaker, 2004; Tolson, 2005). Alterations in the mass balance could be achieved by increasing the use of precision feeding and the exportation of excess manure, and by decreasing the practice of starter fertilization. Future declines in the dairy cow population would also contribute to a more sustainable phosphorus mass balance.

3.5 Power Analysis and Minimum Detectable Treatment Effects

Statistical power is important when designing monitoring studies to detect treatment effects (Lettenmaier, 1976; Hirsch et al., 1982; Ward et al., 1990; Cooke et al., 1995; Galeone, 1999; Loftis et al., 2001; Bishop et al., 2005). Power is defined as
the probability that a statistical test will correctly detect a treatment effect of a given magnitude, and is a function of the magnitude of the treatment effect, \( f \), the required significance level \( \alpha \), and the variance \( \sigma_e^2 \) of the residual errors \( e_i \). A power analysis for the DP model (Eq. 3-3), which was the preferred paired watershed model (Section 3.3), was performed using the simple equation used in Bishop et al. (2005):

\[
Pr[\text{Conclude that } f < 0 \mid f, \alpha, \sigma_f] = \pi(f) \equiv 1 - \Phi\left[z_\alpha + \frac{f}{\sigma_f}\right]
\]

where \( f \) is the regression coefficient on \( k_i \) (Eq. [3-5]), \( \Phi \) is the cumulative distribution function for the standard normal distribution, \( z_\alpha = \Phi[1 - \alpha] \), is the critical value of the normal distribution for a type-I error of \( \alpha \), and \( \sigma_f \) is the standard error of \( f \) that results from the regression analysis with the DP model, and depends on \( \sigma_e \). Figure 3-6 provides a plot of power \( \pi \) vs. load reduction factor for the DP model (Eq. [3-3]). Load reductions were computed using Eq. [2-13] with a range of values of \( f \) and the specific available data from the pre- and post-BMP periods employed in this study. The statistical power is approximately equal to one for a load reduction of 40%, for all values of \( \alpha \) in the plot. Thus, if there really were a 40% reduction in TDP loads, we would be sure to detect it.
Another value commonly reported in monitoring studies is the minimum detectable effect (MDE) (Bunte and MacDonald, 1999), minimum detectable change (MDC) (Spooner et al., 1987), or minimum detectable treatment effect (MDTE) (Bishop et al., 2005), which is defined as the smallest BMP treatment effect that would be statistically significant for a given type-I error, $\alpha$, and type-II error, $\beta$. Based upon Eq. [3-7], the smallest detectable effect for a given sample size and design matrix can be computed from:

$$f_{MDTE} = -(z_\alpha + z_\beta)\sigma_f$$  \[3-8\]
where \( z_\alpha \) is the critical value for a Type-I error of \( \alpha \), \( z_\beta \) is the critical value for a Type-II error of \( \beta \), and \( \sigma_f \) is the standard error of the regression coefficient \( f \) for a particular pre-BMP and post-BMP data set. For example, if \( \alpha = 1\% \) and \( \beta = 20\% \):

\[
f_{MDTE} = -(2.33 + 0.84)\sigma_f = -3.17\sigma_f
\]

Given the value of \( f_{MDTE} \), Eq. [2-13] can be used to compute the actual MDTE for the DP model (Eq. [3-5]). Figure 3-7 provides a plot of MDTE, which is the minimum detectable load reduction factor, versus \( \beta \) for several values of \( \alpha \). For \( \alpha = 1\% \) and \( \beta = 20\% \), the minimum detectable treatment effect for the DP model was 21.8%. The observed load reduction of 41.5% was larger than the MDTE by a factor of two. If one had different numbers of observations in the pre- and post-BMP periods, or the observed values of the independent variables in those periods were different, the power and MDTE values would change.

### 3.6 Single Watershed Models

In some situations, it may not be feasible to do a paired analysis if there are inadequate funds or resources for either maintaining a monitoring station at a physical control site or developing a simulated control site. In these cases, it may be useful to implement a single watershed statistical model, which makes use only of monitoring data from the site of interest. These models are simple to construct and potentially powerful.

The most basic model that one can employ is the following log-linear single watershed model:

\[
\ln(p_i^m) = a + c \ln(Q_i^{m, peak}) + f k_i + \epsilon_i
\]
Figure 3-7: of MDTE vs. $\beta$ for the DP model (Eq. [3-5]) using three different values for $\alpha$ and the full set of biweekly data from the pre- and post-BMP periods
or (written without the natural log transformation)

\[ p_i^m = \left( Q_i^{m\,peak} \right)^c \exp\{a + f k_i + \varepsilon_i\} \]  \hspace{1cm} [3-11]

An alternative single watershed model based on the DP model structure is:

\[ \ln(p_i^m) = \ln\left[ a + b (\text{wwtp}_i) + \left( Q_i^{m\,peak} \right)^{c+g k_i} \exp\{d + f k_i\} \right] + \varepsilon_i \]  \hspace{1cm} [3-12]

or (written without the natural log transformation)

\[ p_i^m = \left[ a + b (\text{wwtp}_i) + \left( Q_i^{m\,peak} \right)^{c+g k_i} \exp\{d + f k_i\} \right] \exp(\varepsilon_i) \]  \hspace{1cm} [3-13]

The model above was tested with and without the BMP-flow interaction term, \( g k_i \).

Removal of the BMP-flow interaction term results in the following model:

\[ \ln(p_i^m) = \ln\left[ a + b (\text{wwtp}_i) + \left( Q_i^{m\,peak} \right)^c \exp\{d + f k_i\} \right] + \varepsilon_i \]  \hspace{1cm} [3-14]

or (written without the natural log transformation)

\[ p_i^m = \left[ a + b (\text{wwtp}_i) + \left( Q_i^{m\,peak} \right)^c \exp\{d + f k_i\} \right] \exp(\varepsilon_i) \]  \hspace{1cm} [3-15]

The three single watershed models will be referred to as log-linear single watershed model (LLS), disaggregated single watershed model (DS), and disaggregated single watershed model with the BMP-flow interaction term (DS-I).

Table 3-5 provides a summary of the single watershed models. All of the single watershed regressions were analyzed with FGLS employing the needed corrections for AR(1) residuals.
Table 3-5: Summary of single watershed regression models

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Eq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLS</td>
<td>(\ln(P_i^{m}) = a + c \ln(Q_i^{m, peak}) + f \ k_i + \varepsilon_i)</td>
<td>[3-10]</td>
</tr>
<tr>
<td>DS</td>
<td>(\ln(P_i^{m}) = \ln [a + b \ \text{(wwtp}_i) + (Q_i^{m, peak})^c \ \exp(d + f \ k_i)] + \varepsilon_i)</td>
<td>[3-14]</td>
</tr>
<tr>
<td>DS-I</td>
<td>(\ln(P_i^{m}) = \ln [a + b \ \text{(wwtp}_i) + (Q_i^{m, peak})^{c+g \ k_i} \ \exp(d + f \ k_i)] + \varepsilon_i)</td>
<td>[3-12]</td>
</tr>
</tbody>
</table>

Table 3-6: Regression coefficients and adjusted \(R^2\) values for single watershed models

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>f</th>
<th>g</th>
<th>Adj. (R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLS</td>
<td>1.12</td>
<td>-</td>
<td>0.83</td>
<td>-</td>
<td>-0.93</td>
<td>-</td>
<td>0.86</td>
</tr>
<tr>
<td>DS</td>
<td>-1.84</td>
<td>0.32</td>
<td>0.92</td>
<td>0.57</td>
<td>-0.70</td>
<td>-</td>
<td>0.89</td>
</tr>
<tr>
<td>DS-I</td>
<td>-2.42</td>
<td>0.38</td>
<td>1.03</td>
<td>0.18</td>
<td>-0.08</td>
<td>-0.17</td>
<td>0.90</td>
</tr>
</tbody>
</table>

The single watershed regression results in Table 3-6 indicate that the models considered for the Cannonsville basin were effective at both explaining variability in the observed data and detecting statistically significant BMP treatment effects. While removing the SWAT covariates from the model did not have a significant impact on model performance, it did result in the loss of important information regarding changes in the dairy cow population, which was incorporated into the SWAT covariates for the paired models. As a result, the disaggregated single watershed models tend to overestimate the effects of BMPs.

Table 3-7 – Comparison of load reductions for DP, DS-I, and LLS models

<table>
<thead>
<tr>
<th>Model</th>
<th>Load Reduction (%)</th>
<th>CI [0.05]</th>
<th>Mean</th>
<th>CI [0.95]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>33.6</td>
<td>41.5</td>
<td>49.2</td>
<td></td>
</tr>
<tr>
<td>DS-I</td>
<td>31.9</td>
<td>56.5</td>
<td>73.4</td>
<td></td>
</tr>
<tr>
<td>LLS</td>
<td>51.3</td>
<td>60.5</td>
<td>68.2</td>
<td></td>
</tr>
</tbody>
</table>
Consider the mean load reductions and confidence intervals for the DP, DS-I, and LLS models in Table 3-7. The mean and confidence intervals for the DP and DS-I model were obtained using the Monte Carlo approach discussed in Section 3.4, and the values for the LLS model were obtained analytically using Eq. [2-11]. The DP model separates the effects of BMPs from those of WWTP upgrades and declines in the dairy cow population. The DS-I model incorporates WWTP upgrades, but does not address the decrease in the cow population and the LLS model does not address WWTP upgrades or changes in the cattle population. The mean load reduction for the DS-I model is larger than the mean reduction for the DP model, and the mean reduction for the LLS model is the largest of the three. Thus, failing to incorporate information regarding WWTP upgrades or changes in the dairy cow population led to significant overestimates of BMP effectiveness. These results demonstrate the importance of including the SWAT covariates ($C_i^s$ or $P_i^s$) in the statistical model in order to separate BMP effects.
CHAPTER 4
CONCLUSIONS

A purpose of this study was to present a rigorous statistical analysis of the effects of BMPs on dissolved phosphorus loads from dairy farms in the Cannonsville Reservoir basin, a drinking water supply for New York City. This paper introduces the use of a simulation model as a control site in a paired watershed analysis as a technique that can be adopted when an actual control watershed is not available or feasible. In our analysis, the simulated control site also played a critical role in separating the effects of BMPs from those of WWTP upgrades and declines in the dairy cow population. A variety of univariate and multivariate models were considered along with different model structures. In addition to paired watershed models, several single watershed regression models were also tested. While the regression residuals did not exhibit heteroscedastic behavior, serial correlation in the residuals was highly significant. Thus, a feasible generalized least squares (FGLS) analysis with a correction for AR(1) residuals was employed.

This study shows that paired watershed analysis with a simulated control site is an innovative statistical method of great value in many applications. The results also demonstrate that multivariate paired watershed models are superior to the univariate approach recommended by EPA, both in terms of model performance and ability to detect statistically significant BMP treatment effects. Modifications to the traditional log-linear regression model structure improved our ability to correct for WWTP upgrades and changes in the dairy cow population, and provided a more physically meaningful statistical model. Single watershed models also performed well, but could not control for changes in the cow population and, thus produced overly optimistic estimates of BMP effectiveness. Results from the best paired model indicate that
biweekly average TDP loads were reduced by 41.5%, with 95% confidence bounds of 33.6 and 49.2%. These results provide strong evidence that the widespread implementation of BMPs in the Cannonsville basin have been successful at improving water quality.
APPENDIX

A.1 Computation of BMP Treatment Effects for Log-linear Paired Models with BMP-Phosphorus Interaction Term

This section of the Appendix includes derivations of Eq. [2-14] and [2-15] in Section 2.4.4.

A.1.1 Load Reduction Factor for LLP-F-I Model

Recall that

\[
\% \text{Reduction} = 1 - \frac{E[P^m]}{E[P^m|f = g = 0]} \quad \text{[A-1]}
\]

and that for the LLP-F-I model (Eq. [2-3]),

\[
\frac{E[P^m]}{E[P^m|f = g = 0]} = \frac{E[(P^s)^b (Q^{m \text{ peak}})^c \exp\{a + f + g [\ln(P^s) - p] + \varepsilon\}]}{E[(P^s)^b (Q^{m \text{ peak}})^c \exp\{a + \varepsilon\}]} \quad \text{[A-2]}
\]

Note that the expectations are taken over the entire period of record in order to maximize the precision of the load reduction factor.

First, consider the expectation in the numerator in Eq. [A-2], which can be written as

\[
E[P^m] = E[(P^s)^b (Q^{m \text{ peak}})^c \exp\{a + f + g [\ln(P^s) - p] + \varepsilon\}] = E[\exp\{a + b \ln(P^s) + c \ln(Q^{m \text{ peak}}) + f + g [\ln(P^s) - p] + \varepsilon\}] \quad \text{[A-3]}
\]
Taking the analytical expectation, and assuming that all regression coefficients and $p$
are constants, only $\ln(p_s)$ and $\ln(Q^m_{\text{peak}})$, and $\varepsilon$ are random variables. Thus,
Eq.[A-3] can be written as

$$E[P^m] = \exp[a + f]E[\exp[b \ln(p_s) + c \ln(Q^m_{\text{peak}}) + g [\ln(p_s) - p] + \varepsilon]]$$  \[A-4\]

which can be further rewritten as

$$E[P^m] = \exp[a + f - gp]E[\exp((b + g)\ln(p_s) + c \ln(Q^m_{\text{peak}}) + \varepsilon)]$$  \[A-5\]

We further assume that $\ln(p_s)$ and $\ln(Q^m_{\text{peak}})$, and $\varepsilon_i$ are normally distributed

$$\ln(p_s) \sim N[\mu_p, \sigma_p^2] \quad \ln(Q^m_{\text{peak}}) \sim N[\mu_Q, \sigma_Q^2] \quad \varepsilon \sim N[0, \sigma_\varepsilon^2]$$  \[A-6\]

Thus,

$$(b + g)\ln(p_s) \sim N[(b + g)\mu_p, (b + g)^2\sigma_p^2] \quad c\ln(Q^m_{\text{peak}}) \sim N[c\mu_Q, c^2\sigma_Q^2]$$  \[A-7\]

and

$$[(b + g)\ln(p_s) + c\ln(Q^m_{\text{peak}}) + \varepsilon] \sim N[(b + g)\mu_p + c\mu_Q, (b + g)^2\sigma_p^2 + c^2\sigma_Q^2 + \sigma_\varepsilon^2]$$  \[A-8\]

Recall that if $Z = \ln(X)$ and $Y \sim N[\mu_Y, \sigma_Y^2]$,

$$E[X] = \exp \left( \mu_Y + \frac{1}{2} \sigma_Y^2 \right)$$  \[A-9\]
Thus, employing the assumption that \( \epsilon \) is uncorrelated with \( \ln(P^s) \) and \( \ln(Q^{m\,peak}) \),

\[
E[\exp\{ (b + g)\ln(P^s) + c \ln(Q^{m\,peak}) + \epsilon \}]
\]

\[
= \exp\{ (b + g)\mu_P + c\mu_Q + 2(b + g)c\sigma_{PQ}
+ 1/2 \left( (b + g)^2\sigma_P^2 + c^2\sigma_Q^2 + \sigma_\epsilon^2 \right) \}
\]  

[A-10]

and

\[
E[P^m] = \exp\{ a + f - gp \} \exp\{ (b + g)\mu_P + c\mu_Q + 2(b + g)c\sigma_{PQ}
+ 1/2 \left( (b + g)^2\sigma_P^2 + c^2\sigma_Q^2 + \sigma_\epsilon^2 \right) \}
\]  

[A-11]

Now consider the denominator in Eq. [A-2]:

\[
E[P^m|f = g = 0] = E[(P^s)^b(Q^{m\,peak})^c\exp\{ a + \epsilon \}]
\]  

[A-12]

Following the same strategy used previously,

\[
E[P^m|f = g = 0] = \exp\{ a \} \exp\{ b\mu_P + c\mu_Q + 2bc\sigma_{PQ} + 1/2 \left( b^2\sigma_P^2 + c^2\sigma_Q^2 + \sigma_\epsilon^2 \right) \}
\]  

[A-13]

Thus,

\[
\frac{E[P^m]}{E[P^m|f = g = 0]}
\]

\[
= \frac{\exp\{ f - gp \} \exp\{ (b + g)\mu_P + 2(b + g)c\sigma_{PQ} + 1/2 \left( (b + g)^2\sigma_P^2 \right) \}}{\exp\{ b\mu_P + 2bc\sigma_{PQ} + 1/2 \left( b^2\sigma_P^2 \right) \}}
\]  

[A-14]
which further reduces to

\[
\frac{E[P^m]}{E[P^m|f = g = 0]} = \frac{\exp\{f + g(\mu_p - p) + 2(b + g)c\sigma_{pq} + 1/2((b + g)^2\sigma_p^2)\}}{\exp\{2bc\sigma_{pq} + 1/2(b^2\sigma_p^2)\}}
\]

[A-15]

**A.1.2 Load Reduction Factor for LLP-I Model**

Consider the derived reduction factor for the LLP-F-I model in Eq. [A-15]. This can be converted to the reduction factor for the LLP-I model simply by removing the flow-related terms from the equation:

\[
\frac{E[P^m]}{E[P^m|f = g = 0]} = \frac{\exp\{f + 1/2((b + g)^2\sigma_p^2)\}}{\exp\{1/2(b^2\sigma_p^2)\}}
\]

[A-16]

**A.2 Regression Outputs for Statistical Models**

This section contains the full regression output tables for the models introduced in this study. The table captions indicate which tables correspond to which regression models.

**Table A-1: Regression outputs for LLP model (Eq. [2-5])**

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.12</td>
<td>0.181</td>
<td>6.21</td>
</tr>
<tr>
<td>b</td>
<td>0.59</td>
<td>0.035</td>
<td>16.68</td>
</tr>
<tr>
<td>f</td>
<td>0.07</td>
<td>0.163</td>
<td>0.40</td>
</tr>
</tbody>
</table>
Table A-2: Regression outputs for LLP-I model (Eq. [2-7])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-0.80</td>
<td>0.354</td>
<td>-2.25</td>
</tr>
<tr>
<td>b</td>
<td>1.10</td>
<td>0.091</td>
<td>12.11</td>
</tr>
<tr>
<td>f</td>
<td>0.66</td>
<td>0.160</td>
<td>4.16</td>
</tr>
<tr>
<td>g</td>
<td>-0.56</td>
<td>0.097</td>
<td>-5.79</td>
</tr>
</tbody>
</table>

Table A-3: Regression outputs for LLP-F model (Eq. [2-9])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.02</td>
<td>0.119</td>
<td>8.55</td>
</tr>
<tr>
<td>b</td>
<td>0.09</td>
<td>0.039</td>
<td>2.26</td>
</tr>
<tr>
<td>c</td>
<td>0.75</td>
<td>0.046</td>
<td>16.30</td>
</tr>
<tr>
<td>f</td>
<td>-0.80</td>
<td>0.120</td>
<td>-6.69</td>
</tr>
</tbody>
</table>

Table A-4: Regression outputs for LLP-F-I model (Eq. [2-3])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.66</td>
<td>0.265</td>
<td>2.49</td>
</tr>
<tr>
<td>c</td>
<td>0.72</td>
<td>0.049</td>
<td>14.81</td>
</tr>
<tr>
<td>b</td>
<td>0.20</td>
<td>0.085</td>
<td>2.38</td>
</tr>
<tr>
<td>f</td>
<td>-0.66</td>
<td>0.153</td>
<td>-4.30</td>
</tr>
<tr>
<td>g</td>
<td>-0.11</td>
<td>0.073</td>
<td>-1.52</td>
</tr>
</tbody>
</table>

Table A-5: Regression outputs for DP model (Eq. [3-5])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-1.862</td>
<td>0.332</td>
<td>-5.604</td>
</tr>
<tr>
<td>b</td>
<td>0.352</td>
<td>0.046</td>
<td>7.724</td>
</tr>
<tr>
<td>c</td>
<td>1.016</td>
<td>0.053</td>
<td>19.047</td>
</tr>
<tr>
<td>d</td>
<td>0.133</td>
<td>0.054</td>
<td>2.472</td>
</tr>
<tr>
<td>h</td>
<td>2.291</td>
<td>0.667</td>
<td>3.437</td>
</tr>
<tr>
<td>f</td>
<td>-0.587</td>
<td>0.083</td>
<td>-7.051</td>
</tr>
</tbody>
</table>
Table A-6: Regression outputs for DP-I model (Eq. [3-3])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-1.952</td>
<td>0.391</td>
<td>-4.997</td>
</tr>
<tr>
<td>b</td>
<td>0.359</td>
<td>0.050</td>
<td>7.161</td>
</tr>
<tr>
<td>c</td>
<td>1.033</td>
<td>0.054</td>
<td>19.147</td>
</tr>
<tr>
<td>d</td>
<td>0.121</td>
<td>0.089</td>
<td>1.359</td>
</tr>
<tr>
<td>h</td>
<td>2.109</td>
<td>1.031</td>
<td>2.045</td>
</tr>
<tr>
<td>f</td>
<td>-0.927</td>
<td>1.094</td>
<td>-0.848</td>
</tr>
<tr>
<td>g</td>
<td>-0.031</td>
<td>0.100</td>
<td>-0.310</td>
</tr>
</tbody>
</table>

Table A-7: Regression outputs for LLS model (Eq. [3-10])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.124</td>
<td>0.112</td>
<td>10.014</td>
</tr>
<tr>
<td>c</td>
<td>0.829</td>
<td>0.028</td>
<td>29.740</td>
</tr>
<tr>
<td>f</td>
<td>-0.933</td>
<td>0.109</td>
<td>-8.613</td>
</tr>
</tbody>
</table>

Table A-8: Regression outputs for DS model (Eq. [3-14])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-1.803</td>
<td>0.315</td>
<td>-5.720</td>
</tr>
<tr>
<td>b</td>
<td>0.315</td>
<td>0.044</td>
<td>7.220</td>
</tr>
<tr>
<td>c</td>
<td>0.919</td>
<td>0.038</td>
<td>23.950</td>
</tr>
<tr>
<td>d</td>
<td>0.571</td>
<td>0.155</td>
<td>3.693</td>
</tr>
<tr>
<td>f</td>
<td>-0.702</td>
<td>0.077</td>
<td>-9.158</td>
</tr>
</tbody>
</table>

Table A-9: Regression outputs for DS-I model (Eq. [3-12])

<table>
<thead>
<tr>
<th>coeff</th>
<th>se</th>
<th>t</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>-2.408</td>
<td>0.398</td>
<td>-6.047</td>
</tr>
<tr>
<td>b</td>
<td>0.378</td>
<td>0.050</td>
<td>7.618</td>
</tr>
<tr>
<td>c</td>
<td>1.028</td>
<td>0.067</td>
<td>15.444</td>
</tr>
<tr>
<td>g</td>
<td>-0.171</td>
<td>0.075</td>
<td>-2.284</td>
</tr>
<tr>
<td>d</td>
<td>0.179</td>
<td>0.252</td>
<td>0.710</td>
</tr>
<tr>
<td>f</td>
<td>-0.083</td>
<td>0.282</td>
<td>-0.295</td>
</tr>
</tbody>
</table>
A.3 Additional Tables and Figures

This section of the appendix contains tables and figures that are relevant but were not included in the body of the thesis.

Table A-10: Summary of all statistical models, where LLP, DP, and DS refer to log-linear paired, disaggregated paired, and disaggregated single, respectively. F stands for flow and I stands for interaction.

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
<th>Eq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLP</td>
<td>( \ln(P_i^m) = a + b \ln(P_{i}^s) + f k_i + \varepsilon_i )</td>
<td>[2-5]</td>
</tr>
<tr>
<td>LLP-I</td>
<td>( \ln(P_i^m) = a + b \ln(P_{i}^s) + f k_i + g k_i[\ln(P_{i}^s) - p] + \varepsilon_i )</td>
<td>[2-7]</td>
</tr>
<tr>
<td>LLP-F</td>
<td>( \ln(P_i^m) = a + b \ln(P_{i}^s) + c \ln(Q_{i}^{m \text{peak}}) + f k_i + \varepsilon_i )</td>
<td>[2-9]</td>
</tr>
<tr>
<td>LLP-F-I</td>
<td>( \ln(P_i^m) = a + b \ln(P_{i}^s) + c \ln(Q_{i}^{m \text{peak}}) + f k_i + g k_i[\ln(P_{i}^s) - p] + \varepsilon_i )</td>
<td>[2-3]</td>
</tr>
<tr>
<td>DP</td>
<td>( \ln(P_i^m) = \ln[a + b (wwtp_i) + (Q_{i}^{m \text{avg}})^c (C_{i}^s)^d \exp[h + f k_i]] + \varepsilon_i )</td>
<td>[3-5]</td>
</tr>
<tr>
<td>DP-I</td>
<td>( \ln(P_i^m) = \ln[a + b (wwtp_i) + (Q_{i}^{m \text{avg}})^c (C_{i}^s)^d + g k_i \exp[h + f k_i]] + \varepsilon_i )</td>
<td>[3-3]</td>
</tr>
<tr>
<td>LLS</td>
<td>( \ln(P_i^m) = a + c \ln(Q_{i}^{m \text{peak}}) + f k_i + \varepsilon_i )</td>
<td>[3-10]</td>
</tr>
<tr>
<td>DS</td>
<td>( \ln(P_i^m) = \ln[a + b (wwtp_i) + (Q_{i}^{m \text{peak}})^c \exp[d + f k_i]] + \varepsilon_i )</td>
<td>[3-12]</td>
</tr>
<tr>
<td>DS-I</td>
<td>( \ln(P_i^m) = \ln[a + b (wwtp_i) + (Q_{i}^{m \text{peak}})^c + g k_i \exp[d + f k_i]] + \varepsilon_i )</td>
<td>[3-14]</td>
</tr>
</tbody>
</table>

46
Table A-11: Sampling protocol for WBDR at Beerston (Longabucco et al., 1998)

<table>
<thead>
<tr>
<th>Stage at Walton (ft)</th>
<th>Sampling Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>≤ 3.00</td>
<td>≤ 0.91</td>
</tr>
<tr>
<td>3.01 - 3.50</td>
<td>0.92 - 1.07</td>
</tr>
<tr>
<td>3.51 - 4.00</td>
<td>1.07 - 1.22</td>
</tr>
<tr>
<td>4.01 - 4.50</td>
<td>1.22 - 1.38</td>
</tr>
<tr>
<td>4.51 - 6.00</td>
<td>1.38 - 1.83</td>
</tr>
<tr>
<td>6.01 - 7.50</td>
<td>1.83 - 2.29</td>
</tr>
<tr>
<td>7.51 - 8.50</td>
<td>2.29 - 2.59</td>
</tr>
<tr>
<td>8.51 - 9.00</td>
<td>2.59 - 2.74</td>
</tr>
<tr>
<td>9.01 - 9.00</td>
<td>2.74 - 2.90</td>
</tr>
<tr>
<td>9.51 - 10.00</td>
<td>2.90 - 3.05</td>
</tr>
<tr>
<td>10.01 - 11.00</td>
<td>3.05 - 3.35</td>
</tr>
<tr>
<td>11.01 - 13.00</td>
<td>3.36 - 3.97</td>
</tr>
<tr>
<td>&gt; 13.01</td>
<td>&gt; 3.97</td>
</tr>
</tbody>
</table>

Table A-12: Observed biweekly average flow rates (cms) during winter, spring, and summer

<table>
<thead>
<tr>
<th>Winter</th>
<th>Winter Pre</th>
<th>Winter Post</th>
<th>Spring Pre</th>
<th>Spring Post</th>
<th>Summer Pre</th>
<th>Summer Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>3.80</td>
<td>3.37</td>
<td>3.60</td>
<td>4.81</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>Q1</td>
<td>10.13</td>
<td>9.37</td>
<td>6.93</td>
<td>15.48</td>
<td>1.65</td>
<td>2.46</td>
</tr>
<tr>
<td>Median</td>
<td>16.05</td>
<td>18.59</td>
<td>10.88</td>
<td>21.31</td>
<td>2.44</td>
<td>5.87</td>
</tr>
<tr>
<td>Q3</td>
<td>28.35</td>
<td>32.78</td>
<td>19.42</td>
<td>26.21</td>
<td>4.51</td>
<td>17.90</td>
</tr>
<tr>
<td>Max</td>
<td>114.43</td>
<td>96.38</td>
<td>53.84</td>
<td>46.56</td>
<td>18.91</td>
<td>67.69</td>
</tr>
<tr>
<td>Mean</td>
<td>25.66</td>
<td>25.79</td>
<td>16.21</td>
<td>22.66</td>
<td>4.09</td>
<td>10.51</td>
</tr>
<tr>
<td>St Dev</td>
<td>27.40</td>
<td>22.75</td>
<td>14.93</td>
<td>11.04</td>
<td>4.25</td>
<td>12.37</td>
</tr>
<tr>
<td>N</td>
<td>28</td>
<td>42</td>
<td>12</td>
<td>21</td>
<td>24</td>
<td>40</td>
</tr>
</tbody>
</table>
Table A-13: Observed biweekly average flow rates (cms) during fall and full year

<table>
<thead>
<tr>
<th></th>
<th>Fall</th>
<th></th>
<th>Full</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Min</td>
<td>1.83</td>
<td>0.83</td>
<td>0.71</td>
<td>0.78</td>
</tr>
<tr>
<td>Q1</td>
<td>4.81</td>
<td>4.71</td>
<td>3.74</td>
<td>5.72</td>
</tr>
<tr>
<td>Median</td>
<td>8.37</td>
<td>11.00</td>
<td>9.00</td>
<td>14.19</td>
</tr>
<tr>
<td>Q3</td>
<td>23.66</td>
<td>24.40</td>
<td>18.36</td>
<td>24.66</td>
</tr>
<tr>
<td>Max</td>
<td>43.65</td>
<td>43.28</td>
<td>114.43</td>
<td>96.38</td>
</tr>
<tr>
<td>Mean</td>
<td>14.73</td>
<td>15.24</td>
<td>15.55</td>
<td>18.57</td>
</tr>
<tr>
<td>St Dev</td>
<td>13.15</td>
<td>13.95</td>
<td>19.77</td>
<td>17.81</td>
</tr>
<tr>
<td>N</td>
<td>20</td>
<td>20</td>
<td>84</td>
<td>123</td>
</tr>
</tbody>
</table>

Table A-14: Observed biweekly TDP loads (kg/day) during winter, spring, and summer

<table>
<thead>
<tr>
<th></th>
<th>Winter</th>
<th></th>
<th>Spring</th>
<th></th>
<th>Summer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Min</td>
<td>8.39</td>
<td>2.60</td>
<td>7.21</td>
<td>4.76</td>
<td>1.93</td>
<td>1.21</td>
</tr>
<tr>
<td>Q1</td>
<td>28.23</td>
<td>10.33</td>
<td>12.98</td>
<td>12.00</td>
<td>7.37</td>
<td>3.97</td>
</tr>
<tr>
<td>Median</td>
<td>48.82</td>
<td>20.07</td>
<td>31.08</td>
<td>26.36</td>
<td>19.57</td>
<td>11.89</td>
</tr>
<tr>
<td>Q3</td>
<td>76.60</td>
<td>45.39</td>
<td>52.80</td>
<td>32.93</td>
<td>31.27</td>
<td>24.30</td>
</tr>
<tr>
<td>Max</td>
<td>574</td>
<td>215</td>
<td>133</td>
<td>68</td>
<td>72</td>
<td>85</td>
</tr>
<tr>
<td>Mean</td>
<td>79.03</td>
<td>37.56</td>
<td>41.32</td>
<td>27.09</td>
<td>21.39</td>
<td>17.16</td>
</tr>
<tr>
<td>St Dev</td>
<td>108.87</td>
<td>42.97</td>
<td>37.57</td>
<td>17.88</td>
<td>16.88</td>
<td>17.73</td>
</tr>
<tr>
<td>N</td>
<td>28</td>
<td>42</td>
<td>12</td>
<td>21</td>
<td>24</td>
<td>40</td>
</tr>
</tbody>
</table>

Table A-15: Observed biweekly TDP loads (kg/day) during fall and full year

<table>
<thead>
<tr>
<th></th>
<th>Fall</th>
<th></th>
<th>Full</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Min</td>
<td>4.34</td>
<td>0.60</td>
<td>1.93</td>
<td>0.60</td>
</tr>
<tr>
<td>Q1</td>
<td>16.97</td>
<td>3.35</td>
<td>14.00</td>
<td>7.35</td>
</tr>
<tr>
<td>Median</td>
<td>29.16</td>
<td>13.93</td>
<td>30.49</td>
<td>16.21</td>
</tr>
<tr>
<td>Q3</td>
<td>70.87</td>
<td>47.59</td>
<td>59.62</td>
<td>36.14</td>
</tr>
<tr>
<td>Max</td>
<td>202</td>
<td>140</td>
<td>574</td>
<td>215</td>
</tr>
<tr>
<td>Mean</td>
<td>49.59</td>
<td>28.16</td>
<td>50.16</td>
<td>27.61</td>
</tr>
<tr>
<td>St Dev</td>
<td>49.25</td>
<td>35.44</td>
<td>72.17</td>
<td>32.23</td>
</tr>
<tr>
<td>N</td>
<td>20</td>
<td>20</td>
<td>84</td>
<td>123</td>
</tr>
</tbody>
</table>
Figure A-1: Residual plot for LLP-F model (Eq. [2-9])

\[ y = 0.52x - 0.00083 \]
REFERENCES


Delaware County Department of Watershed Affairs. 2002. Delaware County action plan DCAPII for watershed protection and economic vitality [Online]. Available at http://www.co.delaware.ny.us/departments/h2o/docs/dcap.pdf (verified June 25, 2008). Delaware County Board of Supervisors, Delhi, NY.


USEPA. 1997a. Techniques for tracking, evaluating, and reporting the implementation of non-point source control measures: Agriculture. EPA 841-B-97-010. USEPA Office of Water, Washington, D.C.

USEPA. 1997b. Linear regression for non-point source pollution analyses. EPA 841-B-97-007. USEPA Office of Water, Washington, D.C.


