MEASUREMENT AND MODELING OF STORMWATER FROM SMALL SUBURBAN WATERSHEDS IN VERMONT

A Dissertation Presented

by

Joel Nipper

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Natural Resources

January, 2016

Defense Date: September 30, 2015
Dissertation Examination Committee:

William B. Bowden, Ph.D., Advisor
Margaret J. Eppstein, Ph.D., Chairperson
Beverley Wemple, Ph.D.
Jamie Shanley, Ph.D.
Cynthia J. Forehand, Ph.D., Dean of the Graduate College

ABSTRACT

Despite decades of U.S. water quality management efforts, over half of assessed waterbody units were threatened or impaired for designated uses in the most recent assessments, with urban runoff being a leading contributor to those impairments. This cumulative research explores several aspects of urban runoff dynamics through a combination of field study and modeling.

Stormwater ponds are ubiquitous in developed landscapes due to their ability to provide multiple forms of treatment for stormwater runoff. However, evolving design goals have reduced the applicability of much of the early work that was done on pond effectiveness. In this study, we instrumented a recently constructed detention pond in Burlington, VT, USA. Flow gaging demonstrated that the pond achieved a 93% reduction in event peak flow rates over the monitoring period. Storm sampling showed that the pond significantly reduced total (TN) (1.45 mg/L median influent, 0.93 mg/L median effluent, $p < 0.001$) and total phosphorus (TP) (0.498 mg/L median influent, 0.106 mg/L median effluent, $p < 0.001$) concentrations over the events sampled. A loading analysis estimated the TN and TP removal efficiencies for the pond to be 23% and 77% respectively. Lastly, temperature data collected from the pond showed that during the summer the pond accumulates considerable heat energy. This study adds to the body of literature on detention pond performance, and raises concerns about the extensive use of stormwater ponds in watersheds where thermal stress is a concern.

EPA SWMM is a widely used urban hydrologic, hydraulic and water quality model, though its application can be limited due to its deterministic nature, high dimensional parameter space, and the resulting implications for modelling uncertainty. In this work, I applied a global sensitivity analysis (SA) and evolutionary strategies (ES) calibration to SWMM to produce model predictions that account for parameter uncertainty in a headwater tributary case study in South Burlington, VT, USA. Parameter sensitivity was found to differ based on model structure, and the ES approach was generally successful at calibrating selected parameters, although less so as the number of concurrently varying parameters increased. A watershed water quality analysis using the calibrated model suggested that for different events in the record, the stream channel was alternately a source and a sink for sediment and nutrients, based on the predicted washoff loads and the measured loads from the stream sampling stations. These results add to the previous work on SWMM SA, auto-calibration, and parameter uncertainty assessment.

Lastly, given the extent of eutrophication impairment in the U.S., I compared TN and TP data collected in these original works with national and regional datasets. TN concentrations sampled in this work were generally commensurate with values reported elsewhere, however TP data were not. Drainage area attributes and an event based rainfall runoff analysis of the study catchments provided circumstantial support for the idea that runoff from lawns is driving the high TP loads in Englesby Brook. The role of pet wastes is considered as a potentially fruitful area for further research.

ACKNOWLEDGEMENTS

I am grateful for the assistance, support, and guidance that I received from many individuals and entities over the duration of this work. First, I would like to acknowledge those individuals who provided direct assistance in this work. Peter Spartos and Khar M. Lau worked as field and lab assistants, ably handling aspects of bottle washing, equipment installation and maintenance, and sample collection sample transport in the 2007 and 2008 field seasons, respectively. Emma Kantrov completed an internship in the Bowden Lab in 2011 conducting spatial analysis in support of directly connected imperious area estimation in the Englesby Brook study watershed that facilitated my research.

Funding for this work was provided in part by the US EPA funded *Redesigning the American Neighborhood* project and by Vermont EPSCoR. Additionally, in kind support was provided by Vermont DEC on the Englesby project through analysis of collected water samples at the LaRosa Laboratory, and by Fitzgerald Environmental who surveyed pipe inverts in the Butler Farms study area.

Aspects of this work and of my professional development benefited immeasurably from the many discussions with my advisor, my committee, and other Bowden Lab and Rubenstein Ecosystem Science Lab researchers over the years. Lastly, I would like to acknowledge my family who were continually supportive of this endeavor over the years.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

1.1 Dissertation Structure

This dissertation is structured into five chapters. Chapter 1 includes a comprehensive literature review and background summary to place the subsequent chapters in context. Chapter 2 includes the details of a wet detention stormwater pond assessment conducted in the Englesby Brook watershed in Burlington, Vermont. This work will be reformatted and submitted to the *Journal of the American Water Resources Association* following the publication of this dissertation. Chapter 3 includes a cumulative documentation of the research work performed in Potash Brook Tributary Seven drainage in South Burlington, Vermont, including flow gaging and storm event sampling, a global sensitivity analysis and evolutionary calibration of a SWMM model for the area, and a drainage area pollutant washoff assessment. The specifics of how this work will be reorganized into discrete journal articles are not yet determined. Chapter 4 includes a comparative summary of national stormwater quality data and the total nutrient data from the studies summarized in Chapters 2 and 3, including the results of an additional event based rainfall runoff analysis. Chapter 4 will be distributed to interested Federal, State and Municipal water quality managers given its relevance to local management issues, and may serve as the basis for further research. Finally, Chapter 5 contains a cumulative summary of this work, including a focus on notable findings and suggestions for further work. A cumulative Bibliography and an appendix with additional data is included following Chapter 5.

1.2 Background Literature Review

My research focusses on measurement and modeling of stormwater runoff from small suburban watersheds. Thus, to begin, I review stormwater sources and the scope of the resulting management problems. Next, I review common stormwater management strategies including detention ponds, which are the focus of Chapter 2. This is followed by a review of stormwater modeling as a research and management endeavor, with a focus on EPA SWMM, which is used in Chapter 3 of this work. Next, I review simulation model sensitivity analysis, including previous applications to EPA SWMM. Lastly, I review the current state of model calibration, including manual and automated approaches, specifically in the context of previous work with SWMM.

1.2.1 *Water Quality*

1.2.1.1 National Water Quality

In the most recent compilation of nationwide water quality assessments, 54% of assessed stream and river miles, 68.4% of assessed lake, reservoir and pond area, and 78.2% of assessed bay and estuary area were found to be threatened or impaired (U.S. EPA 2013). The leading probable source groups include agriculture, atmospheric deposition, urban runoff / stormwater, municipal sewage discharges, and hydromodification, among other sources, reflective of the broad impacts of anthropogenic landscape change on our nation's water resources. Leading attributable pollutants causing impairment include pathogens, sediment, nutrients, metals and toxics, all of which can derive from multiple source groups, both non-point and point sources. This

diversity of source groups and impairing pollutants necessitates a range of management approaches in order to effectively target underlying impairments. In response, a range of federal, state and local regulations have been promulgated in recent decades to begin to address the range of water pollution sources, and yet over half of assessed water units remain threatened and impaired in the most recent assessments. This speaks to relative intractability of diffuse sources of pollution, which in many cases are inherent to human activities at the land surface and are not easily eliminated except at great cost. Stormwater is the focus of this research, which along with agriculture, atmospheric deposition, and unknown sources, is a leading cause of impairment of US waters, and one for which considerable research and management needs exists.

1.2.1.2 Stormwater

Stormwater and urban runoff have been a large contributor to documented water quality impairment and a focus of management efforts to reduce impairment inducing pollution. Federal regulations (40 CFR § 122.26 (b)(13)) define "storm water" as:

> "*Storm water* means storm water runoff, snow melt runoff, and surface runoff and drainage".

Following this definition, stormwater can exist in the absence of anthropogenic landscape change, however its characteristics (both quality and quantity) can be severely altered as a result of landscape change.

First and foremost, unmitigated landscape development can alter the timing and magnitude of streamflow and its contributing components through the construction of impervious surfaces, removal of natural vegetation, and installation of more efficient drainage (Leopold 1968; Booth and Jackson 1997). The construction of impervious surfaces produces one of the most dramatic changes, whereby areas that previously would produce surface runoff infrequently are replace by surfaces that do so routinely. The resulting reduction in infiltration can lead to a reduction in baseflow (Spinello and Simmons 1992), while the more quickly conveyed storm flow can result in flooding, channel erosion and habitat changes (Hammer 1972; Booth 1990; Paul and Meyer 2001). The loss of native vegetation and replacement with impervious surfaces and developed pervious surfaces (e.g., lawns) can also reduce annual evapotranspiration (ET) (Dow and DeWalle 2000), and can produce greater rainfall (and thus runoff) through heat island convection and urban production of cloud condensation nuclei (Huff and Changnon 1973). Cumulatively, these changes can make it such that not only does more water get to receiving waters more quickly by bypassing subsurface flow pathways, but that more water in total can enter receiving waters under a developed land scenario. The cumulative hydrologic effects of urbanization related landscape changes can alter habitat and degrade water quality in the absence of other factors.

In addition to hydrologic change, the change at the land surface can alter the sediment and solute loads that runoff conveys in ways that can adversely affect water quality. Driveways and other pavement surfaces, for example, may have some combination of degraded wear surface (e.g., asphalt, stone), anthropogenic detritus, and

aerially deposited material (Gilbert and Clausen 2006). Managed lawns, in contrast, will lack degrading wear surface material but may have some combination of fertilizers, pesticides, and pet wastes (Garn 2002; Nielson and Smith 2005). Thus, not only does incident precipitation fail to infiltrate, but the extra water that runs off can carry those materials as dissolved and suspended load. Developed watersheds typically have a diverse mixture of land cover characteristics all of which may contribute a range of pollutants at different rates, but all of which can be seen cumulatively by the receiving waters.

Perhaps the largest concerted effort to characterize the quality of collected stormwater was the National Urban Runoff Program (NURP), a large scale stormwater sampling initiative intended to establish a baseline for stormwater quality that could inform water quality management decision making (U.S. EPA 1983). The work was carried out at 28 sites across the U.S. over five years, with sampling of urban stormwater outfalls and data analysis managed by the U.S. Geological Survey (USGS), and in collaboration with state and local partners. NURP assessed the aggregate quality of runoff from the diverse range of urban sources contributing flow to the outfall points. They found a high degree of variability in urban runoff concentrations of pollutants assessed, with relatively high concentrations of metals (e.g., copper, lead, zinc) relative to ecological thresholds, as well as sediment, oxygen demanding substances, and coliform bacteria levels that could be problematic. In general, they were not able to attribute differences in pollutants levels to particular urban cover types within the urban landscape, although the data were still summarized by contributing urban land use and have been

used on that basis. Despite the comprehensiveness of the NURP data set, the data were collected more than three decades ago, and as such are not necessarily representative of contemporary runoff. For example, rates of atmospheric deposition, pesticide and fertilizer use, road surface wear characteristics, automotive deposition (e.g., brake dust, leaded gasoline, among other attributes), and instances of illicit sewer connection, and illegal waste disposal, are all likely to have changed over time, which can affect concentrations at stormwater outfalls. As a result, these data are of limited use for current management challenges.

More recently, Pitt et al. (2004) compiled stormwater outfall sample data from over 200 regulated Phase 1 Municipal Separate Storm Sewer Systems (MS4s) (those serving populations of 100,000 or more) into the National Stormwater Quality Database (NSQD). (By summarizing Phase 1 MS4s sampling data, the resulting data are highly reflective of stormwater from higher intensity development.) These data were generally consistent with NURP for nutrients and sediment concentrations, with the ranges of values recorded encompassing values that could exceed thresholds for sensitive receiving waters (VT-ANR 2014). Preliminary analysis of those data suggested that there were differences in analyte concentrations among the 11 identified contributing land covers present in the data (e.g., residential, mixed commercial, freeways), however additional work remained to confirm those differences given the confounding factors. Lead levels it should be noted, were found to have declined between the NURP and NSQD data sets. The time gap between the studies included the phasing out of lead in gasoline, which was reflected in declining concentrations of lead. Combined, the NURP and the later NSQD

datasets provide a baseline characterization of the quality of untreated stormwater that confirms that development alters not only the hydrology but also the water quality of run off.

Another receiving water attribute that can be affected by stormwater runoff is thermal regime, as a result of impervious surfaces ability to accumulate heat that can then be transferred to incident precipitation as it runs off (Galli 1990). Spronken-Smith and Oke (1998) used remote sensing and surface measurements to assess the thermal variation among different surfaces in urban parks, and the effects of urban park thermal regime on adjacent areas. They noted afternoon temperatures of 45-55 °C for roofs, 36-38 °C for at grade impervious surfaces, and 23 °C for tree canopies and shaded ground. While they did not assess runoff temperatures, these data clearly suggest the potential for heat transfer given a storm event. Thompson et al. (2008) measured surface temperatures and runoff temperature from experimental asphalt and sod plots and found that asphalt surface was on average 20.3 °C warmer than sod, and that the resulting runoff was initially 9.5 °C warmer from the asphalt than from the sod. Van Buren et al. (2000) performed a measurement and modeling study of parking lot asphalt heat transfer, and recorded differences between rainfall temperature (inferred by wet bulb temperature) and asphalt surface temperature. For example, for one summer storm the asphalt at the onset of rainfall was 17 °C hotter than the rainfall. They observed that given sufficient volume of rainfall, nearly all of the heat difference will transfer to runoff, with the effect on receiving waters concentrated in the early part of the storm when runoff temperatures would be greatest.

1.2.2 *Stormwater Management*

In light of the clear impact of stormwater on the quality of receiving waters and the cumulative physical understanding of how landscape development alters stormwater quantity, quality and timing, considerable work has been done in recent decades to mitigate the sources and effects. This includes both regulation that compels action, and research informing the actions to take and practices to employ to target most effectively the known and presumed sources.

1.2.2.1 Regulatory Context

The Federal Water Pollution Control Act (i.e., the Clean Water Act) was originally passed in 1972 to deal with point source water pollution, particularly sewage outfalls and industrial discharges of materials to waters of the U.S. (Ferrey 2004). Stormwater, as previously defined, only came to be federally regulated following the Water Quality Act of 1987 (i.e., amendments to the Clean Water Act), which resulted in promulgation of the Phase 1 Stormwater Rule in 1990. Under Phase 1, medium and large municipal separate storm sewer systems (MS4s), or those serving populations greater than 100,000, were compelled to implement local stormwater management plans, including sampling to inform both their own and broader nationwide efforts. Large construction sites and a subset of industrial operations were also included under the federal National Pollutant Discharge Elimination System (NPDES) framework for the first time under these rules. In 1999 EPA issued the Phase II Stormwater rules, regulating small MS4s, and lowering the regulated construction site threshold to 1 acre of

earth disturbance. EPA also issued the first Multi-Sector General Permit for a broad range of industrial related stormwater discharges in 1995. Cumulatively, these permitting obligations are variously delegated to State and local entities, and administer or overseen by EPA, varying by area.

There have been other broad based regulatory efforts to address stormwater, both preceding and following the Federal regulations just discussed. For example, the City of Bellevue Washington formed a Storm and Surface Water Utility in 1974 to manage for flooding and water quality concerns given contemporary and expected future development (WERF 2010). Similarly, Florida, Maryland and Vermont, among other states, were issuing stormwater discharge permits for large new developments in the 1970s before any Federal requirements to do so existed (NRC 2009). Concurrent with and following promulgation of Federal Rules, many States have continued to go beyond what is federally required for example requiring that stormwater management practices and site design techniques be applied to the maximum extent practicable in Maryland (Md. Code, Env. Art. §4-201.1 and §4-203 2007) and requiring discharge permits for all impervious new development or redevelopment exceeding one acre in Vermont (Vt. Env. Pro. Rules § 18-302 2011). However, in many other jurisdictions, Federal regulations continue to define the upper bound on stormwater management (National Research Council 2009).

Lastly, through the routine water quality assessment process that must occur under the Federal Water Pollution Control Act Section 303(d), waters identified as being impaired due to stormwater or stormwater pollutants (as well as a range of non-

stormwater impairing pollutants) may be subject to Total Maximum Daily Load (TMDL) planning or Water Quality Remediation Plan (WQRP) processes. A TMDL is an accounting system whereby targets are first set based on the pollutant loading that a receiving water can most likely assimilate while still supporting designated uses. Among the existing and expected future loading sources, a reduction plan is then devised to match the assimilative capacity of the receiving water and to thereby meet the water quality goal(s). TMDL planning and implementation can occur at a variety of scales, including for example a stormwater TMDL for Morehouse Brook, a ~1 km^2 watershed in Winooski Vermont (VT ANR 2007b), to the nutrient and sediment TMDL for the 166,000 km^2 Chesapeake Bay drainage area (U.S. EPA 2010). Alternatively, WQRP processes may be used in cases where it is expected that impairing loads can be eliminated quickly by working with a limited number of responsible parties, without the need for the broader framework that TMDL planning provides.

1.2.2.2 Centralized Treatment

Different treatment practices have specific advantages and constraints related to pollutant removal (Geotech Consultants and Wright Water Engineers 2012), hydrologic performance (McCuen and Moglen 1988), cost (Houle et al. 2013), space requirements, and other site conditions. Among stormwater management practices, most can be categorized as either centralized or distributed treatment practices, though this distinction can be somewhat arbitrary. Centralized treatment practices typically collect runoff from multiple land parcels, potentially including different contributing land uses. These

practices can include dry ponds, wet ponds, and various wetland treatment practices, and typically require considerable land area which must be located down slope of the impervious areas to be treated. The characteristics of a centralized treatment practice are dictated by the treatment goals, which can include prevention of overbank flooding, prevention of channel erosion or aggradation, removal of stormwater pollutants, and groundwater recharge.

Traditionally, preventing an increase in overbanking flooding via peak rate control has been the goal of centralized stormwater management (McCuen 1979). However, it was subsequently recognized that by limiting the focus to peak rate control, the larger runoff volumes produced by development can be released at near-peak flow rates for extended periods, which can produce sustained erosive velocities in channel (McCuen and Moglen 1988). Thus, throughout the 1980s and 1990s, different approaches were explored for adding volume control to stormwater ponds, such that the increased runoff volumes caused by development could be slowly released at rates that would not produce sustained erosive velocities in channel (MacRae 1993). Another design objective not central to early ponds, but that has gained in importance, is the removal of pollutants from stormwater. A key enhancement of traditional (i.e., 1970s) pond design is the establishment of a permanent pool of water within a stormwater pond (i.e., a wet pond), which can provide additional pollutant removal capability by allowing pollutants to settle out and to be removed through biological processes (i.e., algae incorporation, denitrification). These processes account for generally higher pollutant

removal rates that have been found in wet ponds, as compared with dry ponds (Fraley-McNeal 2007).

Wet stormwater ponds, which can combine overbank flooding, channel erosion and water quality treatment functionality within a single structure, have been widely constructed and widely studied in recent decades based on their capability to serve multiple stormwater objectives (Wu et al. 1996; Comings et al. 2000; Hossain et al. 2005; Weiss et al. 2007). A large number of earlier detention pond studies provide a basis for differentiating wet ponds from dry ponds, however there are a number of factors that limit the usefulness of previous individual and aggregated wet pond data. First, many studied ponds were built decades ago, using design practices that do not reflect current best practices (Barrett 2008; Wright Water Engineers 2012). Data from those installations cannot be expected to be representative of performance of newer detention ponds, which have incorporated various design enhancements over time (e.g., forebays, aquatic benches). Along with differences in detention pond designs, inconsistent data collection has obscured more conclusive knowledge on pond performance. While the goal of previous studies has consistently been to characterize runoff concentrations and/or loads into and out of detention ponds, the methods employed have varied widely.

For example, previously studied ponds have variously used time-paced or grab (as opposed to flow paced) sampling (Wu et al. 1996; Mallin et al. 2002), or estimated hydrologic flux components indirectly (e.g., by water balance as opposed to direct measurement) (Comings et al. 2000; Hossain et al. 2005). Differences in analytes that have been analyzed also limit the transferability of the data previously collected.

Sediment has been frequently sampled as total suspended sediment (TSS) as opposed to suspended sediment concentration (SSC). This is important because TSS has been shown to significantly underestimate actual sediment concentrations measured by SSC in certain circumstances (Gray et al. 2000). Nutrient sampling has included various subsets of organic and inorganic, dissolved and particulate, and total fractions of interest to the researchers.

Cumulatively, the history of stormwater pond performance assessments is both extensive and of limited use in predicting the performance of planned or newly constructed ponds due to changes in design standards, and differences in sampling regimes and performance assessment methodology (Strecker et al. 2001). Among the monitoring studies that have been conducted, as many as 70% of those studies were conducted prior to 2000, and thus may not be representative of the design features currently being used (Wright Water Engineers 2012). Given the increasing need for stormwater management in the U.S., additional data quantifying the various performance aspects of 'state of the art' wet detention ponds is potentially of value.

1.2.2.3 Distributed Stormwater Management

Distributed treatment practices (sometimes included under the umbrellas of Low Impact Design (LID), Green Stormwater Infrastructure (GSI), and Environmental Site Design (ESD)) are typically at the parcel or finer scale, and capture or convey runoff and associated pollutants on site. They can often be proportionally sized to the areas they are treating, which can increase the fraction of inflow volume that is lost via infiltration and

ET, as opposed to conveyed as effluent. These practices include small-scale bioretention, enhanced treatment conveyances, green roofs, and pervious surface installations among other practices. A key advantage to distributed treatment practices is that they can eliminate or reduce the need for large contiguous land areas to be dedicated to stormwater treatment. For MS4s in particular, the stormwater management regulations were imposed upon large existing developments that may have been largely constructed without consideration for stormwater treatment. Thus, there is a strong incentive to landscape integrate treatment where ever possible, given the challenges in finding large down-slope areas for centralized infrastructure and the potentially great cost for constructing new or newly separated (i.e., from wastewater) conveyance infrastructure.

In addition to offering a degree of spatial convenience, distributed treatment can be as effective at meeting hydrologic and pollutant management goals as traditional centralized infrastructure. A number of practices have been assessed for performance both individually (i.e., per constructed practices) and at the watershed scale. For individual practice types (e.g., bioretention), there are in many cases studies of individual installations, studies considering particular design variants, literature reviews of practices by type, and database compilations of studies for particular practices. Among the latter, the National Pollutant Removal Performance Database (Fraley-McNeal 2007) and the International BMP Database (Geotech Consultants and Wright Water Engineers 2012) contain the largest datasets for predicting how a particular treatment may perform based on past performance of similar practices that were monitored. These compiled results have generally supported the use of distributed treatment. One concern that has emerged

from the compiled research results is that for practices involving engineered soil media or fertilization, relatively high phosphorus effluent concentrations can occur, in some cases greater than influent levels. However, this issue can likely be mitigated in cases where phosphorus is of concern, either through judicious use of low phosphorus alternatives or by directing treatment practice effluent to the subsurface environment (i.e., infiltration) where phosphorus loadings are less problematic.

Other researchers have looked at watershed response to the implementation of distributed stormwater management in comparison with the response given a more traditional management approach (i.e., catch basin, pipes, ponds). For example, Dietz and Clausen (2008) monitored stormwater flow and nutrient loads for two drainage areas concurrent with the development of traditional (i.e., conveyance only, no treatment) and LID residential subdivisions, respectively. Little change was seen in the annual runoff volumes or nutrient loads at the LID subdivision relative the predevelopment forested conditions, while order of magnitude increases were manifested in the traditional subdivision. Wilson et al. (2015) instrumented two commercial subdivisions with traditional treatment (i.e., swale and dry pond) and LID stormwater management approaches and found little difference in resulting sediment and analyte effluent concentrations. However, the traditional site had an empirical runoff coefficient (i.e., runoff divided by rainfall) of 0.49 while the same coefficient was 0.02 for the LID development. While they do not comment on the cost differences for installing the two types of systems, they do document that an LID system can meet stringent runoff and load targets in a commercial setting. Lastly, ongoing research in New Hampshire has

demonstrated that installation of distributed treatment to reduce "effective" impervious area within a developed watershed can shift developed watershed area normalized discharge rates to closely match lower undeveloped watershed discharge rates, however with indirectly comparable watersheds (UNH Stormwater Center 2014).

In summary, properly designed and adequately sized centralized and distributed approaches and practices have been found capable of meeting stormwater management goals. Distributed treatment can be advantageous simply by occupying small underused green space thereby allowing for greater build out, and because distributed treatment can fit many places in the existing developed landscape that larger scale stormwater treatment cannot. Centralized treatment can have the perceived advantage of being out of sight where suitable siting is available, and can often make use of existing but less effective infrastructure (i.e., through retrofits). Ultimately, the most appropriate stormwater system depends upon many site specifics.

1.2.3 *Stormwater Modeling*

The magnitude of stormwater related water quality impairment and the need to investigate sources and evaluate potential management scenarios on both cost and effectiveness bases has led to the development of a suite of modeling tools which can be used in these applications. A selection of the most used U.S. government developed or sponsored tools for stormwater modeling are broadly reviewed in the following section. Section 1.2.3.2 then goes into greater detail on EPA's Stormwater Management Model (SWMM) since that is the tool that was ultimately used in Chapter 3 of this work.

1.2.3.1 Models

Stormwater models can focus on stormwater quantity or stormwater quality, but in practice, many have functionality to simulate both quantity and quality to varying degrees. Mathematical bases and software packages have been extensively reviewed by others (Zoppou 2001; Borah and Bera 2003; Elliott and Trowsdale 2007; Obropta and Kardos 2007) and will only be briefly reviewed here. Ultimately, the user's modeling objective will determine the necessary model functionality required for any particular use case. Factors such as whether or not explicit hydraulic design of stormwater treatment practices are needed, whether water quality accounting is required, and which aspects of the hydrologic cycle need to be included may dictate which modeling tools are most appropriate. In the remainder of this section I review two different models that have been widely used in stormwater applications. Then in the following section, I review the EPA's Stormwater Management Model (SWMM) in greater detail, which is the model I employed in this research.

One of the most common water quantity stormwater design tools is the Natural Resource Conservation Services (NRCS) Runoff Curve Number (CN) hydrology framework, formalized in both Technical Release 20 (TR-20) and Technical Release 55 (TR-55), as well as the U.S. Soils Conservation Service (SCS) National Engineering Handbook Hydrology Section (SCS 1965) and subsequent NRCS revisions. Both TR-55 and TR-20 are available in government distributed computer packages at no cost, and are also incorporated into numerous commercial applications with additional functionality. The basis of these approaches is to predict runoff volume as a function of empirical land

area CNs, which are determined by land cover / land use and a categorical soils parameter developed by NRCS (or rather, their predecessor agency the SCS). Procedures are included to predict both runoff volume and runoff peak rate given precipitation inputs and site specific parameters, and can be used to compute changes in runoff volume or peak flow rate under different scenarios, for example pre- and post-development. In addition to the runoff generation capabilities of TR-55, TR-20 includes a limited representation of reach routing and detention structures among other enhancements. No consideration for water quality is included.

Cited advantages of the CN approach are that it is simple to apply, it can produce reasonable results, and that it has been developed and supported by government agencies over a long period of time (Ponce and Hawkins 1996). For example, the Hydrologic Soils Group parameter (HSG) is mapped by NRCS and thus readily available for modeling applications across the U.S. However, despite its ease of use, the approach has been legitimately faulted for having been developed outside the scope of peer review (Hjelmfelt 1991; Willeke G.E. 1997), for being widely applied to cases for which it was not intended or appropriate (Fennessey et al. 2001), and for providing poor fit to measured data (Kumar and Jain 1982; Fennessey et al. 2001).

One of the broadest scale models used for stormwater applications is Hydrologic Simulation Program Fortran or HSPF (Bicknell et al. 2001). HSPF is an example of a process based, continuous lumped watershed model that can be used in mixed land use watersheds for the simulation of water quantity and water quality processes. Because it can account for urban land hydrology, land surface washoff, and loss and generation of

pollutants from stream channel processes, it can simulate the aggregate effects of urban development in either strictly urban or mixed land use watersheds. However, for strictly urban basins, the lack of detailed urban routing (i.e., sewer hydraulics) and detailed BMP capabilities is a hindrance (Borah and Bera 2003). It has also been noted that HSPF's reliance on a large number of parameters that are difficult or impossible to measure requires a high degree of specialized skill on the part of modelers in order to successfully use HSPF (Whittemore 2004). For example, Doherty and Johnston (2003) calibrated HSPF in four geophysically similar and adjacent watersheds, and found that for a subset of selected parameters there were different (non-unique) sets of values that all adequately fit the data within and across the four sites. While this issue is in no way unique to HSPF, a reliance on numerous parameters that are both abstract/immeasurable and sensitive is problematic. However, in large mixed land use watersheds where detailed routing of urban runoff is not required, HSPF has many advantages and has been used for example as the basis for Chesapeake Bay TMDL modeling (Shenk et al. 2012).

1.2.3.2 SWMM

The EPA Stormwater Management Model (SWMM) is the product of decades of hydraulic and hydrologic water quality model development and currently has routines for surface and subsurface flow, pipe and channel hydraulics, buildup and washoff water quality modeling, and stormwater infrastructure hydrology and hydraulics. SWMM has broad applicability for modeling developed landscape runoff and water quality dynamics and has been used for pre-post development hydrology assessment (Jang et al. 2007),

pollutant washoff water quality analyses (Tsihrintzis and Hamid 1998; Temprano et al. 2006; Lee et al. 2010), combined sewer overflow modeling and assessment (Cantone and Schmidt 2009; Zhang and Li 2015), flood forecasting (Liong et al. 1995; Kim et al. 2014), and stormwater treatment practice modeling and assessment (Aad et al. 2010; Lucas 2010; Burszta-Adamiak and Mrowiec 2013; Zhang et al. 2013), among other applications. Despite the broad applicability of SWMM's computational routines, its use can be limited in many applications by the relatively high parameter and input data needs, its urban focused conceptual model, and the high cost of measured data collection for SWMM calibration purposes. Because this is the model I ultimately decided to use for this research I will review the pros and cons of the SWMM model in greater detail than was provided for the previously discussed models.

The primary unit of spatial discretization within SWMM is the 'subcatchment', a user defined area based on topography. These can vary in size from entire watersheds to sub-parcel scale, depending on objectives. Each subcatchment must be transformed from its actual physical geometry to a rectangular plane representation within SWMM. It is then partitioned between pervious and impervious areas, and the user must select a routing method whereby (1) pervious runoff flows on to the impervious surface, (2) impervious runoff flows on to the pervious surface, or (3) both areas route directly to the subcatchment outlet.

For both the subcatchment impervious and pervious surfaces the user must specify conceptual depression storage, or depth of incident precipitation that can be held prior to the initiation of runoff, and slope and Manning's roughness coefficients

(FIGURE 1). Combined, these parameters specify the nonlinear reservoir conceptual model and the Manning's wide channel flow approximation runoff computation. For pervious areas, an infiltration model must specified through which water captured in depression storage can be lost to the subsurface module. Either an SCS CN based approach, the Green-Ampt equation, or one of two variants of Horton's infiltration model can be used for modeling infiltration in pervious areas. For each of the infiltration models, there are a number of physical and conceptual parameters which must be specified.

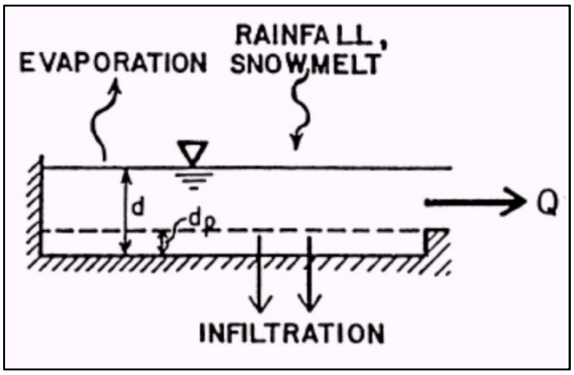

FIGURE 1. Schematic of SWMM surface runoff conceptual model. Parameters shown include depression storage depth (d_p) subcatchment water depth at time t (d), and outflow at time t (Q) as computed by Manning's equation (from Rossman 2010).

Defining the width of the conceptual rectangular cascading plane (*Width*) to which each subwatershed is converted is one of the most problematic parameterization issues for SWMM users. This is because 1) *Width* must be specified in every model, once per subcatchment, 2) *Width* has been previously reported to be a sensitive parameter (Barco et al. 2008; Krebs et al. 2013; Sun et al. 2014), and thus the value used is of

consequence, and 3) *Width* is a conceptual parameter, and is thus both inherently uncertain and unmeasurable. The SWMM User's Manual (Rossman 2010) provides guidance that an initial value can be computed as the drainage area divided by the average maximum overland flow distance for the drainage area, and that the value should be subsequently adjusted to improve hydrograph fit (i.e., calibrated). This is problematic both because observed data to calibrate against do not exist for many modeling applications, and because treating this sensitive parameter as a 'free' parameter to be calibrated allows it compensate for mis-specification in other parameters or model structure.

To address these issues with *Width*, Guo and Urbonas (2009) have used a dimensional analysis to produce a methodology for transforming irregularly shaped natural watersheds into rectangular cascading planes while preserving watershed area and potential energy over the drainage conveyance. This approach provides a quasi-physically based method for specifying the *Width* parameter that is based on measurable subcatchment properties, and thus does not rely on calibration data. To date, this approach has been applied on both hypothetical (Guo and Urbonas 2009) and real (Guo et al. 2012) watersheds, and is currently being incorporated into the SWMM Reference Manual (Rossman 2014), but has not yet been extensively tested or employed in the published SWMM literature.

For subcatchment water entering the subsurface module from pervious area infiltration, the flow can subsequently satisfy void space demand, evaporate, exfiltrate to deep groundwater (i.e., loss) or emerge as subsurface flow to the identified receiving

node. SWMM uses a number of physically based parameters to define the groundwater reservoirs, which are dynamically partitioned between an upper unsaturated zone and lower saturated zone. Specifically, water that infiltrates enters the unsaturated zone from which it can be lost via evaporation or percolated to the lower saturated zone using an implementation of Darcy's Law for unsaturated flow. Water that reaches the lower saturated zone is subject to deep percolation (i.e., lost from the model), evaporation, and outflow, resulting in the depth of the saturated lower zone being re-evaluated at each time step. Outflow is calculated by a user defined power function of the current depth of the lower zone, with the ability to include additional terms accounting for the depth of flow in the receiving channel relative to the depth of simulated groundwater. Each of these fluxes are controlled by one or more user-defined parameters, which are both physically and conceptually based.

For water that is simulated as flowing out of a subcatchment (as opposed to stored, lost to deep percolation, or lost to ET), that flow can ultimately be routed in a number of ways, including to stormwater treatment systems and through natural or engineered conveyances. Steady flow, kinematic wave, and dynamic wave routing options are included, allowing for physically realistic representation of pressurized flow in pipes, reversed flow, backwater effects, and conveyance system surcharging where needed. Standard hydraulic approaches are used to represent storage units, culverts, weirs, orifices, and other common features of engineered drainage networks, including specialized functionality for simulating flow through so-called LID features such as bioretention, permeable pavement and green roofs.

The water quality simulation capabilities of SWMM are based on buildup and washoff algorithms for user defined pollutants. Several functional relationships (e.g., power, exponential, and saturation) are available to specify the accumulation of pollutants between storms as a function of time, and the washoff of pollutants during storms as a function of flow rate. Efficiency of and temporal parameters for street sweeping can be specified, as well as static removal efficiencies for other simulated BMPs. Stream channel processes are not explicitly accounted for, however pollutants can be lost (i.e., deposited) at nodes within the conveyance network based on a user defined function. However, deposited pollutants are not subsequently available for transport, nor is channel erosion simulated at all.

1.2.4 *Sensitivity Analysis*

Sensitivity analysis (SA) includes a collection of methods for assessing the degree to which a model's predictions are affected by model parameters and /or inputs. For highly parameterized models, SA plays a critical role in focusing data collection, calibration, and uncertainty estimation efforts. For example, a parameter to which model outputs are not sensitive would be a poor candidate for costly data collection and verification, or for computationally expensive calibration and uncertainty extrapolation efforts. In contrast, a sensitive parameter may warrant any of those actions based on the modeler's objectives. The sensitivity of a model can be definitively assessed for a fixed model structure and parameter domain. However, for many environmental models, there is a range of feasible structural configurations and parameterizations for different

systems, or even for the same system. Thus, study-specific sensitivity analyses can be of critical importance in cases where previous analyses are not available, but can also provide insight for previously assessed models.

1.2.4.1 Local Methods

Approaches to SA have been generally categorized as being either local or global (Saltelli et al. 1999). Local methods individually assess model parameters relative to a base model, typically varying parameters one at a time by a fixed percentage or fixed factor of each parameters' assumed variance (Downing et al. 1985). Parameters are then ranked on sensitivity based on the relative change in model output resulting from individual changes in model input, with all changes being relative to the base scenario. The advantages to this approach include that it is conceptually simple, practical to implement for many models, computationally inexpensive, and can produce useful results within the limited context of the model parameterization being considered. The primary limitations of local methods are that only a small portion of the model parameter space is explored, and that interactions among parameters are not assessed.

1.2.4.2 Global Methods

Global SA is distinguished from local SA primarily by the concurrent variation of parameters. The resulting variation in model output(s) can then be attributed to individual model parameters averaged over the explored parameters space, rather than relative to the base case (Saltelli 2002). This has the advantage of providing insight when

parameters interact and of being able to concurrently assess sensitivity to the entire parameter space, given a fixed model structure, in a way that local SA does not. The primary disadvantage of global methods is that they can be more complicated and time consuming to implement and interpret.

While there are numerous approaches to global sensitivity analysis that have been applied to hydrologic and water quality modeling, these have generally been based on either Monte Carlo filtering or variance decomposition. The first approach to Monte Carlo filtering was presented by Spear and Hornberger (1980) and is referred to as Generalized Sensitivity Analysis (GSA). In this approach, a uniform random sampling of the parameter space is performed resulting in large number of model simulations. Each simulation is next characterized as being either behavioral (i.e., an acceptable simulator of the underlying system) or non-behavioral (i.e., an unacceptable simulator), based on an application-specific performance criteria threshold. The differences (if any) between a model's parameter values in the behavioral and non-behavioral model runs are then taken as indicating model sensitivity to a given parameter. Subsequently, Freer et al. (1996) expanded the GSA technique to include many categories of model fitness, as opposed to the binary behavioral / non-behavioral classification scheme used in GSA. Advantages of this approach, termed Regional Sensitivity Analysis (RSA), are that it can display finer scale parameter sensitivity and it is not dependent on a single performance threshold, as GSA is. This methodology has been employed for analysis of conceptual rainfall runoff modeling (Wagener et al. 2001; Wagener and Kollat 2007), distributed catchment modeling (Sieber and Uhlenbrook 2005), channel flood modeling (Roux and Dartus

2008), and hyporheic exchange modeling (Naranjo et al. 2012), among other applications.

The second set of approaches are based on variance decomposition, whereby the parameter space is randomly sampled and the resulting variation in model output is partitioned among the input parameters and input parameter interactions contributing to the output variance. Among hydrologic and water quality modeling applications, variations on Sobol' indices (Sobol' 2001) are among the most frequently employed variance decomposition methods (Song et al. 2015). To compute Sobol' indices, model output variation is first generated by sampling the model parameter space. First order sensitivity indices can then be calculated as the ratio of the output variance contributed by an individual model parameter to the total model output variance. Total sensitivity indices, accounting the individual and interacting contributions of parameters to total variance can be computed as well. Explicit calculation of parameter interaction terms has been cited as an advantage of Sobol's method over RSA for example, which only deals with parameter interactions indirectly (Tang et al. 2007). This feature has led to increasing use of Sobol's method in recent years, including applications in watershed modeling (van Werkhoven et al. 2008; van Werkhoven et al. 2009), semi-arid flash flood forecasting (Yatheendradas et al. 2008), and forested basin scale runoff modeling (Zhang et al. 2013).

1.2.4.3 SWMM Applications

There have been numerous published SAs of SWMM, including both local and global approaches. For example, Barco et al. (2008) individually varied seven surface runoff parameters and found that simulated runoff was most sensitive to percent imperviousness and depression storage depths. However they used parameter perturbations of +/- 90% that in many cases would be greater than the uncertainty in underlying parameters. Peterson and Wicks (2006) included a local sensitivity analysis in a study that used SWMM to simulate flow and solute transport is karst systems and found conduit Manning's n to be most sensitive out of the subset of parameters assessed. As a final example of local SA applied to SWMM, Krebs et al. (2013) conducted a one-at-a-time SA of subcatchment surface parameters and a Green-Ampt infiltration parameter to inform a subsequent multi-objective genetic algorithm calibration of SWMM. Their SA was constrained by the presumed uncertainty in their particular parameterization. They found impervious depression storage and impervious Manning's coefficient to most strongly affect modeled runoff dynamics.

Several global SAs have also been conducted on various SWMM components, using the Monte Carlo scatter plot approach demonstrated by Duan et al. (1992) and Beven (1993), the GSA or RSA procedures of Spear and Hornberger (1980) and Freer et al. (1996), respectively, and variance decomposition in at least one case. For example, Aronica et al. (2005) ran 10,000 simulations of a twelve hectare drainage area model, sampling a complete random surface parameter set for each run and subsequently computed a hydrologic performance metric for binary classification of simulations

between behavioral (i.e., acceptable) and non-behavioral (unacceptable) results. Scatter plots of parameter values versus model performance revealed the extent to which variation in model performance was correlated with individual parameter values and found model performance to be most affected by the conduit Manning's coefficient, impervious depression storage and the surface Manning's coefficients. In contrast, the Horton infiltration parameters and pervious area runoff parameters were not sensitive despite the 32% pervious subcatchment, likely due to the use of a peak flow oriented objective function (i.e., Nash-Sutcliffe efficiency or NSE).

Zhang and Li (2015) applied the same Monte Carlo scatter plot approach to a 130 hectare combined sewer drainage area in Shanghai. Of the five surface hydrology parameters varied, percent impervious and impervious Manning's roughness were found to most affect the NSE performance metric. In another study, Gaume et al. (1998) applied the RSA procedure of Freer et al. (1996) to an application of SWMM's exponential buildup and washoff algorithms for modeling suspended solids in a combined sewer system. The buildup and washoff models were generally sensitive to the input parameters with the exception of the buildup exponent, where acceptable model performance was found to be distributed across most of the parameter's range. Lastly, Sun et al. (2014) calculated variance decomposition based first order sensitivity indices for selected surface runoff parameters following a global Monte Carlo sampling of the model application's parameter space, using a modification of SWMM that explicitly accounted for 'Trees' as a subcatchment area that is distinct from 'Pervious'. They demonstrated differences in the sensitivity of simulated flow volume versus simulated

peak flow rate, with most of the assessed parameters showing a degree of sensitivity in one or both cases.

Cumulatively, these works provide a baseline sensitivity reference for the components that have been assessed to date (particularly the surface hydrology component). However, they also demonstrate how sensitivity can vary based on the SWMM model structure, the methods used (i.e., local vs global), and by the degree of parameter variation employed, consistent with the application to application variation in sensitivity seen in other models (e.g., van Werkhoven et al. 2008). Therefore, a user needs to assess previous SAs to determine whether they adequately address their structure and parameterization under consideration. For some aspects of SWMM simulations, including pervious area characteristics, infiltration models, and subsurface flow parameters, the parameters and structures have been minimally assessed in the past, if at all. Thus, if those model components are thought to be important in representing a system, project specific SA may be warranted.

1.2.5 *Calibration*

Calibration is the process of changing model parameters within predefined constraints to improve agreement between model output and observed data (Moriasi et al. 2007). Where only a limited number of parameters are to be calibrated, manual methods can be employed by iteratively adjusting parameters individually or in concert while monitoring for changes or improvement in model performance. However, for even moderately complex or parameterized models, the process of manual parameter

adjustment and assessment can become impractical (i.e., the "curse of dimensionality"), which has spurred the development of automated approaches.

1.2.5.1 Manual versus Automatic

The role of modeling in the investigation and management of watersheds has led to considerable research on model calibration, both endogenously by watershed modelers and by drawing on calibration work from the modeling of other systems. The reasons for calibrating a model include adjusting for watershed conditions not believed to be accurately captured by the model's base parameterization (e.g., fragipan soils via CN modeling, see Peterson and Hamlett 1998; Benaman et al. 2005), or where uncertain empirical parameters are present, such that any value within a predefined range can be used that will maximize model performance (e.g., SWMM conceptual / empirical parameters, see Tsihrintzis and Hamid 1998; Temprano et al. 2006). A general distinction can be drawn among various calibration approaches by classifying them into either manual calibration or automatic calibration approaches. In manual calibration, a modeler will systematically change one or more parameters to produce a desired change in the initial model output, typically in a laborious fashion. Manual calibration can be effective if approached systematically (e.g., stratified sampling approaches, McKay et al. 1979) and when the number of parameters to be calibrated is limited, however it quickly becomes impractical as the number of parameters being calibrated increases.

It has long been recognized that this process of manual calibration ranged from tedious to intractable, spurring interested in automated approaches (Ibbitt and O'Donnell

1971). Early approaches to automated calibration of watershed models (generally conceptual rainfall runoff models at that time) were for the most part derivative-free (but see Gupta and Sorooshian (1985) for a derivative based approach) direct and pattern search algorithms (Ibbitt and O'Donnell 1971; Pickup 1977). However, it was generally found that these approaches were unable to find global optima and as such were not ideal for rainfall runoff models given the known presence of suboptimal minima in the typical objective function response surface (Gupta and Sorooshian 1985; Hendrickson et al. 1988).

The next generation of approaches were evolutionary in approach, meaning that the algorithm properties were based on principles of biological evolution (i.e., reproduction, survival of the fittest). For example, Wang (1991) applied a canonical genetic algorithm, following the work of Holland (1975), to calibrate a conceptual rainfall runoff model and reported that it identified the global optimum in 8 of 10 runs, with near optimal results in the two runs that did not identify the global minimum. Duan et al. (1992) combined aspects of evolution search with existing direct search approaches to formulate the Shuffled Complex Evolution-University of Arizona (SCE-UA) algorithm, which was demonstrated to outperform several non-evolutionary predecessor algorithms. SCE-UA consists of multiple concurrent direct search simplexes (following the work of Nelder and Mead 1965) which periodically exchange information (i.e., shuffle), resulting in relatively fast and consistent convergence toward the global optima. In introducing SCE-US, Duan et al. (1992) did not directly compare SCE-UA to a canonical GA, however work by others (Jeon et al. 2014) has suggested that both have

advantages and disadvantages that are dependent on the particular application. It should be noted that the need for user configuration of the algorithms (e.g., population size, number of complexes, mutation rates, etc.) complicates any direct comparison of these methods. Since the publication of SCE-UA and the GA application of Wang (1991), there have been hundreds of applications using these approaches for single objective calibration of watershed models.

Evolution strategies (ES) are a particular variant of biological evolution-inspired algorithms (i.e., evolutionary algorithms), which were initially distinct from the more commonly used GAs (Eiben 2003). ES was initially devised for real valued parameters, and is therefore not directly applicable in some circumstances, for example when variables must be represented in binary or as permutations (Bäck and Schwefel 1993). There are important distinctions between ES and GAs, which were developed concurrently over decades. For example, ES relies more heavily on mutation or random perturbations of parameter values (as opposed to recombination or shuffling of discrete values between solutions) for the introduction of variation into parameter sets, and includes properties of self-adaptation. That is, in addition to evolving solutions to the problem being optimized (e.g., hydrologic model parameters that improve simulation performance), an ES can also evolve its own parameters to increase the efficiency of the search over generations. This typically includes the variation operator or mutation rate, which controls the level of variation allowed in the evolving parameters and in particular how large a change can be made in the search for better performance (i.e., fitness function / objective function minimization).

Of the published ES applications to environmental model optimization, the majority have used variants of ES, reflecting the evolving nature of the field. For example, Ostermeier et al. (1994) introduced de-randomized mutative step-size control ES (DES), which replaces a random aspect of object parameter mutation with a deterministic approach. The approach purports to solve the problem of inefficiency of random draws from the mutation distribution, whereby relatively large mutations can be drawn (with low probability) from mutation distributions that have evolved to have small standard deviations. That large mutations can occur where small mutation are known to be most profitable has been demonstrated to introduce inefficiency into searches for a number of theoretical test problems (Ostermeier et al. 1994). The DES approach has been successfully used in at least two studies for the optimization of groundwater remediation scenarios (Yoon and Shoemaker 1999; Bayer and Finkel 2004), where DES was found to outperform other optimization approaches (e.g., simple GAs and direct search).

Another commonly used variant is the so-called covariance matrix adaptation ES (CMA-ES) first introduced by Hansen and Ostermeier (1996). Similar to DES, this approach seeks to de-randomize the mutation of strategy parameters, but further makes use of the strategy parameters' performance over the full generational time to optimally adjust the strategy parameters at the current generation. Thus, this approach both eliminates the possibility of randomly drawn large changes where small changes are profitable and can draw on the series of preceding mutations to specify deterministically the strategy parameters at each generation. While this can produce vastly greater search

efficiency in many theoretical scenarios, it comes at the expense of greater algorithm complexity and parameterization with the ES itself. Further, Beyer and Arnold (2003) demonstrate that these changes to mutation approach are not always beneficial (i.e., can lead to premature convergence) and note that the cumulative theoretical results on ES variants (including their own) do not necessarily transfer to "real world optimization problems". Nonetheless, applications of CMA-ES have been found to perform well in two applied water resources applications (Bayer and Finkel 2007; Maier et al. 2009), suggesting the efficiency gains demonstrated in the theoretical cases do not cause the algorithm(s) to fail in real world applications. Still, performance comparisons of different ES variants across real world optimization problems appear to be lacking.

Other recent work has focused on the inherently multi-objective nature of watershed model calibration, whereby a user may need to calibrate or optimize for uncorrelated measures of model performance concurrently, for example water balance and peak flow accuracy (Gupta et al. 1998). The goal then is to identify a Pareto optimal or non-dominated set of solutions, which are those among the feasible space which cannot be improved with respect to one objective without loss of goodness of fit in other included objectives. For example, Yapo et al. (1998) developed multi-objective complex evolution algorithm (MOCOM-UA), which uses downhill simplex evolution of sets of solutions based on the Pareto ranking scheme following Goldberg (1989). Continued developments in multi-objective optimization algorithms that can be applied to water resources modeling have included sharing of information among complexes (see for example the widely used Non-dominated Sorted Genetic Algorithm-II or NSGA-II of

Deb et al. 2002), incorporation of self-adaptive features (e.g., Kollat and Reed 2006; Vrugt and Robinson 2007), and a number of other approaches recently reviewed by Reed et al. (2013).

1.2.5.2 Generalized Likelihood Uncertainty Analysis (GLUE)

An inherent aspect of traditional optimization algorithms applied to watershed models has been the goal to identify a single best or best sets of parameterizations. A related but distinct school of thought, with respect to model calibration, follows from the so-called equifinality thesis of Beven (1993). The equifinality thesis proposes that, given measurement error, limits in hydrologic understanding that propagate into model structural error, uncertainty in model parameters and inputs, and the relatively low information content of a streamflow record for identifying a range of distributed physical parameters, efforts to identify optimal parameter sets are misplaced. Thus, many different parameterizations and even model structures may be acceptable simulators of the system. It follows that the calibration objective should be to identify a collection of acceptable simulators of the system as opposed to a single best simulator, since they may all have value as predictors of future states of the system.

Beven and Binley (1992) described a procedure for dealing with multiple acceptable models, the so-called Generalized Likelihood Uncertainty Estimation (GLUE) procedure, which has been both widely adopted and widely debated since its initial introduction (Beven and Binley 2014). Under GLUE, the parameter space is first sampled by Monte Carlo sampling a large number of times and each set of parameters is

then given a likelihood weight of either 0 (unacceptable simulator) or a positive value, increasing proportional to the quality of the simulation. As additional data become available (e.g., subsequent years of hydrologic data), the likelihood weights can be recalculated to further restrict the parameter space. Thus, GLUE can produce a large range of model predictions which in some cases may not encompass observations for all or part of the record. Beven (1993) observes that the GLUE procedure is analogous to classical optimization, except that in classical optimization the best simulation is given a likelihood value of 1, while all others are assigned a value of zero. A key feature, however, is the rejection of classical optimization framework of seeking a single global optimal solution.

Despite its widespread adoption by environmental modelers, the GLUE framework has been the subject of considerable debate and criticism focused primarily on the efficiency of sampling and its statistical validity (Mantovan and Todini 2006; Stedinger et al. 2008). Reliance on uniform random sampling in most applications has yielded low search efficiency, such as the 37,000 acceptable parameter sets found from 2.7 million sets evaluated in the work of Brazier et al. (2000). Thus, in cases where a large number of parameters are to be included in a GLUE analysis, adaptive search approaches have been incorporated to improve the efficiency of behavioral parameter set identification (Blasone et al. 2008).

The criticisms of GLUE's statistical validity stems from how the range of model predictions generated by GLUE can and cannot be interpreted. Under a restrictive set of circumstances in which errors in inputs and model structure are known, a formal

likelihood function can be used in GLUE resulting in model predictions with statistically interpretable properties (e.g., model predictions will bracket future observations with X% probability, given the cumulative uncertainty in inputs, model structure and model parameters). However, as discussed by Beven et al. (2008), in a typical hydrologic modeling application there will be unknown input errors (e.g., in rainfall data) that are processed nonlinearly through a model structural containing unknown errors, precluding definition of a formal likelihood function. Thus, a typical GLUE application will produce prediction intervals accounting for the subjective narrowing of parameter uncertainty via behavioral model classification, but otherwise lacking in statistical interpretability. Formal likelihood measures have been incorporated into GLUE and applied to rainfall runoff models in greatly simplified or hypothetical examples (Mantovan and Todini 2006; Stedinger et al. 2008). However, methods for extension of these approaches to more complicated watershed models with unknown errors have not yet been established.

1.2.5.3 Previous SWMM Work

A number of automated calibration approaches have been applied to SWMM in the past, towards identification of best parameter sets from a large feasible parameter space. For example, Barco et al. (2008) used the complex method direct search algorithm (Box 1965) to optimize four SWMM parameters, while keeping all others fixed at best estimates. Other researchers have employed evolutionary approaches, including Liong et al. (1995), who reported successful use of a simple GA to optimize subcatchment surface

parameters and Horton infiltration model parameters for the minimization of peak flow rate errors. Similarly, Balascio et al. (1998) used a GA for a set of surface runoff and Horton infiltration model parameters, and found the algorithm to be successful at minimizing a multi-objective hydrologic error term. Fang and Ball (2007) also explored the use of a GA for minimizing the error in simulated flow in a 133 hectare drainage area in Australia. However, rather than simply search for a best parameter set, they used the GA to narrow the feasible parameter space, in recognition that identifying a single best parameterization was conceptually problematic.

More recently, Krebs et al. (2013) used the NSGA-II algorithm (Deb et al. 2002) to optimize surface depression and conduit roughness parameters while keeping all other parameters fixed at best estimates. Their approach repeatedly evolved the included parameters to the same values (suggesting a global optimum), however most parameters were repeatedly evolving to a boundary of the allowed parameter range. Thus, while the NSGA-II algorithm was successful at identifying an optimum within the allowable parameters space, the results strongly suggested that out of range parameters could further optimize performance, indicating that either parameters or model structure were misspecified for the application.

The GLUE procedure for model parameter uncertainty estimation has been applied to many environmental models since its conception (Beven and Binley 2014), including at least two SWMM applications. For example, Sun et al. (2014) applied GLUE to a modified version of SWMM that accounted separately for trees and lawns (as opposed to default SWMM approach of lumping all 'pervious'). Using a combination of

volume and rate based flow performance measures they demonstrated how different performance measures can be used to constrain the acceptable regions of the parameter space, and used GLUE to propagate the confined parameter space through their model predictions. Zhang and Li (2015) applied the GLUE methodology to eight surface and infiltration parameters in a combined sewer drainage area and assessed those parameters on their ability to simulate water levels at a pump station suction well, while also exploring both different objective functions and the subjective acceptance thresholds used for accepting candidate models. They reported the GLUE procedure to be capable of restricting the predictive uncertainty in water levels such that they very closely corresponded with observed water levels, albeit with a large volume of in-line storage initialized into the model as opposed to produced by SWMM.

In summary, despite the solid basis of SWMM calibration and parameter uncertainty analysis work summarized in the preceding sections, parameter uncertainty estimation efforts to date have generally focused on the surface runoff model or selected components of the surface runoff model. Further, calibration efforts have seldom been informed by a robust application specific SA to focus the work. Given the high dimensionality of the SWMM parameter space, these efforts may significantly underestimate the variability in SWMM predications attributable to parameter uncertainty on account of the large number of potentially sensitive parameters that remain at fixed values. (This is neglecting model structural uncertainty, uncertainty in inputs/forcings, and uncertainty in calibration data, which are more difficult to assess and heretofore unexplored in the context of SWMM modeling.) In the following work, I more fully

assess parameter uncertainty in a SWMM model for the primary purpose of aiding our hydrologic and pollutant transport assessment, but secondarily for the benefit of other SWMM users.

CHAPTER 2. NUTRIENT PERFORMANCE, HYDROLOGIC PERFORMANCE, AND TEMPERATURE DYNAMICS OF A MUNICIPAL RETROFIT WET EXTENDED DETENTION POND IN BURLINGTON, VT

2.1 Abstract

We instrumented and studied a wet extended detention pond constructed as a retrofit installation in Burlington, Vermont. Flow gaging at the pond inlet and outlet demonstrated that the pond achieved a median reduction in peak flow of 93% over 89 discrete events identified in the flow record between 2007 and 2009. Storm event samples collected at the inlet and outlet via auto-sampler showed that the pond significantly reduced total nitrogen (TN) (1.45 mg/L median influent, 0.93 mg/L median effluent) and total phosphorus (TP) (0.498 mg/L median influent, 0.106 mg/L median effluent) concentrations over the events sampled. Using the event mean concentrations and a set of dry weather (baseflow) concentrations, I conducted a Monte Carlo loading analysis that estimated the TN and TP removal efficiencies for the pond as 22.5% and 76.7% respectively. These removal efficiencies, which were measured at a pond located in the humid-continental (cool summer) climate zone, are consistent with reported values for similar facilities located elsewhere. Temperature data collected at the study pond's inlet and outlet suggests that during warmer months, stormwater retained in the pond warms considerably. This suggests that summer warming in large-scale detention ponds could be an issue in areas where receiving waters are already thermally stressed.

2.2 Introduction

Urban runoff is a widely recognized source of water pollution. EPA's most recent cumulative assessment of state data found 53% of assessed river and stream miles to be impaired or threatened, with urban and municipal discharges as the second leading probable cause of impairment, behind agriculture (U.S. EPA 2013). The pollutants most frequently identified as contributing to U.S. river and stream impairment include those that tend to be associated with urban runoff and stormwater (e.g., pathogens, nutrients, metals, sediment) (U.S. EPA 2013). The management of urban runoff and associated pollutants is addressed under the Federal National Pollution Discharge Elimination System (NPDES) and Total Maximum Daily Load (TMDL) frameworks, as well as various state and local programs, often with the goal of restoring runoff quality and quantity back toward predevelopment levels, or to within the assimilative capacity of the receiving water (Ferrey 2004).

Unmitigated stormwater discharges to receiving waters can result in adverse hydrological, biological, geomorphic, chemical, and thermal change (Paul and Meyer 2001). Development alters the timing and magnitude of flow through the construction of impervious surfaces, removal of natural vegetation and installation of more efficient drainage (Poff et al. 1997). Changes in flow can in turn deliver higher loads of solutes and sediment bound pollutants (Pitt et al. 2004), and can alter the thermal characteristics of runoff by making it hotter (Galli 1990; Herb et al. 2008), which can further degrade water quality and habitat (Walsh et al. 2005). Changes in hydrology associated with urbanization can also cause stream channels to erode (Wolman 1967), creating a

continual supply of sediment within the stream channel that is delivered downstream (Trimble 1997), including the potential for mobilization of toxic legacy sediments (Singer et al. 2013). The cumulative effect of these changes is often that streams become degraded, with a loss of valued taxa and a resulting shift to more pollutant-tolerant aquatic community (Cook 1976; Pratt and Coler 1976).

The negative effects of stormwater and associated pollutants can be addressed through various structural treatment practices (e.g., detention and infiltration) and through source reduction (e.g., restrictions on fertilizer use, street sweeping) management practices. Among infrastructure practices, most can be further categorized as either centralized or distributed treatment practices. Centralized treatment practices typically collect runoff from multiple land parcels, which may have different land uses and covers. These practices can include dry ponds, wet ponds, and various wetland treatment practices. Distributed treatment practices are typically at the parcel or finer scale and detain runoff and associated pollutants on-site. These include small-scale bioretention, pervious surface installations, treatment conveyances, and other landscape integrated practices. Different treatment practices have specific advantages and constraints related to pollutant removal (Geotech Consultants and Wright Water Engineers 2012), hydrologic performance (McCuen and Moglen 1988; Wilson et al. 2015), cost (Houle et al. 2013), required space, and other site conditions. Consideration of these factors will determine the best or most feasible treatment practice for a particular set of circumstances.

Stormwater ponds have been a widely used method to treat urban runoff because they reduce peak flows and treat multiple parcels at a centralized location. Due to their wide use, there have been numerous opportunities to monitor the efficiency of detention ponds (Wu et al. 1996; Comings et al. 2000; Hossain et al. 2005; Weiss et al. 2007) and so they are readily accepted by regulators. However, despite the history of stormwater pond performance assessments, there is still considerable uncertainty about how any particular pond will perform in the field over time, due in part to differing pond inputs, changes in design standards, and differences in sampling regimes and performance assessment methodology (Strecker et al. 2001). For instance, many older ponds did not include sediment forebays, elongated geometry, aquatic benches, or substantial permanent pool volumes, all of which can facilitate pollutant removal from transport. Further, among the monitoring studies that have been conducted, as many as 70% of those studies were conducted prior to 2000 and thus may not be representative of the design features currently being used (Wright Water Engineers 2012).

The International BMP Database, a central repository for stormwater BMP studies, identifies only one wet stormwater pond that was studied within New England (Wright Water Engineers 2012). That pond was a retention pond[1] located at the University of New Hampshire (UNH) Stormwater Center, in Durham, NH (UNH

[1] Stormwater pond naming conventions are not consistent across the U.S. Ponds can generally be categorized as either wet ponds or dry ponds. Wet ponds maintain a permanent pool of water between storm events by design, whereas dry ponds are designed to store runoff volumes only temporarily and to empty between events. Vermont's Stormwater Management Manual calls wet ponds (i.e. those with permanent pools) detention ponds, while ponds designed to drain fully are referred to as dry ponds. Convention within the International BMP Database and elsewhere is to refer to ponds with permanent pools as 'retention ponds', and to call fully draining dry ponds 'detention ponds'. With the exception of this single reference to the University of New Hampshire Retention Pond, naming convention throughout this work will use the Vermont Stormwater Manual terminology.

Stormwater Center 2009). The UNH pond is atypical in two ways. First, the pond drains a land area of 0.4 hectares (i.e., one acre), which is a relatively small drainage area for a wet pond. The Vermont Stormwater Management Manual suggests a minimum of 4 hectares of drainage area be routed to a wet pond to ensure preservation of a permanent pool (VT-ANR 2002). New Hampshire in contrast requires a documented hydrologic budget demonstrating that the permanent pool will be maintained (Comprehensive Environmental and NH-DES 2008). Under either of those regulatory frameworks, a 0.4 hectare drainage area is likely at the low end of the range of drainage areas for wet detention due to regulatory constraints and the availability of alternative treatment practices to treat small areas. Second, the pond was constructed with steep side slopes in clay soil which led to erosion within the pond and thus endogenous sediment and nutrient supplies by the second year of operation. Given that these design limitations can usually be avoided, the resulting low pollutant removal rates calculated at the UNH pond should not be viewed as broadly representative of modern design ponds. Nonetheless, the UNH retention pond appears to singularly comprise the New England regional wet pond performance data set. Thus, the goal of this study was to evaluate the nutrient performance, hydrologic performance, and temperature dynamics of a recently constructed large wet extended detention pond in Burlington, VT.

2.3 Methods

2.3.1 *Site Description*

The stormwater pond that we studied is located in the Englesby Brook watershed in Burlington Vermont. Englesby Brook is one of the most highly developed watersheds in Vermont with moderate density residential neighborhoods, commercial and light industrial uses, and institutional and municipal land covers. This 2.45 km^2 watershed drains to Lake Champlain, directly adjacent to a public swimming area (TABLE 1). The Vermont Department of Environmental Conservation has determined that both Englesby Brook and Lake Champlain are impaired by stormwater and phosphorus, respectively, and that flow from Englesby Brook has contributed to bacterial swimming closures in the past. Englesby Brook has been listed on the U.S. Environmental Protection Agency's 303(d) listed since 1992, alternately due to bacteria and other contaminants, and multiple impacts associated with excess stormwater runoff. Restoration work within Englesby Brook has been underway since at least 2000 (Center for Watershed Protection 2000; Medalie 2012), including most recently through the City of Burlington's Municipal Separate Storm Sewer System (MS4) Phase II stormwater management plan and through a flow-based stormwater TMDL approved in 2007 (VT ANR 2007).

TABLE 1. Land cover attributes for the Englesby Brook Watershed, and for the study pond's drainage area. Land cover and canopy cover estimates are calculated from 2011 National Land Cover Dataset products. Additional detention pond drainage area impervious estimates were hand digitized from 2004 color orthophotos.

Land Cover Attribute	Englesby Brook	Detention Pond Drainage Area
Area (km^2)	2.45	0.47
Impervious Cover		
Total NLCD 2011 (%)	28.1	44.2
Total Hand Digitized 2004 (%)	---	39.0
Directly Connected (%)	---	32.7
Land Cover		
Developed (%)	87.6	89.5
Forest (%)	11.9	10.4
Pasture/hay (%)	0.4	0.2
Canopy Cover (%)	31.5	29.6

The Englesby Brook TMDL was one of twelve flow-based stormwater TMDLs issued in Vermont between 2006-2009. The framework applied in each of these watersheds was to first match each stormwater impaired watershed with one or more non-impaired watersheds with similar development patterns and geophysical attributes that were meeting state-defined biocriteria standards (so called "attainment" watersheds, see Foley and Bowden 2005). Ten years of P8 modeling (Walker 1990) were then run for each attainment and stormwater impaired watershed, and synthetic flow duration curves for each watershed were tabulated from the P8 output (Tetra Tech 2005). The TMDL targets for Englesby Brook and the other stormwater impaired watersheds were then calculated as the average Q 0.3% flow rate among the statistically matched attainment watershed subset for each particular impaired watershed, using mean daily flow rates.

An additional reduction to account for future growth was included, resulting in a target reduction of 34.4% in the Q 0.3% for Englesby Brook, relative to the baseline modeling (VT ANR 2007).

There were several stormwater management activities implemented in the watershed from 2001 to present, both structural and non-structural, toward watershed restoration goals. These included the construction and retrofitting of multiple detention ponds, construction of a shallow marsh wetland, and numerous measures under the MS4 management including street sweeping, catch basin cleaning, and illicit discharge detection and elimination. The largest structural control measure, on a treatment area basis, is an extended wet detention pond (hereafter "the pond") constructed near the middle of the watershed in 2005. This pond's drainage area is 48.7 hectares, or 19.9% of the total Englesby Brook watershed area. The land cover attributes within the pond's drainage area are similar to the rest of the watershed, though slightly more developed (TABLE 1). The pond's drainage area is more impervious than the watershed as a whole, and the impervious surfaces are largely connected to the stormwater drainage infrastructure (32.2% directly connected impervious cover (DCIA), though a DCIA estimate is not available for the entire watershed). Water that infiltrates pervious areas within the drainage area does not route to the pond except through infiltration inflow (I/I) to the subsurface storm pipes.

2.3.2 *Pond Design*

The study pond is an extended wet detention pond near the center of the Englesby Brook watershed (FIGURE 2). The design is generally consistent with the extended wet detention pond design included in Vermont's Stormwater Management Manual (VT-ANR 2002). A water quality storage volume for the pond was sized based on the expected runoff from a 0.9" storm over the impervious area contributing flow to the pond and is allocated between the forebay, the permanent pool, and 38.1 cm of extended detention storage between the permanent pool level and weir notch cutout on the outlet riser. The pond was also designed to detain the runoff volume from the 1-year design storm (2.1" over 24 hours, SCS Type II distribution) for 4.6 hours on a center of mass basis and to reduce the peak rate from the 10-year storm (3.2") by 55%.

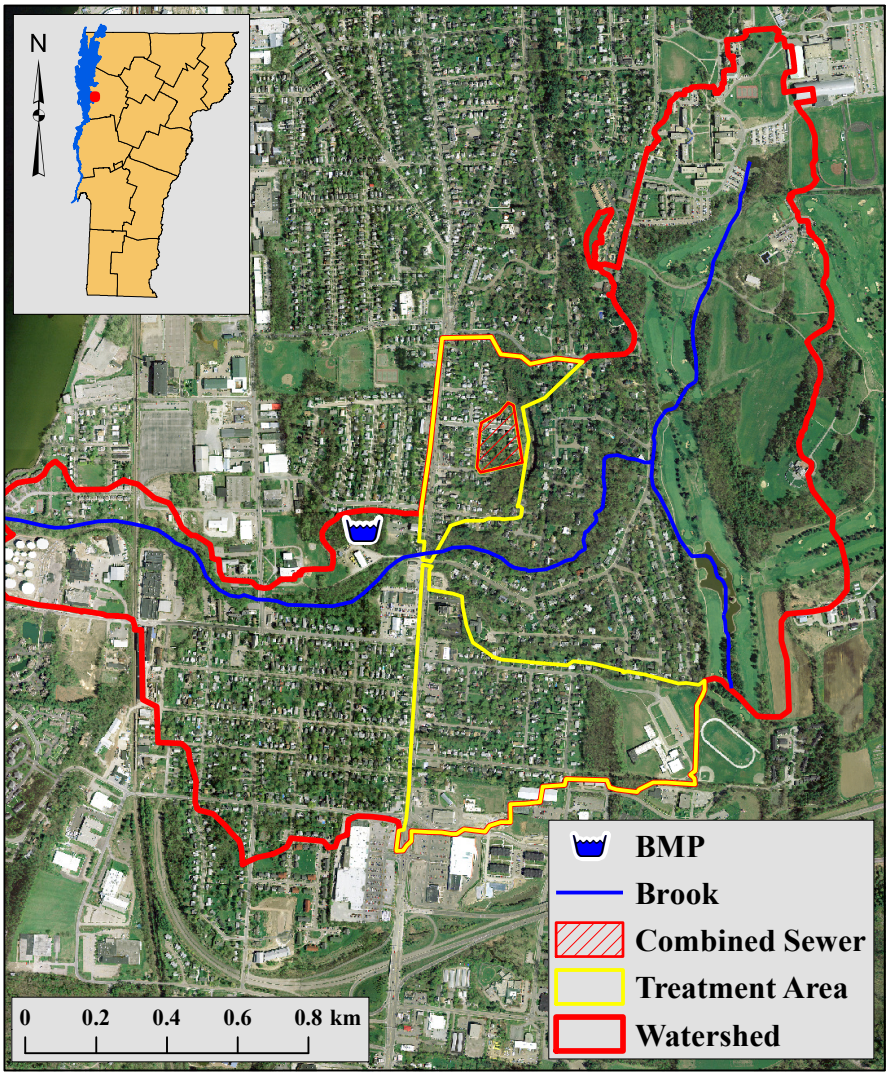

FIGURE 2. Map of the study area identifying the Englesby Brook watershed boundary and stream channel, location of the study pond and its contributing area, and area within the treatment area draining to a municipal wastewater treatment facility. (Imagery date is May 2004, downloaded from Vermont Center for Geographic Information.)

Prior to construction of the pond in 2005, piped stormwater from the 48.7 ha treatment area discharged directly to Englesby Brook. Inflow to the pond is through a 0.91 m diameter collector pipe that connects to a concrete inlet riser adjacent to the pond forebay. The inlet riser redirects flow into the forebay and includes a flow splitter /

overflow diversion designed to divert runoff above the 10-year flow directly to Englesby Brook. Flow into the pond enters the forebay and travels 55 m before exiting the forebay at the opposite end via an all-around spillway in the forebay outlet riser. Outflow from the forebay outlet riser discharges to a 35 m stone-lined channel, terminating in the main pond's permanent pool. The main pond is irregularly shaped, and underlain by a geomembrane liner atop a constructed clay core. Outflow from the main pond occurs at four stage thresholds, with a 1.83 m permanent pool (at its deepest) maintained below the lowest outlet. Outflow occurs first at a proprietary vortex orifice installed near the bottom of the outlet riser. A rectangular notch weir is located 38.1 cm above the bottom of the vortex orifice, which is cut into an all-around spillover outlet in the outlet riser with an invert at 63.5 cm above the permanent pool level. Lastly, there is an emergency side spillover from the main pond designed to accommodate either the 100-year design storm or backed up flows in the event of clogging of the regular outlet orifices. Outflow from the outlet riser box flows 168 m through a 0.61 m diameter pipe before discharging to a stabilized bank adjacent to the Englesby Brook channel.

2.3.3 *Flow gaging and water sampling*

USGS personnel installed monitoring equipment at the inlet and the outlet of the detention pond in the summer of 2007. Flow gauging consisted of Sutron bubble gages in PVC stilling wells installed inside the inlet riser and on the outside of the main pond outlet riser. Stage and computed flow were recorded at a 5-minute interval. Stage-discharge ratings were developed using a combination of design drawings, site surveys,

and temporary weir plates. This instrumentation worked well, except for a tendency for the low flow outlet vortex to clog. On at least one occasion during our period of sampling, the vortex clogged substantially causing the pond stage to rise to the outlet weir notch in the period following a small storm. This clog was cleared manually. However, water balance calculations led us to believe that partial clogging occurred at other times as well, resulting in less outflow than the corresponding stage reading would suggest. Analysis of the flow record suggested this occurred infrequently and corrections to the flow record were made using USGS standard procedures.

Storm event samples were collected using ISCO 3700 autosamplers positioned on top of the inlet diversion structure and outlet riser, linked to and triggered by the flow gaging through Campbell Systems Dataloggers. Continuous temperature and conductivity readings were also collected at these inlet and outlet locations. Water samples were collected in proportion to flow volume, into a single composite jug per sampling location and storm event. The start of event sampling was either triggered by exceedance of a predefined stage threshold or was triggered manually during a site visit in advance of a storm. The auto-sampler program ran until either the composite jug was filled, or a site visit occurred for collection of the sample.

Composite and grab dry weather samples were also collected at the pond inlet and outlet. To collect the composite dry weather samples, I reprogramed the autosamplers to fill the composite jug via flow proportional sampling over a period of approximately 24 hours. The resulting composite samples included between 18-75 ISCO aliquots per daily sample and were collected on days when it had not rained more than 2 mm in the 48

hours preceding the onset of sampling. These 1-day dry weather composite samples were collected once at the inlet and outlet during spring, summer, and fall seasons. Several single grab samples were also collected during summer months. These were collected by positioning a sample container where free discharge entered the forebay for the inlet and from within the permanent pool adjacent to the vortex orifice for the outlet.

After collection, all storm and dry weather samples were either transported directly to the State of Vermont's National Environmental Laboratory Accreditation Program (NELAP) accredited analytical laboratory for analysis, or were preserved and stored at a University of Vermont laboratory until subsequent transport to the state lab. The collected samples were analyzed for total nitrogen (SM-4500 N C persulfate digestion) and total phosphorus (EPA-4500-P F), both on an unfiltered basis.

2.3.4 *Flow and Loading Analysis*

I post processed the flow records to enable an analysis of period of record (POR) peak flow rate reductions. Event based hydrologic analysis was constrained by the lack of identifiable outlet peak flows for many of the smaller inlet peak flows and by the multiple consecutive inlet peaks that could produce a composite outlet peak response. To proceed, an inlet flow threshold of 0.10 $m^3 s^{-1}$ was defined based on a visual assessment of the flow record since storms with flow rates above that level typically had clearly defined outlet peaks. I then ran a sliding window over the inlet flow record to select any peaks over the 0.10 $m^3 s^{-1}$ threshold, with a maximum of one peak per eight hour window selected. Outlet peaks were then identified as the maximum outlet flow rate in the six

hour window following each identified inlet peak, lagged by 15-minutes. All inlet and outlet hydrograph pairs were plotted and visualized as these peak identification parameters were iteratively adjusted. The results of this analysis allowed for event based peak flow rate reduction calculations over the POR.

For the loading analysis, the sampled storm data allowed for direct calculation of TN and TP loads only for the storms that were sampled. To extend this analysis to the many storms that were not sampled, I fit probability distributions to each of the analyte datasets (TN and TP) at each of the sites (inlet and outlet) as a basis for estimating event mean concentrations of storms that were not sampled. The fitted probability distributions were then assumed to characterize the storm driven concentrations during the entire POR. Thus, for each storm for which samples were not physically collected and analyzed, the storm concentrations were instead estimated as random draws from the fitted distributions. For non-storm periods, the collected sample concentrations were not of sufficient size to fit probability distributions. Instead, for each discrete non-storm period of the flow record (by definition, the periods between discrete storm events), a simple resampling scheme was used. That is, inlet and outlet non-storm concentrations from the sample dataset (n=8 TP, n=7 TN, composite and grab, TABLE 3) were randomly selected and used to compute the dry weather load for each discrete non-storm interval.

To estimate the cumulative TN and TP fluxes into and out of the pond over the POR I (1) assigned the actual storm and non-storm composite sample concentration values to the periods during which they were collected, (2) drew random TN and TP values from the fitted distributions for each inlet and outlet event in the record during

which no water sample was collected, and (3) randomly selected, with replacement, discrete non-storm sample concentrations to apply to each non-storm period that was not sampled. Due to the small number of winter and early spring samples in the dataset, this analysis was limited to the periods between April 1 and December 1. This approach allowed us to assign TN and TP concentrations and to calculate fluxes for 617 days of flow record. These individual storm and non-storm load components were then summed over the period of record to give an aggregate estimate of wet weather, dry weather, and total TN and TP fluxes into and out of the pond. Finally, I repeated this entire sampling scheme iteratively to produce 100,000 unique estimates of the fluxes into and out of the pond.

2.4 Results

2.4.1 *Hydrology*

Flow gaging at the pond inlet and outlet provides a basis for assessing the pond's hydrologic performance. Peak flow performance of the pond is presented on an event basis in FIGURE 3. The storm selection criteria used resulted in 89 identified storm events over the POR (2007-06-08 to 2009-10-15). Of those 89 events, the majority were fully conveyed by the low flow vortex orifice, with an average peak flow rate reduction of 94.9% (FIGURE 3; TABLE 2). Twenty six of the storms were at least partially conveyed by the rectangular weir notch, with an average peak flow reduction of 82.1%. For the 12 events that reached the upper cut out in the outlet riser, the average peak flow reduction was 50.1%. The three storms with the highest peak inlet discharge are shown in FIGURE 4, all of which reached the upper most outlet conveyance.

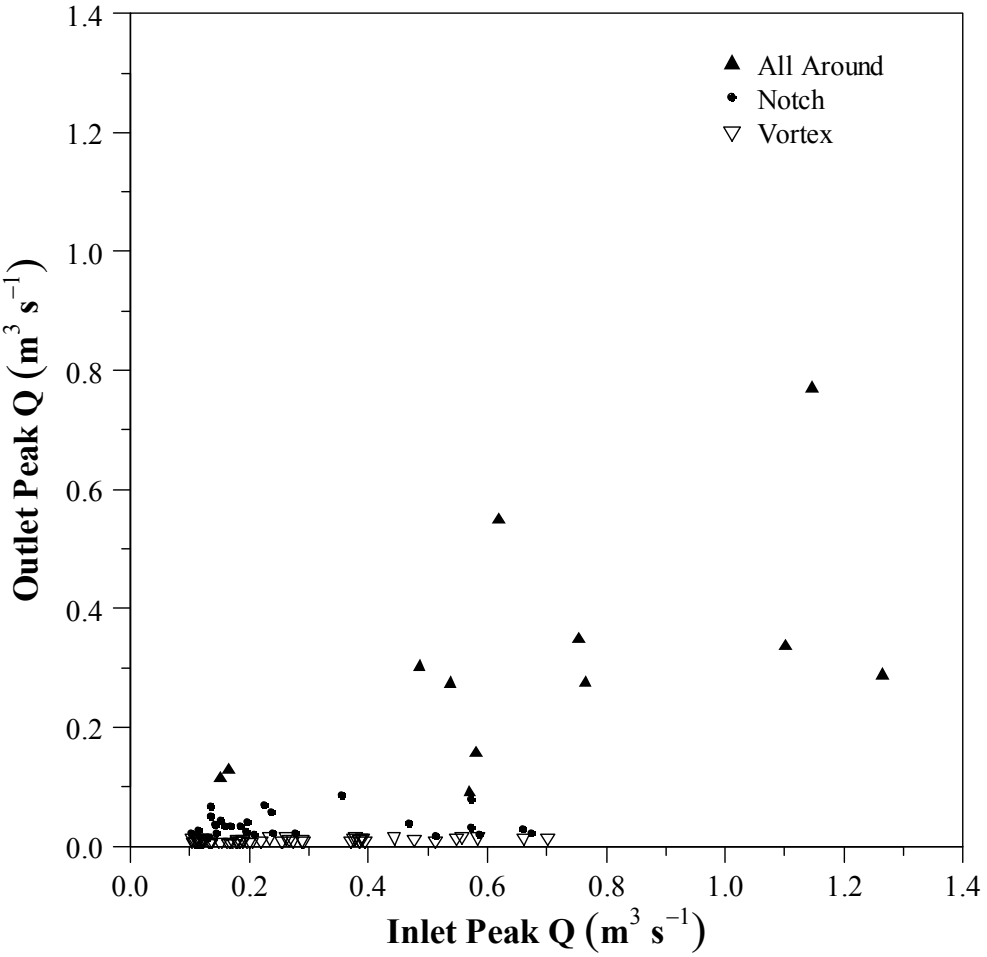

FIGURE 3. Peak inlet and outlet flows for the study pond. Events were identified as the highest inlet peak rate over 0.10 m^3/s within any 8 hour sliding window, resulting in 89 discrete events over the period of record. Outlet controls are described in section 2.3.2.

TABLE 2. Peak flow reduction summary. Eighty nine discrete storm events are grouped by the highest outlet conveyence reached at peak outflow.

Outlet Control	N Storms	Mean Inlet Peak Flow ($m^3 s^{-1}$)	Mean Outlet Peak Flow ($m^3 s^{-1}$)	Mean Peak Flow Reduction (%)
Vortex	51	0.265	0.010	94.9
Notch	26	0.290	0.038	82.1
All Around	12	0.679	0.302	50.1

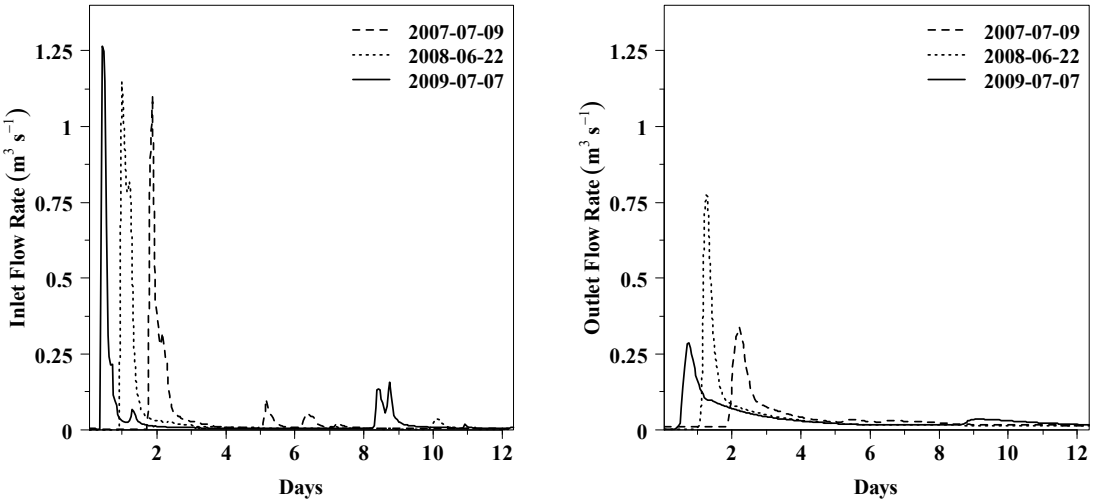

FIGURE 4. The three largest storms within the POR, based on the peak inlet peak flow. Time zero corresponds to 2007-07-09 07:30 EST (21.3 mm of rain in 1 hour), 2008-06-22 17:35 EST (18.0 mm of rain in 6.2 hours), and 2009-07-07 15:15 EST (< 5 mm recorded for the day). The large inflow from the 2009-07-07 event most likely resulted from an intense localized thunderstorm.

I characterized the long-term hydrologic performance of the pond using flow duration curves (FDCs), highlighting the differences in the percent of the time that inlet and outlet flow rates exceeded various magnitudes. FIGURE 5 shows the inlet and outlet FDCs for the 860-day overlapping POR, illustrating several relevant aspects of the pond's hydrologic performance. First, the pond is effective at reducing peak flow rates from the inlet to the outlet, consistent with the event based analysis. This is clearly

demonstrated by the separation between inlet and outlet curves at the flows exceeded less than ~3.4% of the time. As should be expected for a detention pond, for the majority of the record excluding peak flows the outlet flow rate was greater than the inlet flow rate for a given exceedance percentile. However, it was unexpected that the inflow to the pond did not cease entirely for extended periods of time. For example, the 95% flow at the inlet is equivalent to 12 L min^{-1}, which we attribute to groundwater interception by the conveyance network (so called infiltration / inflow or I/I) and non-stormwater discharges to the conveyance network.

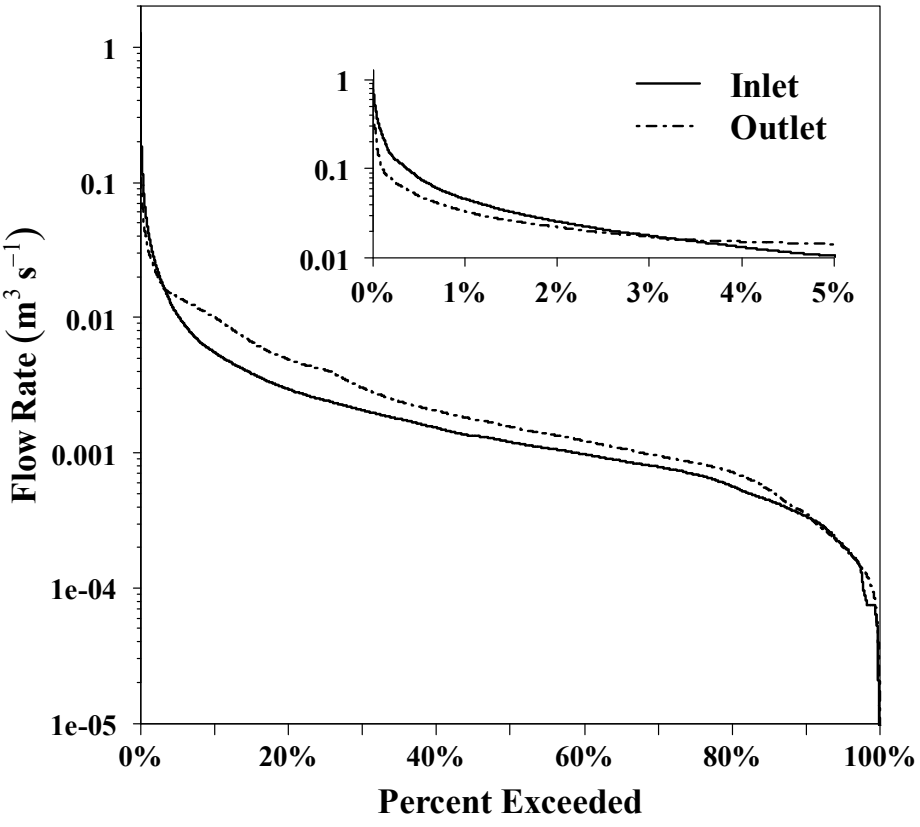

FIGURE 5. Inlet and outlet flow duration curves with a high flow inset in the same units.

2.4.2 *Event mean concentrations, TP*

We collected and analyzed 42-paired inlet and outlet storm samples for unfiltered total phosphorus (TP) (FIGURE 6a). The median inlet and outlet TP storm concentrations were 0.498 and 0.106 mg/L, respectively. I used Levene's test for equality of variances and found variance of the inlet samples to be significantly greater than corresponding samples at the outlet ($p < 0.001$). The concentrations of TP storm samples were approximately log normally distributed at the inlet and outlet and were compared using the nonparametric Mann-Whitney U test for paired means. Inlet TP storm concentrations were significantly higher than the corresponding outlet samples ($p < 0.001$).

2.4.3 *Event mean concentrations, TN*

We collected and analyzed 43-paired inlet and outlet storm samples for unfiltered total nitrogen (TN) (FIGURE 6b). The median inlet and outlet TN storm concentrations were 1.45 and 0.93 mg/L, respectively. As for TP, Levene's test for equality of variances indicated that the variance of the inlet samples was significantly greater than corresponding samples at the outlet ($p < 0.001$). The concentrations of TN storm samples were also approximately log normally distributed, and were compared using the nonparametric Mann-Whitney U test for paired means. I found inlet TN storm concentrations to be significantly higher than the corresponding outlet samples ($p < 0.001$).

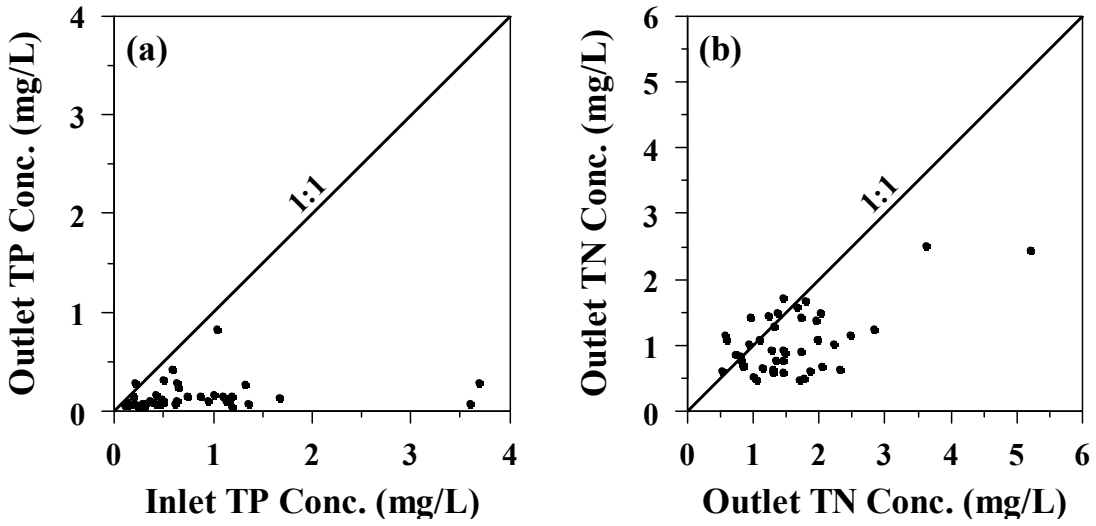

FIGURE 6. Scatter plots of inlet and outlet TN and TP storm composite event mean concentrations.

2.4.4 *Dry Weather Sampling*

Details of the dry weather samples are given in TABLE 3. The small number of dry weather samples (TABLE 3) limits the analysis that can be conducted, however a few findings can be inferred. For TN, the effluent concentration was lower than the influent concentration for all except a single grab sample pair, while the averages of the composite samples were close to the average storm concentrations at both the inlet and outlet. For TP, the pond effluent concentration was higher than the corresponding inlet samples for two of the three daily composites, whereas effluent had lower TP concentration for three of the four grab samples, relative to the inlet. Overall, the TP concentrations of the dry weather pond effluent were close to the mean value measured during storm events, while the dry weather influent concentration was lower than storm event levels.

TABLE 3. Dry weather sample characteristics.

Time of First Sample	Time of Last Sample	# Aliquots	Total Nitrogen (mg L^{-1})	Total Phosphorus (mg L^{-1})	Sampled Flow (m^3)	Days Since Storm > 2 mm
Inlet						
1 Day Composite						
2008-08-23 14:14	2008-08-24 16:01	45	1.85	0.167	102.9	4
2008-09-23 14:04	2008-09-24 19:39	75	1.21	0.052	35.4	9
2009-04-15 14:02	2009-04-16 13:43	75	2.18	0.109	166.9	8
Single Grab						
2007-08-15 13:50	---	1	2.55	0.100	---	9
2010-08-26 18:00	---	1	---	0.035	---	3
2010-08-28 19:15	---	1	0.10	0.050	---	5
2010-08-31 7:00	---	1	3.00	0.013	---	8
Outlet						
1 Day Composite						
2008-08-23 15:13	2008-08-24 15:17	18	0.98	0.243	83.8	4
2008-09-23 14:12	2008-09-24 19:55	75	0.59	0.095	40.6	9
2009-04-15 14:37	2009-04-16 17:46	75	1.71	0.034	194.8	8
Single Grab						
2007-08-15 13:30	---	1	0.59	0.090	---	9
2010-08-26 18:05	---	1	---	0.028	---	3
2010-08-28 19:25	---	1	2.70	0.021	---	5
2010-08-31 7:05	---	1	0.60	0.049	---	8

2.4.5 *Long term flux estimates*

All four EMC datasets were right skewed and were iteratively fit with lognormal, Weibull, gamma, and exponential probability density functions. In all four cases, lognormal distributions were best fit based on comparisons of log-likelihood values from the sets of distribution fits (FIGURE 7). Using these distributions and the dry weather sample concentrations, I proceeded to construct the POR loading analysis as described in Section 2.3.4.

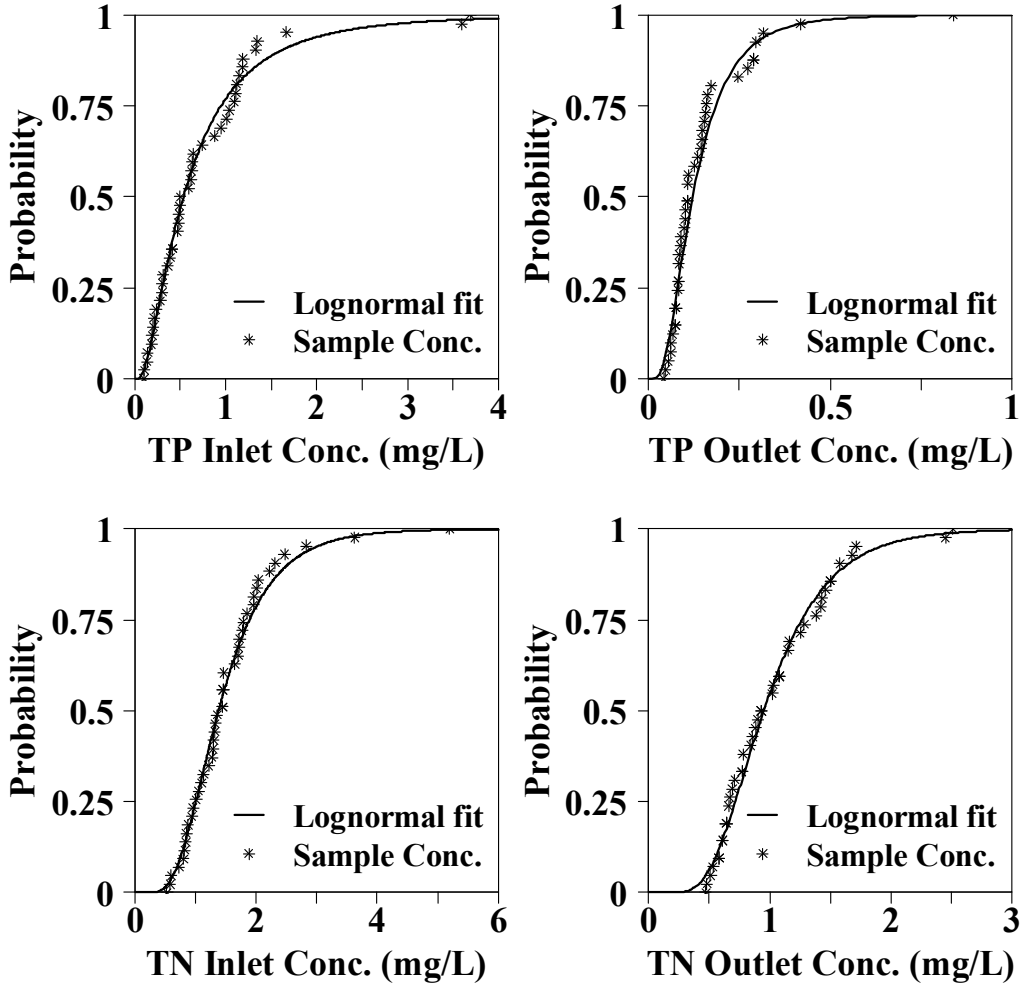

FIGURE 7. Lognormal distribution fits to the TN and TP storm sample sets at the inlet and outlet.

For TN, a relatively large percentage of the total loads at both the inlet and outlet (38.3% and 28.5% respectively) was attributed to non-storm driven loadings (TABLE 4). This is a result of the non-event TN sample concentrations having often been higher than the median storm event TN concentrations. For TP, the total loads at the inlet and outlet were primarily driven by storm events (79.2% and 83.7%, respectively). The total flux estimates are presented as histograms for TP and TN, in FIGURE 8 and FIGURE 9 resp.,

showing the summed results of the 100,000 iterations of total loads into and out of the pond. There was no overlap between the total TP load estimates into and out of the pond, while there is a small amount of overlap in the TN total loads at the inlet and outlet.

TABLE 4. Prediction percentiles of long-term load estimates for TN and TP into and out of the detention pond. The period of analysis includes 2007-06-08 through 2007-12-01, 2008-04-01 through 2008-12-01, and 2009-04-01 through 2009-10-15 (617 days total).

Total Nitrogen	2.50%	50%	97.50%	Total Phosphorus	2.50%	50%	97.50%
Pond Inlet				Pond Inlet			
Storm (kg)	172.1	186.5	203.9	Storm (kg)	81.0	93.9	113.6
Non-storm (kg)	96.1	116.0	135.3	Non-storm (kg)	15.7	24.2	35.3
Total (kg)	277.7	302.8	328.5	Total (kg)	102.4	118.5	140.7
Pond Outlet				Pond Outlet			
Storm (kg)	152.0	167.2	187.6	Storm (kg)	20.0	23.1	27.8
Non-storm (kg)	54.7	66.8	81.9	Non-storm (kg)	3.4	4.5	5.9
Total (kg)	214.2	234.5	259.2	Total (kg)	24.3	27.6	32.5
Reduction (%)		22.54		Reduction (%)		76.69	

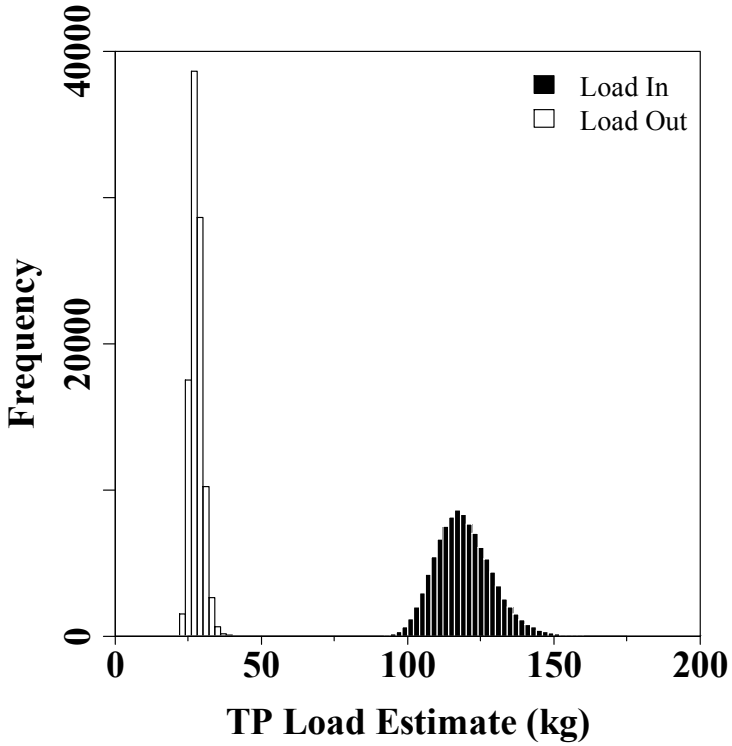

FIGURE 8. Frequency distribution of total phosphorus loads estimated over the 617 day period of analysis into and out of the pond.

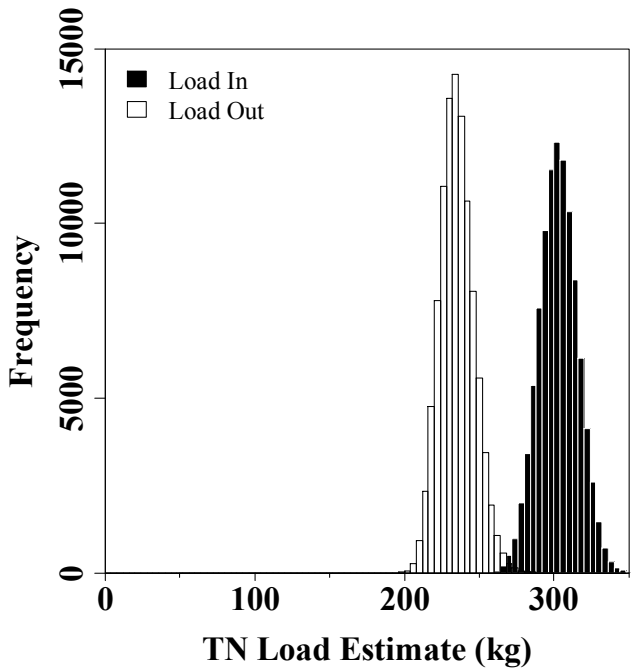

FIGURE 9. Frequency distribution of total nitrogen loads estimated over the 617 day period of analysis into and out of the detention pond.

2.4.6 *Temperature*

Water temperature exhibited strong event, inter-event and seasonal dynamics. On an event basis, storm driven influent in warmer months tended to be warmer than the inter-event I/I inflow, with the shape of the temperature inflow graph closely tracking inflow storm hydrographs (FIGURE 10). A reversed pattern was evident in the late fall and early winter, where colder event driven stormwater caused sharp decreases in inflow temperature relative to inter-event flow (data not shown). For the sustained I/I inter-event flow that was maintained into the pond for most of the POR, the water temperature roughly tracked air temperature dynamics on a daily time scale, with warmer inflow at mid-day in the warmer months, even in the absence of event flow. Water temperature at

the outlet showed small spikes during warmer months that can be attributed to warm storm flow at the inlet, however much larger differences were evident on a daily basis during inter-event periods. The period of 2008-06-07 through 2008-06-11 included in FIGURE 10 illustrates this dynamic over a period of days where air temperature exceeded 30°C over three consecutive daily peaks. These diel fluctuations resulted in 7-10 degree daily swings in effluent temperature, with pond effluent reaching daily maximums near 16:00 and daily minimums near 06:00.

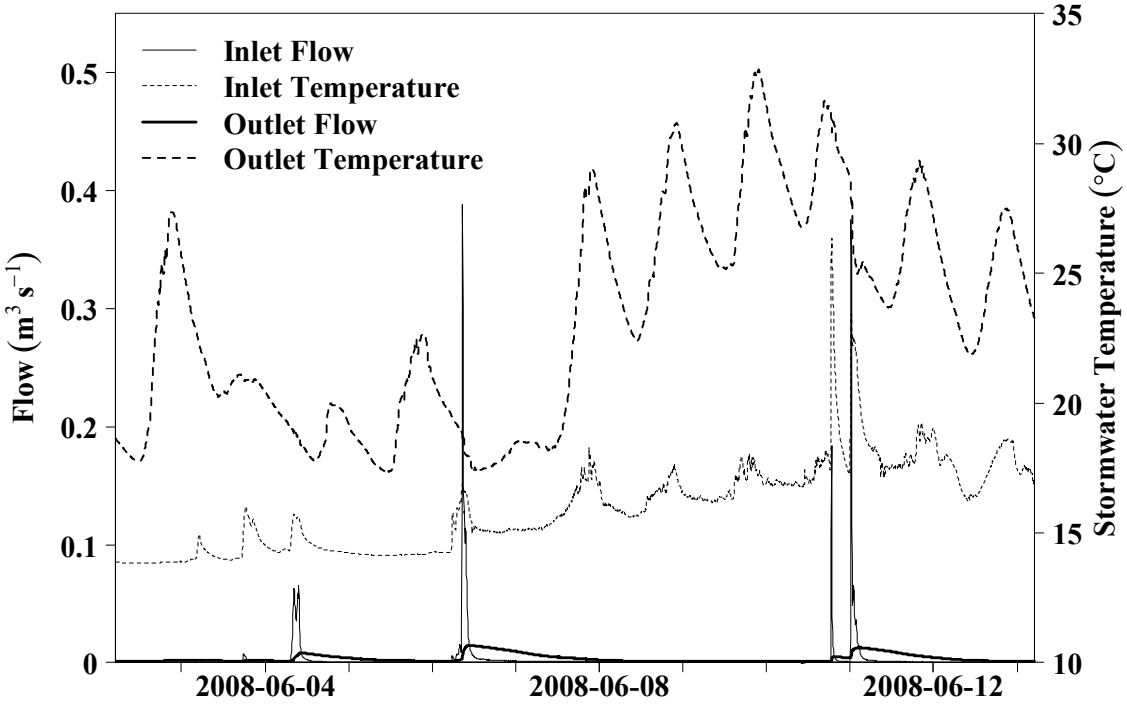

FIGURE 10. Flow rate and temperature into and out of the pond from 2008-06-02 to 2008-06-13, including the POR maximum temperature at the outlet.

On a seasonal basis, inflow and outflow tracked ambient temperatures, but with a greater seasonal amplitude at the outlet, which more closely tracked air temperature. The

annual inlet record (FIGURE 11) showed a seasonal trend driven by air temperature, which was superimposed with storm event spikes (summer) and troughs (fall winter), along with the diel variation seen in FIGURE 10. The outlet also exhibited seasonal dynamics, but with additional variation largely attributable to diel temperature differences. It can also be seen that inlet flow temperature was slower than the outlet to drop to 0 °C. In contrast, the outlet was at or near 0 °C for most of the winter, though it warmed more quickly than the influent through early spring.

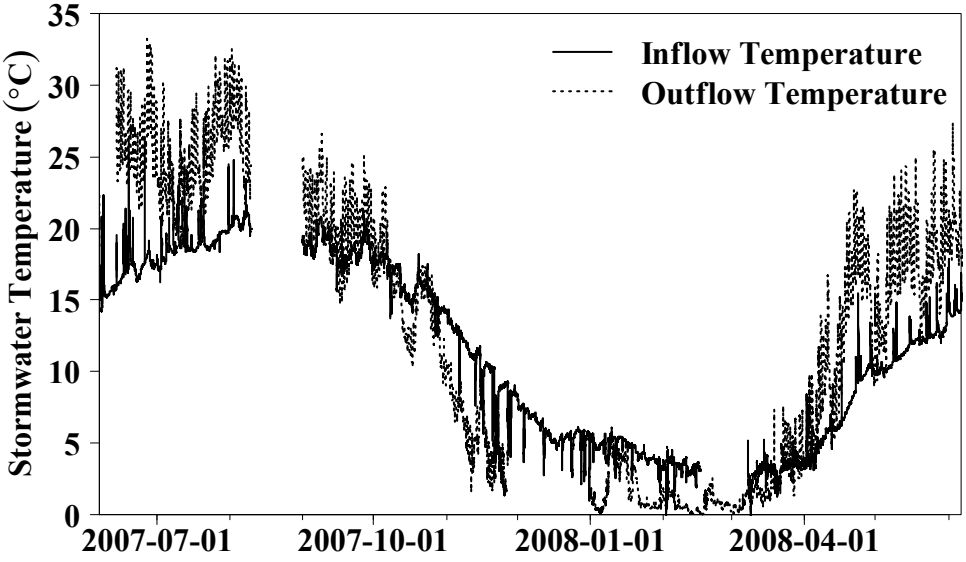

FIGURE 11. Seasonal temperature trends at the pond inlet and outlet.

2.5 Discussion

2.5.1 *Hydrology*

As expected, the flow gauging at the pond inlet and outlet showed the pond to be generally effective at reducing peak flow rates (FIGURE 4 and FIGURE 10). The

broader context for the pond's hydrologic performance is the Englesby Brook watershed TMDL, which targets a 34.4% reduction in the watershed Q 0.3% flow rate relative to a pre-TMDL baseline (VT ANR 2007). The TMDL also includes a base flow remediation target quantified as an 11.2% increase in the Q 95% flow rate, relative to the pre-TMDL baseline. Given that the study pond drains surface flows from 19.9% of the Englesby Brook watershed, this retrofit pond is expected to play a substantial role in meeting the peak flow reduction target. The TMDL flow metrics computed from the TMDL baseline modeling, the study pond flow record, and the streamflow record for the Englesby Brook watershed are summarized in TABLE 5. The Englesby Brook streamflow series is from USGS gage #04282815, which was operated near the mouth of Englesby Brook, approximately 0.8 km downstream of the pond outfall from 1999-10-01 through 2010-09-30.

TABLE 5. Flow metrics used in the Englesby Brook TMDL, calculated from modeled and measured time series. Pre and Post refer to periods of record (POR) before and after construction of the study pond.

Flow Series	Q 0.3% mm/d	Q 95% mm/d
TMDL Synthetic FDC (Daily Flows)		
Englesby Brook Baseline	14.61	0.180
Mean Attainment Target (inlcuding future growth allocation)	9.58	0.200
Pond Inlet		
Daily Mean	6.45	0.043
Instantaneous (5 min.)	20.55	0.036
Pond Outlet		
Daily Mean	6.36	0.037
Instantaneous (5 min.)	11.44	0.036
Englesby Brook		
POR Daily Mean	9.91	0
Pre Daily Means	12.34	0
Post Daily Means	8.77	0.001
POR Instantaneous (5 min.*)	13.65	0
Pre Instantaneous (5 min.*)	15.74	0
Post Instantaneous (5 min.)	12.60	0

* First 8 months of record are 15 minute data

TMDL flow metrics (Q 0.3% and Q 95%) are presented on the basis of daily mean flows and instantaneous[2] flows, by necessity. That is, the synthetic FDCs used in the TMDL modeling were reported at a daily time step, as are the resulting baseline metrics and targets (Tetra Tech 2005). While this temporal resolution is the explicit in the TMDL analysis, a more refined analysis can be done using the existing 5-minute records. This is important because the flow data from the study pond and the watershed

[2] Instantaneous is used in this section to refer to the highest resolution data available and is meant to distinguish those flows from daily aggregated flows. For the study pond, data were collected at a 5 min interval and so instantaneous refers to 5 min data in this context. Englesby streamflow data were initially reported by USGS at 15-minute interval, and later at a 5-minute interval. Thus, instantaneous Englesby streamflow data refers to this mixed resolution time series.

gage show the Englesby Brook drainages to have rapid storm dynamics at the sub-daily time scale, not unexpected given the small area and high level of development within the watershed (TABLE 1). Aggregating the measured flow records to daily mean flows produces comparable metrics to those that were generated in the TMDL analysis, but has the effect of averaging sub-hourly event dynamics with sustained inter-event periods of low or no flow. Aggregation to the daily time step as used in the TMDL analysis can greatly misrepresent the actual flow conditions experienced by the receiving stream. For this reason we have included daily flow as well as instantaneous flow metrics in TABLE 5 and in this discussion, but we emphasize that the daily flow metrics are of limited validity outside of the TMDL regulation context.

Under the TMDL, the high flow reduction target (Q 0.3%) is set at 5.03 mm/d (or 34.4%) on the basis of the whole watershed. Using daily mean flow into and out of the study pond, the Q 0.3% was reduced by only 0.9 mm/d or 1.4% over the study period. Calculating the same quantity from the instantaneous flow record results in a 44.3% reduction in Q 0.3% for the pond's drainage area. Given the fraction of the total watershed treated by the pond, this 44.3% reduction in the pond's instantaneous Q 0.3% corresponds to an 8.8% reduction on a whole watershed basis, or 25.6% of the total TMDL target being met from only 19.9% of the contributing area. The same calculation based on daily mean flows would credit the pond with having achieved only 0.3% of the total TMDL target. As discussed above, the instantaneous Q 0.3% better corresponds to the hydrograph dynamics of the pond, and the difference between instantaneous and daily

mean metrics highlights the misrepresentation that can result from aggregating storm flows to daily resolution.

The pond was not expected to have a large effect on the Q 95% target since it was not designed as an infiltration practice, nor was baseflow restoration an explicit design goal for the site. However, given that the pond detains storm flow volumes from 19.9% of the watershed, it is conceivable that it could contribute to the targeted Q 95% augmentation through temporal redistribution of surface conveyed storm flow. On the basis of daily mean flows, the low flow metric (Q 95%) at the pond inlet was slightly higher than at the outlet, while on an instantaneous basis there was no difference. I also calculated Q 95% values for the pond from a series of median daily flows since the distribution of flows on any given day were typically skewed (median daily flow metrics not included in TABLE 5). From these data, Q 95% increased slightly at the pond outlet relative to the pond inlet. Thus, depending on whether and how the flow data were aggregated the pond can be shown to increase slightly, decrease slightly or have no effect on Q 95%. The most useful summary of these data is therefore simply to note that the empirical Q 95% stream flow for Englesby Brook corresponds to a dry channel (TABLE 5), and that the empirical Q 95% for the pond corresponds to low sustained flow into and out of the pond (approximately 7 liter s^{-1}). It is unlikely that this contributes meaningfully to baseflow augmentation in the Brook.

While there was little difference in Q 95% from the inlet to outlet, the temporal redistribution of storm flows is evident in the empirical FDC (FIGURE 5). Within the 3% and 90% probability of exceedance interval, the outlet flow rate was uniformly

greater than the inlet flow rate for any given percent exceedance. This demonstrates that except for the highest 3% and lowest 10% of flows, the pond generates higher instantaneous outflow relative to inflow. It is only at the relatively infrequent lowest flows that the difference between inlet vs outlet flow approaches zero as corresponding flow rates decline to zero.

The Englesby Brook streamflow gage records provide another indirect means of assessing the extent to which the study pond has contributed to watershed restoration targets. The streamflow data at USGS gage #04282815 were collected over an 11 year period (1999-2010), during which the study pond and number of other large and small BMPs were constructed. The study pond and a shallow marsh wetland were constructed in 2005, while two smaller stormwater ponds (on a contributing areas basis) were constructed in 2006. Other smaller scale stormwater management installations came online throughout the period of record. The construction of the larger stormwater treatment systems in the middle of the gage record provides a natural break for the analysis of the stream flow records, albeit complicated by other small scale changes within the watershed throughout the gage record. The Q 0.3% and Q 95% metrics were calculated over the stream flow record for (1) the entire 11 year POR, (2) the period before which the study pond went online ('pre', 1999-2005), and (3) the period after the study pond went online ('post',2005-2008). These data are summarized in TABLE 5 on a daily mean flow and instantaneous flow basis.

The value of the Q 0.3% flow calculated based on the POR as well as the USGS daily mean flow prior to construction of the pond ("pre-" data) did not match the baseline

Q 0.3% derived from the daily resolution TMDL modeling. This was not a surprise given the differences in methodologies (e.g., measured vs modeled, significant BMPs implementation within the measured POR that were not included in the baseline modeling), producing what should be incomparable metrics. Considering just the measured data, on both the daily and instantaneous bases, the pre- Q 0.3% exceeded the POR Q 0.3% values, while the post- Q 0.3% values were lower than those for the entire POR. This suggests that peak stream flow during the post construction period was lower than in the preceding period, which we attribute at least in part to stormwater management.

The landscape within Englesby Brook was fairly static over the decade of stream flow gaging, except for stormwater management. The area was already largely built out at the onset of flow gaging in 1999, and after 2002 any substantial new or redeveloped impervious area within the watershed was subject to State of Vermont stormwater management regulations, and later the Phase 2 MS4 permitting. Thus, with no major greening or reforestation initiatives within the watershed, the primary land use change over this period of time was the implementation of stormwater management retrofits. The other factor with obvious potential to manifest the observed change in Q 0.3% is precipitation. Medalie (2012) noted in a watershed scale BMP assessment for Englesby Brook over the same period of record that the post- period was significantly wetter on the basis of paired monthly median precipitation totals, though not on an annual total basis due to one relatively wet year in the pre-installation period. The magnitude of low probability flow events (i.e., Q 0.3%) is inherently sensitive to occurrence of low

probability precipitation events in the period for record, including not only the magnitude but the antecedent watershed condition which can influence flow rates. We therefore cannot exclude the possibility that the apparent decline in watershed Q 0.3% from the pre- to post- periods is an artifact of the relatively short periods of record over which they were calculated.

In contrast to the Q 0.3%, the streamflow based assessment of the Q 95% target does not suggest that substantial progress has been made toward watershed hydrology restoration goals, but rather highlights the extent to which the TMDL modeling is incomparable to the measured stream flow data. The TMDL baseline identifies a daily mean Q 95% of 0.18 mm/d for Englesby Brook. However, Medalie (2012) noted that during the pre-BMP construction period, there were between 47 and 64 days per year with mean daily flows equal to zero. These dry periods occurred primarily during summer, and result in empirical Q 95% metrics of zero. This extended dry dynamic was not captured in the TMDL modeling, with the result being that an increase in base flow from even the post- period would have to be by a factor of 500 or greater to meet the attainment watershed target, rather than the 11% increase referenced in the TMDL.

2.5.2 *Event Mean and Grab Sample Concentrations*

Analysis of the baseflow samples is limited by the small sample size, however there did appear to be difference in dry weather versus storm driven TP. While the dry weather TP effluent generally grouped well with the storm driven TP effluent, the influent baseflow and storm samples did not. It appeared the influent TP was lower

during dry weather periods, which I attribute to a lack of sediment and particulate bound TP loading during dry weather periods. The combination of linear forebay and main pond flow paths was designed to settle out suspended loads resulting in outlet loads that would be comprised of dissolved and colloid bound P during storm and non-storm flow conditions. Thus, these dynamics generally corresponded with expected removal mechanisms within the pond.

I also compared the EMC storm data collected in this study with other urban runoff concentration data reported in the literature. This study's median storm TN EMC calculated from all sampled storms at the inlet was 1.45 mg/L, which is lower than most literature estimates for piped urban runoff (TABLE 6). The median storm influent TP concentration of 0.498 mg/L measured in this study, however, was higher than most reported values for piped urban runoff from the literature. The elevated TP concentrations, combined with TN concentrations that were commensurate or lower than other published estimates suggests that there was a disproportionate source of TP within the study pond's contributing drainage area.

TABLE 6. Urban runoff TN and TP concentrations from this study and values reported in the literature.

	Median Urban Runoff EMCs	
Reference	**TN (mg/L)**	**TP (mg/L)**
This Study	1.45	0.498
Previous Studies		
NURP (1983)	2.18*	0.27
Pitt et al. (2004)	2.0*	0.27
Smullen et al. (1999)	2.0*	0.26
International Stormwater BMP Database (2012)	0.75 - 2.37^	0.11 - 0.36^
Steuer et al. (1997)	1.87*†	0.29†
Bannerman (1993)	---	0.66†

*** TKN + NO_2 + NO_3**

^ Range of inlet EMCs reported for different BMP types

† Geometric mean

Analysis of the different land cover types contributing to urban runoff has previously identified residential lawns as a disproportionate contributor of TP, relative to other urban land covers (Bannerman et al. 1993; Steuer 1997). The sources of TP from residential lawns have generally been attributed to lawn fertilizers and pet waste, both of which are plausible contributors within this study's drainage area. Vermont enacted consumer education and voluntary phosphorus fertilizer restrictions in 2011, however no restrictions on residential lawn phosphorus application were in place concurrent with the sampling in this study (2007-09) (Vt. Stat. Ann. tit. 10, § 1266b 2011) ("Application of phosphorus fertilizer,"). We also speculate that pet waste loading could be relatively high within the contributing drainage area. Vermont has one of the highest pet ownership rates in the country (AVMA 2012), and the study's contributing area is majority

residential lots (56% by area) averaging 0.24 acres, and is served by sidewalks. Lastly, through field observation we have noted the presence of drop inlet yard drains within the contributing drainage area, presumably connected to the pond's collection drainage network. The combination of these factors (lack of regulations addressing phosphorus fertilizer, potentially high pet density, and direct lawn and street connection) suggest a possible explanation for the elevated TP EMC data measured in this study.

2.5.3 *Irreducible Effluent Concentrations*

The concept of irreducible effluent concentrations provides further context for considering the effluent quality of the pond in this study (Schueler 2000). Irreducible effluent concentrations refer to the observations that for similar types of stormwater BMPs, there are minimum achievable effluent concentrations beyond which further reductions are not likely. This can be due to re-suspension, desorption, and biological and chemical processing that prevent effluent concentrations from reaching zero or dropping below other thresholds. It follows that influent with a high concentration relative to a BMPs intrinsic irreducible concentration level may be easily reduced, and will compute as a high removal efficiency, whereas influent concentrations close to the irreducible effluent concentration will not be further reducible, resulting in low removal efficiency. Thus, high BMP removal rates can be indicators of either high influent concentrations, good BMP performance, or both.

A comparison of effluent from this study with literature estimates of irreducible wet pond effluents suggests general agreement. Schueler (2000) aggregated data from

the National Pollutant Removal Database (Winer 2000) and estimated 0.13 mg/L TP, and 1.3 mg/L TN as the irreducible effluent limits for wet ponds. Similarly, aggregated data from the International Stormwater BMP Database (ISBMPD 2012) estimated lower bounds on the 95% confidence intervals around effluent EMC medians as 0.19 mg/L for TP and 1.75 mg/L for TN, suggesting that it is infrequent that a wet pond produces effluent concentrations below those values (TABLE 7). The median storm effluent from this study of 0.106 mg/L TP and 0.93 mg/L TN are both slightly lower than the aggregated estimates for wet ponds in general. This may be attributable to the study pond being a recently constructed extended detention wet pond, a variant of wet pond that has not been broken out in previous analyses of irreducible concentrations. Overall, this consideration of irreducible effluent concentrations, combined with the preceding discussion of characteristic runoff concentrations suggests that the study pond both received higher than average TP influent and produced lower than average TP effluent. For TN, lower than average influent was generally released from the pond as lower than average effluent, suggesting the pond was at least commensurate with the performance of other studied wet ponds in this regard.

2.5.4 *Pond Efficiency*

An initial attempt was made to develop predictive models based on the characteristics of the sampled storms that would allow for event specific predictions of TN and TP concentrations into and out of the pond. A large set of predictor variables related to flow, precipitation, and time of year were investigated using linear regression

and best subsets multiple regression. The best predictors identified (i.e., month of year, days since storm with peak flow of at least X) were generally of low predictive value for TN and TP event mean concentrations at the inlet and outlet. A suspected reason for this is that many of the sampled events in our data set included what could be identified post-hoc as multiple discrete events. Due to the relatively small area and high connected impervious cover of the treatment area, discrete events at the inlet can occur within periods of a few hours, so that during a day's sampling more than one discretely identifiable inlet hydrograph was often sampled. Thus, a single composite sampled event mean concentration often included multiple discretely identifiable inlet events, each of which would differ in event characteristics (e.g., peak q, time since previous event). This makes it difficult to detect the effect of these predictors and would complicate the application of predictive models to the hydrograph record. Given these challenges, I instead constructed pond loading estimates using the distribution fitting and Monte Carlo sampling approach described in Section 2.3.4, from which removal efficiencies were calculated.

The EMC reductions and long-term nutrient removal rates estimated in this study using the distribution fitting and Monte Carlo sampling approach are generally in agreement with published estimates for wet detention pond performance. Direct comparison with other studies is complicated by differences in pond design and sizing, contributing land use and influent characteristics, and differences in sample collection methods (Strecker et al. 2001). Nonetheless, there are two published compilations of BMP performance studies, namely the International Stormwater BMP Database

(ISBMPD) (2012), and the National Pollutant Removal Performance Database (NPRPD) Version 3 (Fraley-McNeal 2007), which allow my estimates to be placed in context with previous work (TABLE 7). The most recent aggregate assessment of the data compiled within the ISBMPD characterizes influent and effluent storm EMCs by BMP type (Geotech Consultants and Wrigth Water Engineers 2012). The median influent and effluent TP EMCs from the current study are above and below, respectively, the 95% confidence intervals from the ISBMPD wet pond data. For TN, sampled influent and effluent were both below the 95% CI estimates for wet pond EMCs.

TABLE 7. Wet pond performance estimates from the ISBMPD and NPRPD, and from the Englesby Brook study pond. ISBMPD data are from Geotech Consultants and Wright Water Engineers (2012) summary of retention ponds included in the database. NPRPD data are from Fraley-McNeal (2007). This study's removal efficiencies were calculated from the long-term storm estimates in TABLE 4.

	Total Nitrogen			Total Phosphorus		
	Inflow	Outflow	Removal Efficiency (%)	Inflow	Outflow	Removal Efficiency (%)
ISBMPD						
Median EMC (mg L^{-1})	1.83	1.28	---	0.3	0.13	---
95% CI*	1.60 - 1.98	1.19 - 1.36	---	0.27 - 0.31	0.12 - 0.14	---
n storms	19,259	19,272	---	46,657	48,654	---
NPRPD						
Median	---	---	31%	---	---	52%
Q1-Q3	---	---	16-41%	---	---	39-76%
n studies	22			45		
This study						
Median EMC (mg L^{-1})	1.45	0.93	10.5%	0.498	0.106	75.4%
95% CI*	1.22 - 1.69	0.77 - 1.08	---	0.39 - 0.74	0.08 - 0.14	---
n storms	44	43	---	43	42	---

*** Confidence intervals estimated by bias corrected accelerated bootstrapping.**

Fraley-McNeal (2007) summarized 45 wet detention studies from the NPRPD, on a mass reduction efficiency rather than EMC basis. The studies included in their analysis included only those for which at least five storms were flow or time composite sampled,

and for which the mass efficiency methodology was documented. (It should be noted that some studies are included within both the ISBMPD and the NPRPD dataset, such that these do not constitute independent datasets.) For comparison purposes, these removal efficiency estimates are compared against the POR storm flow flux estimates made in this study (TABLE 4). TABLE 7 includes both of these estimates, and the median and inter-quartile range removal rates compiled by the NPRPD.

The NPRPD estimated median TP reduction was 52%, while our estimated storm TP removal efficiency was 75.4%, near the upper bound of the IQR computed from the NPRPD. We attribute this greater TP percent removal to the relatively high influent concentrations measured in this study, which given the relatively invariant TP effluent concentrations, translates directly into a high percent removal. The NPRPD estimated median reduction for TN was 31%, which is higher than the 10.3% storm-event TN removal estimated in this study. However, our storm TN removal efficiency appears low in part due to an event vs. non-event volume imbalance in the long term analysis. That is, in separating the inlet and outlet flow records into event versus non-event driven flows for the analysis, 23% more water was classified as being event driven at the outlet, compared to the inlet. This was due to the long drain time for the pond, such that consecutive discrete events at the inlet could be intersected by a period of non-event driven flow, while the corresponding outlet flow record was continuously classified as event driven over the same period. This resulted in a greater volume of water at the outlet being classified as event driven and thus included in the event based efficiency

accounting, with the unequal inlet and outlet volumes having the effect of driving down the computed removal rate for event flow.

Lastly, our estimates of pond efficiency are qualified based on a couple of factors. First, the pond was relatively recently constructed at the onset of monitoring and routine maintenance will be required to keep it operating as designed. The forebay was dredged of approximately ~130 m^3 of sediment in 2011, removing some fraction of the TN and TP that we document as removed from transport in this study, and we expect the performance of the pond over time to continue to be dependent on this loss pathway. Another consideration is that our event samples address winter conditions only to a limited degree. This was due in part to freezing and snow conditions making it difficult to collect composite samples during the winter. Thus, it is unknown if the pond performs at the same level during those periods which constitute a third of every year. Finally, source reduction activities, including more frequent street sweeping and catch basin, and improved pet waste, leaf fall, and fertilizer management have the potential to reduce TP and TN loads into the pond. Given the relatively invariant nature of the pond effluent chemistry, we expect that would correspond directly to reduced removal efficiency within the pond.

2.5.5 *Temperature*

The relatively high impervious cover within the pond's contributing drainage area (44.2%) is likely to produce substantially hotter surface runoff than would occur under undeveloped or lower density of development (Galli 1990; Thompson et al. 2008),

however heating of surface runoff was likely mitigated by two factors in this study. First, the contributing area includes a moderate degree of deciduous tree cover (29.6%) that shades patches of impervious surface throughout. Previous work on the temperature of runoff from pavement has not specifically quantified the thermal effects of ambient temperature shaded impervious (Thompson et al. 2008; Janke et al. 2011; Kertesz and Sansalone 2014), however it is likely to be diminished relative to unshaded impervious surfaces (Spronken-Smith and Oke 1998). The second factor is the relatively long conveyance distances through which flow is piped within the contributing drainage area. Sabouri et al. (2013), in an investigation of sewer pipe effects on runoff temperature, identified 'longest pipe length' as a key parameter affecting the cooling of surface heated runoff through the conveyance network, though the upper bound on their analysis was 975 m. The farthest catch basin from our study pond's inlet riser is 1,335 m upslope in pipe distance at relatively shallow slope (2.7% at land surface), while the closest is 120 m. Additionally, interception of groundwater within the conveyance network provides a continual source of relatively cool inflow to the pond. We do not expect this to have a strong effect during large events, when the groundwater volume would be small relative to the storm volume. However, small afternoon storms during summer months have been found to have large thermal effect on receiving waters, given the combination of warm precipitation, warm pavement, and small volume such that the majority of storm flow is comprised of first flush (Herb et al. 2008). For those storms, which are characteristic for this study area, dilution and mixing of stormwater with I/I flow is likely to provide a degree of mitigation. Combined, these factors can lessen heat accumulation by runoff at

the land surface and provide opportunity for heat loss prior to stormwater reaching the pond.

Thermal mitigation within the contributing area and conveyance network appeared to be largely negated by the pond itself. The permanent pool surface area of 2,050 m^2 (550 m^2 as forebay) provides considerable opportunity for warming via solar radiation and atmospheric conduction. There is little shading of the pond water surfaces except for a line of trees on an adjacent lot ~12 m east of the north-south oriented linear forebay, which may provide a degree of shading during morning hours. However, the larger main pond, and the forebay during all except for early morning hours receive no shading except for the emergent vegetation within the pond. There is also a constructed berm extending 1.52 m above the permanent pool water surface within the main pond which could provide a degree of buffering from wind that might cool water within the pond. These factors largely explain the diel temperature fluctuations shown within the main pond in FIGURE 10.

The design features of the pond contributing to its ability to accumulate heat are typical of these types of installations. The only explicit mitigation used in regulatory detention ponds in Vermont is a lowering of the detention time target for effluent that drains to cold water fisheries. Specifically, where 1-year volume control requirements apply, a center of mass detention time of 24 hours is required in warm water fisheries while only 12 hours is to be provided in cold water fisheries. While the study pond was not designed to either standard (rather, it was retrofit on a peak flow control basis), design sizing reported 4.6 hours center of mass detention time for the 1-year design

storm, suggesting a greater degree of thermal accumulation might have occurred were the pond designed to meet 1-year volume control standards.

2.6 Conclusions

The performance assessment conducted in this work adds to the body of research on stormwater pond effectiveness. The hydrologic assessment demonstrated the pond's effectiveness in reducing peak flows from the contributing areas, and the overall role that this large wet extended detention pond serves in the broader flow-based Englesby Brook TMDL. The total nitrogen and total phosphorus influent, effluent, and mass efficiency results were found to be generally in agreement with previous published estimates, with a few exceptions as noted in the Results and Discussion sections. While the hydrologic and nutrient performance was in general as good or better than expected based on previous research, the temperature data collected at the inlet and outlet demonstrates the heating that can occur while water is stored within a stormwater pond. In watersheds where thermal pollution is of concern and stormwater management is required, under-drained gravel trench outlet pond modifications (ME-DEP 2006), rock crib treatment trains (Thompson et al. 2008), or use of alternate stormwater management strategies less likely to accumulate heat (Long and Dymond 2014) to the same level seen in this study may be warranted.

Lastly, as the consideration of the Englesby TMDL targets made clear, the temporal resolution of hydrologic analysis in flashy systems can have a profound effect on calculated metrics. While we could document the hydrologic performance of the pond

in isolation, there were challenges in constructing a meaningful relationship from those data to watershed daily flows. A more robust regulatory and management framework would explicitly account for higher resolution flow dynamics at multiple channel points along the watershed to ensure that individual and cumulative management efforts have the intended effect on base flows and peak flows along the channel length as opposed to exacerbating hydrologic impairments (McCuen 1979). However, such a framework is unlikely to be adopted unless required (e.g., if the existing frameworks is demonstrably failing to meet management objectives), given the cost and other complications of adopting this approach across the many stormwater impaired watersheds of Vermont.

CHAPTER 3. GLOBAL SENSITIVITY ANALYSIS AND EVOLUTIONARY CALIBRATION OF SWMM HYDROLOGY AND WATER QUALITY FOR A MIXED LAND USE AREA

3.1 Abstract

EPA SWMM is a widely used hydrologic, hydraulic and water quality model, though its application can be limited due to its deterministic nature, high dimensional parameter space, and the resulting implications for modelling uncertainty. In this work, I apply a global sensitivity analysis and evolutionary strategies (ES) calibration to SWMM to produce model predictions accounting for parameter uncertainty for a headwater tributary case study in South Burlington, Vermont. I also assess two different methods to specify subcatchment width, a key SWMM parameter, including the novel methodology developed by Guo and Urbonas (2009). SWMM parameter sensitivity was found to vary based on model structure and demonstrated both sensitivity and lack thereof among the numerous subsurface hydrology parameters SWMM employs. The ES approach was generally successful at calibrating selected parameters, although less so as the number of concurrently varying parameters increased. Lastly, a watershed water quality analysis using the calibrated model suggested that for different events in the record, the stream channel was alternately a source and a sink for sediment and nutrients, based on the predicted washoff loads and the measured loads from the stream sampling stations. Cumulatively, this adds to the volume of previous work on SWMM sensitivity analysis, auto-calibration, and parameter uncertainty assessment, and demonstrates the implications of high dimensional parameter uncertainty in a typical SWMM modeling context.

3.2 Introduction

In the most recent compilation of nationwide water quality assessments, 54% of assessed stream and river miles were found to be threatened or impaired (U.S. EPA 2013), with urban runoff and associated pollutants being a leading attributed stressor. Urban development contributes to the impairment of waterbodies through hydrologic modification of the land surface and through changes in loading to receiving waters due to different activities at the land surface (Lenat and Crawford 1994; Paul and Meyer 2001; Wissmar et al. 2004). These urban nonpoint loads are highly variable (U.S. EPA 1983; Pitt et al. 2004), as is the effect of urban hydrologic modifications such as percent impervious cover (Booth and Jackson 1997; Schueler et al. 2009; Fitzgerald et al. 2012). As a result, the presence of urbanization and water quality impairment does not deterministically define the most appropriate management strategy.

The EPA's Stormwater Management Model (SWMM) is a widely used hydraulic and hydrologic water quality model for developed landscapes. With routines for simulating surface runoff, subsurface runoff, pipe and channel hydraulics, stormwater treatment practices, and buildup and washoff (Bu/Wo) dynamics, it is both widely useful and very highly parameterized. Many of the parameters are either empirical, conceptual, or typically lumped in practice such that there is no singularly correct SWMM parametrization (or structure) for a given drainage area. Thus, while EPA SWMM is inherently a deterministic model predicting a single set of model outputs from a single set of parameters and inputs, there are typically many different parametrizations and structures which could be used to represent a system of interest, potentially yielding

different simulation results. That there are multiple disparate model structures and parameterizations which may be equally good at simulating a particular period of monitored record is the concept of equifinality (Beven 1993), which has shaped a considerable body of environmental modeling research in recent decades. In the case of SWMM, simply relying on modeler's best judgement point estimates for uncertain parameters where the feasible parameter space can generate a large range of outputs (i.e., ignoring equifinality) can limit the use and acceptability of SWMM results for management purposes.

Previous research efforts have explicitly addressed issues related to SWMM's high dimensional parameter space, for example conducting sensitivity analyses on model components of interest (Gaume et al. 1998; Aronica et al. 2005; Barco et al. 2008; Krebs et al. 2013; Sun et al. 2014; Zhang and Li 2015) and using evolutionary and Monte Carlo approaches to engage the high dimensional parameter space (Balascio et al. 1998; Barco et al. 2008; Krebs et al. 2013; Knighton et al. 2014; Zhang and Li 2015). Nonetheless, there remains a need for sensitivity analysis work on previously unassessed model components, as well as for further sensitivity analyses of other components when they are applied to model structures or modeling objectives that differ from those used in previous assessments. Additionally, while several approaches have previously been employed to account for SWMM's high dimensional space in predictive modeling applications, this area remains both relatively unexplored and without a definitive preferred approach.

3.2.1 *Goals*

The goals of this study were to:

1) Assess the sensitivity of SWMM outputs to surface hydrology, subsurface hydrology, and water quality input parameters using a global approach;
2) Apply a combined evolutionary calibration and Monte Carlo parameter uncertainty estimation approach to a SWMM model of a developed headwater drainage area, informed by the results of the sensitivity analysis; and
3) Estimate the contributions of neighborhood surface washoff to watershed loads using a combination of calibrated SWMM results and measured data.

3.3 Methods

3.3.1 *Study Site*

Our study site was a headwater tributary to Potash Brook, in South Burlington, Vermont. Since 1989, state biomonitoring within Potash Brook has frequently scored the watershed as fair and poor condition which led to its identification as an impaired stream on the U.S. Environmental Protection Agency's 303(d) list in 2004 and ultimately to the development of a stormwater total maximum daily load (TMDL) in 2006 (VT ANR 2006). Approximately 53% of the Potash Brook watershed is developed with much of the development built without modern stormwater management. Consequently, stormwater management has been focused on remediation efforts, in addition to treatment of new development and redevelopment.

This project was part of a larger program of research called "Redesigning the American Neighborhood" (RAN), which was focused on involving homeowners in a process to reimagine how to manage stormwater in the context of emerging regulations that were more stringent than existed in the past. A detailed explanation of the overall RAN project is provided by McIntosh et al. (2006), and so only the aspects relevant to this study will be reviewed here.

The study area is a tributary of Potash Brook and originates in an agricultural field which then flows north through a residential subdivision (the Butler Farms and Oak Creek Village neighborhoods, hereafter BF/OCV). The intermittent channel at the south end of the BF/OCV neighborhood is fed by a 49 hectare upslope field area, which is 71% agricultural (primarily hay) with the remaining 29% consisting of single family residences along the drainage area perimeter. From the south end of the BF/OCV neighborhood, the channel flows ~ 900 meters north through the subdivision, over which the channel has been straightened and passes under four roadways via culverts. Over this distance, the channel accumulates an additional 68.3 ha of drainage area, 53.3% of which is the neighborhood. Most of the 245 lots within BF/OCV were constructed on top of fill of unknown characteristics, because the underlying native soils are a combination of poorly draining clay and slit clay textures (i.e., Hydrologic Soil Group D). The neighborhood's drainage system consists of curb, gutter and catch basin conveyance, most of which discharged directly to the brook at numerous discrete outfalls. Near the north end of the neighborhoods, there were two dry detention basins that had limited treatment capacity and did not comply with current design standards. These basins

provided only a modest degree of storage and attenuation, due to the relatively large diameter outlet pipes that were installed at the low points within the ponds. The remainder of the lower catchment area outside of the neighborhood consists of portions of a golf course and meadow.

3.3.2 *Data Collection*

Data were collected in four discrete locations; two in-stream discharge monitoring and sampling stations and two closed conveyance stormwater outfall monitoring and sampling stations (FIGURE 12). The closed conveyances (East Drain and West Drain) and the upper in-stream site (SW1) were hydrologically nested within the area measured at the lower in-stream site (SW2). Flow gaging at the in-stream sites was implemented by embedding concrete paving stones along the two cross-sections of the streambed, and mounting ISCO 720 submerged probes on plates affixed to the concrete blocks. Stage-discharge relationships were developed by area-velocity discharge measurements taken at the cross-sections. The ISCO 720 submerged probes were connected to ISCO 6712c auto-samplers, which were programmed with the stage-discharge ratings. In-pipe sampling consisted of In-Situ Level Troll 500 vented pressure transducers mounted inside of the 47 cm diameter PVC storm pipes using ISCO mounting rings. The pressure transducer cables were routed through flexible conduit to protect against abrasion, and were connected to ISCO 6712 auto-samplers positioned on the stream bank above the outfalls. The storm drains terminated at flared concrete aprons, with at least 10 cm of drop from the apron invert to the stream bed during typical flow conditions. Flow rates

in the pipes were calculated as a function of depth using Manning's equations with a variable roughness as described by Wong and Zhou (2003). In-stream data were collected at 5-minute interval, while in-pipe flow was recorded at 1-minute interval. Rainfall data were collected by a tipping bucket rain gage installed near the centroid of the SW2 drainage area and were recorded as the number of tips per 5-minute interval. A switch was made from a HOBO bucket to a Rainwise Bucket in 2009, changing the precision from 0.2 mm per tip to 0.254 mm per tip. Overall periods of hydrologic record and the numbers of storms sampled per site are given in TABLE 8.

.

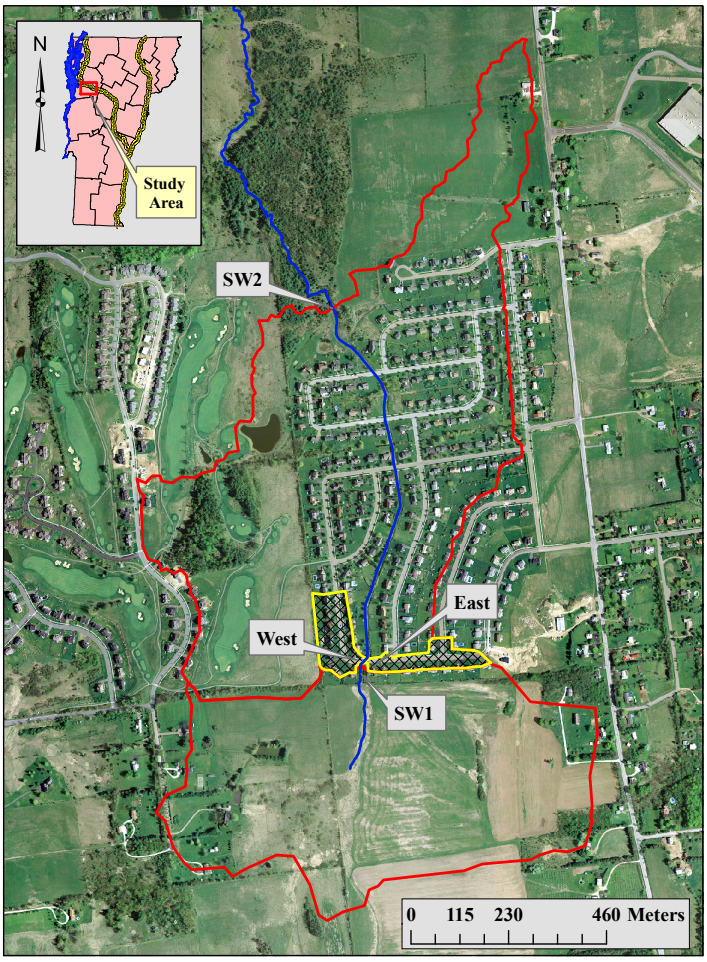

FIGURE 12. Study watershed showing the nested drainage areas within SW2. (Imagery date is May 2004, downloaded from Vermont Center for Geographic Information.)

TABLE 8. Periods of hydrologic record and number of storm samples per site.

Sampling Location	Years of Flow Record	Days of Flow Record	Storms Sampled
SW1	2007 - 2009	582	27
SW2	2007 - 2009	653	40
East	2008 - 2009	337	13
West	2008 - 2009	405	20

Stream water quality samples were collected using a flow-weighted sampling program, triggered by stage exceedance and stage rate of change thresholds. Stream samples were collected by sample lines attached to fence posts installed ~1 m upstream of the flow measurement probes and were pumped into a single composite jug per site and per storm. Samples from the storm pipes were collected by sample intake lines affixed to the cable conduit, approximately 0.25 m downslope of the sensors, near where the PVC storm pipes terminated at the beginning of the concrete aprons. In addition to the composite storm samples, we collected samples during one storm using 24-bottle kits in the auto-samplers, allowing us to sample analyte concentrations throughout an event. All collected samples were transported to an EPA-certified commercial lab for analysis, where they were analyzed for total suspended sediment (TSS) (EPA 160.2), unfiltered total phosphorus (TP) (EPA 365.1), nitrate (EPA 300.0), nitrite (EPA 300.0) total Kjeldahl nitrogen (EPA 351.3 / 350.1), and chloride (EPA 300.0). Nitrite was dropped from further analyses in 2008 following repeated low values and non-detections.

3.3.3 *Model Parameterization*

A base parameterization SWMM model for the 117 hectare area draining to the SW2 monitoring station was constructed using a combination of GIS data, site specific survey data, and parameter guidance from the SWMM User's Manual. Three-meter LIDAR data were used to delineate the overall catchment area draining to the SW2 outlet location and to estimate the slope parameter for each SWMM subcatchment. A municipally funded drainage survey of the BF/OCV neighborhoods was completed in

2005, which provided detailed data on storm drain conveyances for the neighborhood. Using these data, I adjusted drainage area delineations for portions of the neighborhood that topographically drained to the study brook, but which are piped to an adjacent drainage area.

The combined LIDAR delineation and site-survey data resulted in 21 discrete SWMM subcatchments within the 117 hectare drainage area. Impervious area was manually digitized for each subcatchment by overlaying drainage area boundaries on aerial imagery to create a total impervious layer. A directly connected impervious surface layer was subsequently created by assuming roads, sidewalks approaches, driveways, and roof areas that pitched toward the street all drain to catch basins, while backward sloping roof areas were assumed to drain to pervious areas. A combination of total imperious area, directly connected impervious area, and spatial configuration of those areas within each subcatchment was used to define subcatchment impervious area and internal routing. Surface parameters for each subcatchment, including Manning's coefficient and depression storage depths for pervious and imperious surfaces were initially defined using suggested values from the SWMM User's Guide.

For each discrete SWMM subcatchment, SWMM utilizes a width parameter to define the shape of the kinematic wave cascading plane used in the Manning's overland flow runoff model. Per the SWMM User's Manual, the width parameter can be estimated as the subcatchment area divided by the average overland flow path and subsequently calibrated to improve model performance (Rossman 2010). More recently Guo and Urbonas (2009) have derived a parabolic shape function relating watershed shape to

kinematic plane width, length and slope. The relationships make use of the subcatchment area, length of collector channel, and relative position of the collector channel within the subcatchment to compute a kinematic plane width and slope that preserves both area and vertical fall relative to the natural irregular subcatchment. Guo et al. (2012) field tested this approach and found that the shape function generated widths a priori that were similar to those derived from calibration of the flow data. Following this approach, I have utilized the SWMM Manual and the Guo Methods in this work to further test the relatively new Guo methodology and out of necessity for a calibration-free approach to specify width for most of the subcatchments within our model for which we did not have gauged flow data. To do so, I estimated upper and lower probable bounds on collector length and area skewness parameters for each subcatchment from the previously discussed mapping, which allowed for a Monte Carlo sampling approach to the Guo Method.

Conveyances within the subcatchments were simply represented using subcatchment internal routing. More explicit routing, including finer subcatchment discretization routed to the catch basin and storm line conveyance network was explored using the data collected in the drainage survey. However, based on initial exploration of SWMM model structures for the East Drain and West Drain outfall areas, improved model performance was not obtained by explicitly including catch basins and storm lines. Thus, I opted to use a lumped-per-outfall discretization, as described above, and accounted for the rapid transport provided by the conveyance network using subcatchment internal routing and calibration of surface parameters.

Stream channel conveyance was defined by measuring the linear channel segments between culverts via aerial imagery, supplemented with field measurements of channel cross section geometry. Channel geometry measurements at the SW1 and SW2 stations did not differ substantially and so I assumed the SW1 geometry represented the upper half of the channel with the SW2 cross section geometry representing the lower half of the channel. Field observations of the culverts in the neighborhood channel were used to specify culvert loss coefficients and to confirm culvert dimensions as reported in a state database. The two dry detention ponds at the lower end of the neighborhood were parameterized using original design drawings, confirmed by aerial imagery and LIDAR data. A geometric wetland area located within the conveyance network in the lower part of the neighborhood was separately specified as storage node based on LIDAR data and field observations.

Given the study objectives to model flow from the constructed neighborhood in addition to the pervious areas contributing flow to the monitoring stations via surface and subsurface flow, I included the SWMM subsurface flow components within the model structure. Detailed data were not available to characterize the existing subsurface physical environment, nor are detailed data over the vertical profile easily incorporated into SWMM's subsurface model. Instead, the SWMM subsurface model was simply treated as a conceptual model that provides a degree of attenuation to infiltrated water before potentially returning flow to the stream channel. Subsurface parameters were initially defined using soil survey data and SWMM suggested values and were considered

to be free parameters within physically plausible ranges for calibration, as needed to improve model performance during hydrograph recessions and inter event periods.

While the subsurface model includes a physically based representation that I have treated as a conceptual model, SWMM water quality load generation and washoff models are inherently conceptual. Among the options SWMM provides, I used the exponential buildup and washoff functions, applied to total nitrogen, total phosphorus, and total suspended sediment. Initial estimates of maximum buildup coefficients were estimated by dividing developed land annual export rates for TN, TP, and TSS (Novotny 2003) by 24, limiting total buildup between events to the load that would accumulate in half a month if total annual loads accumulated evenly through time over the year. Ranges of buildup exponents and washoff model parameters were initially estimated using a manual trial and error process, but were subsequently given broad feasible ranges in which they could be calibrated.

Shaw et al. (2010) have discussed how several example applications of exponential buildup models in the literature can be equivalently modeled using a constant available mass (CAM) approach, based on the observation that available washoff loads are frequently not dependent on time between events. Further, they identified rainfall kinetic energy and runoff volume as being stronger predictors of particulate washoff loads than antecedent dry days. While we did not incorporate these additional predictors into our model given the lack of means for incorporating them into the SWMM engine, by allowing the buildup exponent (*BuCo*) the opportunity to evolve to an upper bound

during calibration, we allowed for an approximate form of the CAM model to be evolved, should it fit the data.

Evapotranspiration (ET) was initially calculated using the SWMM internal implementation of the Hargreaves equation (Hargreaves and Samani 1982), using air temperature, Julian day, and latitude to compute daily potential ET, from which actual ET can be realized. However, initial exploratory modeling suggested that the Hargreaves approach generated more ET than was suggested by measured flow water balance calculations, which is consistent with previous research findings that the Hargreaves method benefits from calibration (Aguilar and Polo 2011). Lacking the means to calibrate the Hargreaves method (i.e., SWMM‘s implementation is not easily user modifiable), I next implemented the more data intensive Penman-Monteith equation (Monteith 1965) using available atmospheric data and an assumed 10 cm grassland, following Dingman (2002). Without any parameter adjustment or calibration outside of the base parameterization, this approach reduced annual ET estimates by approximately 10%, bringing simulated ET closer in line with what was calculated using a water balance approach. Thus, I proceeded by computing potential ET using a spreadsheet based Penman-Monteith model and input these values to SWMM using the climate input file option.

3.3.4 *Sensitivity Analysis*

In this work, I use the Regionalized Sensitivity Analysis (RSA) methodology (Freer et al. 1996), due to its global properties, its prior applications to SWMM and other

environmental models, and its informative and easily interpretable results. To construct the RSA plots, a large number of simulations were run by randomly sampling the parameter space for the parameters of interest. A performance measure was then calculated for each run, and all simulations were ranked by the performance measure and separated into *n* equal sized bins, for example with the best 1/*n* simulations in first bin and the worst 1/*n* in the *n*th bin. For each parameter of interest, the cumulative marginal distributions for each of the *n* bins were then plotted together. A parameter to which the performance measure was not sensitive will be equally represented among the best and worst bins, and all bins will plot as straight lines. In contrast, sensitive parameters may have different levels of model performance associated with different parts of the parameters' range, even while allowing for concurrent random sampling of other dimensions of the parameter space. These parameters may show separation in the cumulative distributions among the *n* bins, thereby allowing sensitivity to be visualized graphically.

I designed the RSA application to inform the subsequent calibration work and to investigate SWMM parameter sensitivity in a way that would produce generalizable results for other SWMM applications. Further, the approach had to accommodate many parameters and multiple model responses (e.g., storm flow, event recessions, water quality) based on the scope of the assessment. Therefore, to focus this work toward meeting those objectives, I used the RSA methodology sequentially to test the sensitivity of 1) surface hydrology within the East Drain subcatchment, 2) surface and subsurface hydrology within the SW1 subcatchment, and 3) buildup washoff parameters within the

East Drain subcatchment. Using these divisions, I could independently assess SWMM sensitivity using one high impervious cover and one high pervious cover subcatchment for which we had collected measured response data. Details of the RSA batches are described in the following sections.

3.3.4.1 East Drain Subcatchment Surface Hydrology

The East Drain subcatchment area includes 1.2 hectares of relatively homogeneous single lot residential development, with a curb and catch basin closed drainage system. As previously discussed, I did not include the catch basin and storm line system in our parameterization, but rather, used SWMM's overland flow model with routing of pervious areas to impervious areas to represent the flashy runoff dynamics of this area. The surface hydrology parameter ranges used for the East Drain subcatchment RSA runs are presented in TABLE 9. I ran a total of 40,000 simulations for 91 days of record (2008-7-28 through 2008-10-27), producing 40,000 simulated hydrographs for comparison with the measured flow record. Root mean squared error (RMSE) was used as the objective function for this component, given the flashiness of the observed record and relative importance of peak flows in simulating the area, calculated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(sim_i - meas_i)^2}$$

where *N* is the number of discrete measurements, *sim* is simulated flow rate, and *meas* is measured flow rate.

TABLE 9. Uniform distribution parameters for the East subcatchment SWMM parameters varied in the RSA runs. Abbreviations used in the text are included.

Parameter	Abbreviation	Units	Min	Max
Area	*Area*	*ha*	1.17	1.30
Impervious Cover	*ICPct*	*%*	39.6	59.4
Width	*Width*	*m*	5	250
Slope	*Slp*	*%*	1.04	3.74
Impervious Manning's n	*nImp*	*unitless*	0.001	0.02
Pervious Manning's n	*nPrv*	*unitless*	0.05	0.45
Impervious Depression Storage	*DSPrv*	*mm*	0	5
Pervious Depression Storage	*DSImp*	*mm*	0	15
% Zero Impervious	*ZeroIC*	*%*	0	30
Max. Infitration Rate	*fMax*	$mm\ hr^{-1}$	5	102
Min. Infiltration Rate	*fMin*	$mm\ hr^{-1}$	0.001	25.40
Infiltration Decay	*fk*	hr^{-1}	1	8
Soil Drying Time	*fDry*	*days*	1	14

3.3.4.2 SWMM Subsurface Flow

The SWMM subsurface flow component has as many as 18 unique parameters per aquifer and subcatchment pair, many of which are interdependent. Behavior of those elements is also dependent on delivery of flow to the subsurface environment from the surface and is thus dependent on climate forcings and surface parameters as well. Given the large number of uncertain and / or empirical parameters used and the added capability for simulating hydrograph recessions and inter-event flow provided, SWMM subsurface modeling was an important priority for the RSA in this work. However, while the RSA is performance based, no commensurable data were collected which could be

used to directly assess performance of SWMM simulated groundwater flow. Instead, I opted to use the SW1 streamflow data to assess model performance as the basis for the RSA.

Flow from the SW1 drainage area was generally not well represented by the SWMM surface runoff model in the various plausible configurations considered. Similarly, rainfall runoff data analysis for the SW1 area suggests that event runoff volumes were only weakly predicted by event rainfall volumes. It is likely that time variant aspects of the agricultural management of this area (a corn / hay rotation over the period of sampling), including plowing, cutting, and variable evapotranspiration demand, accounts for additional variation in the rainfall runoff response. However, I have neither detailed management data for the upland area, nor would such data be easily incorporated into the SWMM conceptual model. As a result, the use of SWMM subsurface flow components to improve event response modeling, particularly for hydrograph recessions, must be viewed in part as a misspecification of algorithm processes to the physical processes they represent. Nonetheless, this application framework provides the best opportunity from the available data to assess simulated subsurface flow sensitivity and may in fact function as an effective conceptual surrogate for the processes being represented.

The objective function used for the subsurface RSA under this framework was the mean absolute percent error (MAPE) between simulated and measured streamflow, calculated as:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{meas_i - sim_i}{meas_i}\right|$$

where *N* is the number of discrete measurements, *meas* is measured flow rate, and *sim* is simulated flow rate. This measure was selected due to its relatively high weighting of lower flows rather than peaks, which is better suited to the assessment of subsurface flow relative to RMSE. I ran a total of 152,000 simulations for 247 days of record (2009-4-1 through 2009-12-4), producing 152,000 simulated hydrographs for comparison with the measured flow record. Parameter ranges for the varied parameters are given in TABLE 10.

TABLE 10. Uniform distribution parameters for the SW1 SWMM parameters varied in the RSA runs. Abbreviations used in the text are included.

Parameter	Abbreviation	Units	Min	Max
Subcatchment				
Area	*Area*	*ha*	42.53	47.01
Width	*Width*	*m*	50	1400
Slope	*Slp*	*%*	0.69	3.63
Pervious n	*nPrv*	*unitless*	0.05	0.45
Pervious Depression Storage	*DSPrv*	*mm*	0	25.4
Max Infiltration Rate	*fMax*	*mm hr^{-1}*	5	76.2
Min Infiltration Rate	*fMin*	*mm hr^{-1}*	0.001	25.4
Infiltration Decay Constant	*fk*	*hr^{-1}*	0.1	8
Soil Drying Time	*fDry*	*days*	0.1	14
Surface Elevation	*SElev*	*m*	0.001	2
Aquifer and Groundwater				
Porosity	*Por*	*frac*	0.2	0.5
Field Capacity	*FC*	*frac*	0.2	0.45
Wilting Point	*WP*	*frac*	0.05	0.25
Sat. Hyd. Conductivity	*KSat*	*mm hr^{-1}*	1	254
Conductivity Slope	*KSlp*	*unitless*	0.1	20
Tension Slope	*TSlp*	*mm*	0.1	20
Upper Evap. Fraction	*UEvap*	*frac*	0	1
Lower Evap. Depth	*LEvap*	*m*	0	1
Lower Groundwater Loss Rate	*GLoss*	*mm hr^{-1}*	0	5
Bottom Elevation	*BElev*	*m*	-2	0.5
Initial Unsat. Zone Moisture Content	*MCInit*	*frac*	0.05	0.45
Groundwater Flow Coeff. (A1)	*A1*	*m sec^{-1}*	0.00001	0.5
Groundwater Flow Exponent (B1)	*B1*	*unitless*	0.4	4

3.3.4.3 Buildup and Washoff

The exponential implementations of the SWMM Bu/Wo models were assessed for sensitivity using the data collected at the East Drain subcatchment. The exponential Bu/Wo algorithms are not highly parametrized, with only two parameters per algorithm per analyte. The parameter ranges used for Bu/Wo RSA runs are included in TABLE 11.

A total of 30,000 model runs were executed, producing 30,000 simulated pollutographs for comparison with the measured load estimates.

TABLE 11. Uniform distribution parameters for the East Drain Buildup and Washoff parameters varied in the RSA runs. Abbreviations used in the text are included.

			TSS		TN		TP	
Parameter	**Abbreviation**	**Units**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
Buildup Coeff	*BuCo*	$kg\ ha^{-1}$	0	40	0	3	0	0.3
Buildup Exp	*BuEx*	day^{-1}	0	6	0	6	0	6
Washoff Coeff	*WoCo*	*unitless*	0	5	0	5	0	5
Washoff Exp	*WoEx*	*unitless*	0	5	0	5	0	5

As with subsurface flow, there is the issue of commensurability between what we measured and what SWMM simulates that must be dealt with before employing a performance-based assessment (such as RSA or calibration). Water quality samples in this study were collected primarily on a flow weighted composite basis, producing a single concentration per analyte per site per sampled storm. In contrast, SWMM computes pollutant concentrations continuously at a user defined temporal scale, as fine as 1 second. Conceptually, the most direct comparison between the measured data and the simulated record would be to extract from the simulated record and then average all instantaneous concentrations corresponding to the instants in time that the autosampler collected an aliquot during a sampled event. Perfect model performance would correspond to a match between the average of those instantaneous simulated concentrations and the lab analyzed composite concentration for that storm. Despite the

conceptual correctness of this approach, I did not use this approach for the following reasons.

First, while a perfect simulation would exactly match the measurements using the approach just described, so would a nearly infinite number of incorrect simulations that just so happened to average to the measured value. While this concern cannot be alleviated in any scenario given how the data were collected, it does undermine the above approach as being singularly correct. A issue second stems from the lack of temporal precision in the water quality sampling. Given the flashiness of both measured and modeled data, the resulting instantaneous simulated concentrations could vary by an order of magnitude within a 2-minute window on an event rising limb. The total time for the ISCO autosampler to purge, rinse, and collect a sample could be in excess of 2 minutes, complicating the selection of a temporally matching instantaneous value from the simulated record. Thus, based on these considerations I ruled out using this approach.

The approach used here was instead based on total loads. For each sampled event in the record I computed a total load for each analyte as a function of cumulative flow and storm composite concentration that was sampled. The comparable simulated quantity was defined analogously as the simulated load (simulated flow times simulated concentration) during that same sample period (including all concentrations predicted at inter-aliquot time steps). Further, to isolate the buildup and washoff models and minimize the load differences solely attributable to hydrologic model error, the buildup and washoff RSA runs were simulated using a calibrated ensemble of hydrologic models

for the East Drain area (i.e., the order that work was carried out differed from how it is summarized here).

A disadvantage to this approach stems from the nature of our composite sampling. By collecting and analyzing a single composite EMC per storm, we have limited the data available to the calibration approach. That is, many different SWMM load predictions can match total measured loads perfectly, while also incorrectly predicting instantaneous concentrations through time. However, this was determined to be the best approach given the available data. Thus, I defined the objective function as the sum of absolute load errors (SAE) over a set of seven sampled events, calculated as:

$$SAE = \sum_{i=1}^{N} |sim_i - meas_i|$$

where *N* is the number of sampled storms considered, *sim* is the modeled load, and *meas* is the measured load.

3.3.5 *Calibration, Validation and Prediction*

My approach to calibration was inspired by previous work on both equifinality and parameter optimization. However, in recognition of the different importance of and bases for the numerous parameters in this SWMM application, a three level classification scheme was adopted to facilitate calibration and uncertainty extrapolation. For many parameters (e.g., percent impervious, drainage area), it is recognized that the feasible

parameter space cannot be constrained based on model performance, due to noncommensurable predictions and measurements, errors in inputs, and errors in model structure. The only way to constrain those parameters is through higher precision field measurement. For other parameters that are immeasurable (i.e., conceptual or empirical), it may be legitimate to optimize (or to constrain via a GLUE approach) values within a defined range, provided the underlying empirical model is applicable to the process being simulated. A third class of parameters can then be construed, which are those parameters to which the model is not sensitive. These insensitive parameters can be handled in any number of ways with little consequence, although it is computationally cheapest to simply fix them. Lastly, it should be stated that while this categorization approach is inherently subjective for many parameters, it does provide a potentially useful framework for dealing with the high dimensional uncertain parameter space typical of SWMM and other watershed model calibrations. Thus, given the sensitivity analysis work described in the previous sections, I proceeded to categorize the SWMM parameters into three groups for the calibration process. These are:

(1) Parameters to which the SWMM models were not sensitive. These parameters can simply be fixed at best estimates during calibration and subsequent modeling with minimal consequence.

(2) Sensitive parameters that are empirical, not directly measurable, or for which a solid basis for estimation does not exist. These are parameters where it may be acceptable to vary values within their conceptual limits toward identifying values that maximize agreement between simulated and observed (i.e., some

feasible values of these parameters can be ruled out as being acceptable simulators of the system).

(3) Sensitive parameters that can be measured or physically estimated and for which useful estimates exist. These parameters may have known uncertainty, but that uncertainty should be extrapolated through the model predictions for conceptual soundness, and would be inappropriate to optimize to 'best' values.

In summary, the goal of the calibration was then to identify values of sensitive calibration parameters (2) that maximize agreement between simulated and observed while concurrently accounting for the estimated uncertainty in the parameters classified as (3). To accomplish this, I developed a MATLAB implementation of an evolution strategies (ES) search algorithm (described below) to be applied to the SWMM calibration parameters (identified as (2) above), concurrent with a Monte Carlo sampling implementation for the sensitive, non-calibration parameters (identified as (3) above). This approach was designed to avoid the problem described by A. Stirling, as presented in Saltelli (2002), specifically, that arbitrarily restricting the input parameter space produces deceptively precise model output.

For the ES algorithm we used a canonical (μ, λ) implementation as described by Eiben and Smith (2007). In brief, from an initial parent population of parameter sets of size μ, a set of offspring parameter sets of size λ (typically $7*\mu$) was created through a combination of mutation and recombination of the parent population. Each of the λ offspring were evaluated as a SWMM realization and ranked based on fitness (i.e.,

objective function value) with the best μ from the offspring population retained as parents for the next generation. The creation of the λ offspring was biased toward Gaussian mutation (80%), as opposed to recombination (20%, implemented as a combination of discrete and intermediary recombination following Eiben and Smith (2007)), with mutation step sizes concurrently evolved with the SWMM parameters such that the intra-generational parameter search space can change during an ES run. This co-evolution of the mutation step size, or self-adaptivity, allows the algorithm to tune itself in this one respect, lowering the possibility of algorithm failure or inefficiency due to user misspecification. This feature, coupled with ES's natural handing of continuous variables parameter space made it a suitable candidate for this work.

An important variation on this approach, as compared to traditional optimization, is the concurrent Monte Carlo sampling of physically based parameters that were not being optimized. This sampling introduces noise into the search, which would be expected to make the search task for the ES algorithm harder. Arnold and Beyer (2003) have examined the efficiency of five search algorithms on a theoretical optimization test case where Gaussian noise had been incorporated into the objective function to explore the implications on search efficiency. Of the algorithms tested, the ES (using a derandomized ES variant) was the most efficient strategy given a high dimensional parameter space and large amounts of noise in the objective function. Thus, I concluded that the ES strategy was a good choice for the nature of this case study.

I typically used a population size of two times the number of varied SWMM parameters (evolving and Monte Carlo sampling), and created subsequent generations

sized as seven times the parent population through a combination of mutation and recombination. I ran the ES algorithm for a fixed number of generations, given that the concurrent Monte Carlo sampling would be expected to prevent the populations from converging on single parameter sets. For each grouping of parameters to be evolved, I ran a total of four ES runs at 30 generations per run. For each collection of ES runs, I then retained as the calibrated set the last generation of each of the four runs, plus the best parameter sets in all other generations and runs (excluding the last generations) to create a total retained population of parameter sets equal to 10% of all parent parameter sets selected over the runs. Many of these values (e.g., population size, number of runs, number of generations per run) were subjectively set based on initial algorithm testing in which I sought to balance computational run times with the need for proper functioning of the algorithm.

Similar to the approach used in the RSA, I divided the parameters to be calibrated and the data available for doing so. The major groupings included (1) East and West Drain surface hydrology, (2) East and West Drain Bu/Wo algorithms, and (3) full watershed subsurface hydrology. Given the available measured data, I assumed these components were sufficiently isolated for purposes of independent calibrations, and thus did not warrant a (concurrent) multi-objective calibration approach. Results of calibrating these components will provide useful data for parameterizing the entirety of the drainage area model using locally derived parameter values.

3.3.5.1 East Drain and West Drain Surface Hydrology

Flow gaging within the East and West Drain outfalls in 2008-09 provided the basis for the neighborhood subcatchment hydrologic calibrations. The approach used was to calibrate each of these two drainage areas using one year of record, with the second year of record used for validation. For the fitness function, I used the ratio of root mean squared error to the standard deviation of measured flow (RSR), calculated as:

$$RSR = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(meas_i - sim_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(meas_i - meas_{mean})^2}}$$

where N is the number of discrete measurements, *meas* is measured flow rate, *sim* is simulated flow rate, and $meas_{mean}$ is the mean of the measured flow record. The resulting parameter values from these calibration runs were then pooled, both within and between sites, with the calibrated values taken to represent those parameters' values for the entire neighborhood. For example, values of *nImp* that were found to produce good agreement between simulated and measured within the two instrumented subcatchments were taken to be good values of *nImp* for other parts of the neighborhoods that were not discretely assessed.

The results from the calibration runs included hundreds of unique calibration parameter sets that maximized hydrologic fitness under the concurrent Monte Carlo sampling of uncertain parameters. However, these hundreds of parameter sets had to then be paired with high dimensional non-calibration Monte Carlo parameter space to

fully represent the calibrated results. Thus, the last calibration step was to resample from the hundreds of discrete calibration parameter sets while continuing to Monte Carlo sample the uncertain parameter space to generate a large number of realizations of the calibrated model. To do so, I ran an additional 15,000 simulations (i.e., the calibration blowout) using this approach to more fully characterize the calibrated model's performance over the full accepted parameter space. Lastly, for validation I simply ran the same 15,000 blowout parameter sets over the alternate year of available flow record and computed the hydrologic fitness from the resulting flow records.

Finally, for my assessment of the comparability of subcatchment *Width* parameters specified by the SWMM User's Manual ($Width_{SUM}$) versus the Guo Method ($Width_{Guo}$), I took two additional steps. First, I computed large samples of $Width_{Guo}$ and Slp_{Guo} by randomly sampling the needed inputs from uniform distributions representing collector channel lengths, area skewness coefficients, and slopes for both the East Drain and West Drain subcatchment areas. I was then able to compare evolved $Width_{SUM}$ values calibrated by the ES with the range of values computed by application of the calibration-free Guo Method ($Width_{Guo}$). Second, I ran an additional batch of ES calibration runs where the Guo Method was employed to treat *Width* and *Slp* as uncertain parameters as opposed to calibration and uncertain parameters, respectively, thus eliminating *Width* as a calibration parameter. The purpose of these runs was to assess whether the ES would evolve different values of other evolving parameters (e.g., Manning's coefficients, depression storage) when alternately coevolved with $Width_{SUM}$ versus evolving while Monte Carlo sampling of values for $Width_{Guo}$ and Slp_{Guo}.

3.3.5.2 East Drain and West Drain Water Quality

We collected event mean concentration (EMC) and discrete concentration data at all four locations (East Drain, West Drain, SW1,and SW2). However, water quality calibration was not attempted at the upper (SW1) and lower (SW2) monitoring stations because I did not have a suitable hydrologic model for the SW1 site nor did I have a suitable model for channel processes relevant to the SW2 site. The drains, in contrast, represented two homogeneous residential closed system outfalls, with impervious surface washoff and pervious lawn run-on as the primary sources of sediment and nutrients. Thus, I considered them to be good candidates for SWMM's simple buildup and washoff algorithms and these sites formed the basis of the water quality modeling.

The data available for calibration and validation included 17 composite sampled storms at for the West Drain and 11 composite sampled storms for the East Drain. Storm composites consisted of between 4 and 38 aliquots per storm, with a mean of 22 aliquots per sampled storm. Additionally, for a single storm we successfully deployed the autosamplers for discrete storm sampling using a 24 bottle sampling kit. This enabled us to analyze for changes in constituent concentrations through the storm event. Samples from the discretely sampled storm contained valuable data for Bu/Wo calibration and validation that are not included in the composite EMC data. Because many parameterizations can closely match event loads but miss the intra-event dynamics (e.g., improbably high peak concentrations), I found I had to use these discrete concentration data to constrain the calibration. Further, in addition to providing instantaneous as opposed to event mean concentrations, the sampling during the discretely sampled storm

actually occurred over two distinct event hydrographs, with event peaks separated by 4.5 hours. Thus, in addition to providing key data to constrain instantaneous constituent concentrations, these data provided the best basis to estimate the inter-event build up rates, since the other sampled storms were generally too far apart in time to be sensitive to this value. Similarly, the discrete sampled storm offered the strongest opportunity to validate the Bu/Wo model, given the event concentrations through time in addition to total load estimates. Thus, I elected to split the discretely sampled storm, using the most informative second event peak for calibration, while reserving the first pulse for validation. I concluded that this was the best compromise given the modeling objectives and available data, and allowed for both calibration and validation against discrete concentrations.

For the composite sampled storms, I approximately equally split the data into calibration and validation sets for each of the sites, while trying to ensure a representative split (e.g., small storms vs large storms). For the fitness to be minimized, I computed the sum of the absolute errors (SAE) of calibration event loads as:

$$SAE = \sum_{i=1}^{N} |sim\ load_i - meas\ load_i|$$

where *N* is the number of storm events, *sim load* is the SWMM predicted load for a given analyte, and *meas load* is the measured load estimate for a given analyte. An important feature of this fitness measure is that overestimation on some events cannot not be

compensated for by underestimation on others, rather, the best value is one that exactly matches the loads for all events. Lastly, to make use of the discrete concentration data that were collected, I computed a penalty as two times the sum of the total absolute load error if the maximum instantaneous concentration for the discrete calibration peak was more than three times the maximum observed concentration for that event. This was implemented in response to observations during initial calibration efforts, where the evolved solutions tended to have very high concentrations for brief periods that matched the total load estimates, but that conflicted with our data and expectations for intra-event dynamics.

Concurrent with the evolution of Bu/Wo parameters, the underlying hydrologic models were subject to the hydrologic model blowout sampling as described in Section 3.3.5.1. This ensured that the evolved Bu/Wo model parameters would be the best that could be identified with respect to the full range of hydrologic prediction scenarios to be used for these areas. After running the ES calibrations, I used the same steps here as for the East Drain and West Drain hydrologic calibrations, pooling and then resampling the calibration parameters while concurrently resampling the uncertain parameters from their assumed uniform distributions. I conducted a total of 10,000 calibration blowout simulations to more fully evaluate the calibrated model in this manner, and subsequently ran the same 10,000 parameter sets to evaluate the calibrated parametrization for the subset of validation events.

3.3.5.3 Full Model Pervious Area and Subsurface Hydrology

I initially planned to calibrate the subsurface hydrology parameters using the SW1 drainage area, which has a high level of pervious cover and was found to be poorly represented by the SWMM surface runoff model. Given the lack of extensive impervious cover or engineered drainage infrastructure in this area, it provided the best opportunity to isolate the role of SWMM's pervious area and subsurface processes on runoff responses within our study area. However, I found this upland agricultural area draining to the SW1 station to be a challenge to acceptably represent using SWMM's routines.

While I was able to match specific events, and even seasonal sets of events with plausible SWMM model structures and parametrizations, no models were identified which could adequately represent the consecutive seasons within a year with an acceptable level of model performance. In particular, the drying of the intermittent SW1 channel over the summer and resumption of flow in fall were challenging to capture, in part due to lack of specific knowledge about or direct SWMM capability to represent agricultural land surface changes in the subcatchment over time. Further, the best model structures and parameterizations that were identified routed most flow through subsurface pathways, including for larger events. While this improved the hydrograph fit (although it remained relatively poor), it precluded modeling water quality as those events would deliver zero modeled washoff load, and contradicted our hydrologic understanding of the system in which the clay soils would primarily produce runoff through infiltration excess and saturation excess surface flow, as opposed to interflow. Thus, in subsequent modeling I simply routed the measured SW1 flow record into the upper most node of the

SWMM model, while using SWMM subcatchments and conveyance infrastructure to model the lower 68.3 hectares of neighborhood and other watershed contributing area. This resulted in a flow record at the model outlet (SW2) consisting of the SW1 measured flow record routed through the modeled stream channel and culverts to which I added the SWMM-modeled runoff from the neighborhood and the remainder of the lower drainage area.

To proceed with calibration of the subsurface and pervious area parameters given the poor results at SW1, I opted instead to use the lower watershed stream response (SW2). Preliminary assessments suggested that modeled runoff from the impervious surfaces combined with the routed SW1 flow record could predict event peaks at SW2 with acceptable accuracy, but that inter-event periods and hydrograph recessions in particular could be improved. Thus, I aimed to use the receding and low flow portions of the SW2 streamflow record as the basis for calibrating the pervious area and subsurface parameters. To do so, I used the $MAPE_{Low}$ as the fitness function, which was the previously defined *MAPE* function computed over only the lower 75% of SW2 flows. The longest year of record, 2009, was selected for calibration while the 2007-08 years were used for validation.

I used the same steps here as for the East Drain and West Drain calibrations, pooling and then resampling the calibration parameters while concurrently resampling the uncertain parameters from their assumed uniform distributions. I conducted a total of 40,000 calibration blowout simulations to more fully evaluate the calibrated model, and

subsequently ran the same 40,000 parameter sets to evaluate the calibrated parametrization over the validation period.

3.4 Results and Discussion

3.4.1 *Sensitivity Analysis*

The results of my sequential application of Freer et al. (1996) RSA methodology to SWMM's hydrology and water quality, as described in Methods, are included in the following sections.

3.4.1.1 Surface Hydrology

RSA plots calculated from 40,000 East Drain surface hydrology runs are plotted in FIGURE 13, using RMSE of East Drain outflow as the performance measure. Sensitivity among parameters varied greatly, with *Width, DSImp, nImp,* and *ZeroIC* each showing moderate degrees of sensitivity. A slight degree of sensitivity was seen in *Slp* and *ICPct*, while pervious area parameters (e.g., *DSPrv*, *nPrv*, and the Horton infiltration parameters) exhibited no sensitivity in this context. These results generally conformed with my expectations for this system given its physical attributes. The high proportion of connected impervious surface within the East Drain, collected by curb and gutter closed drainage, was assumed to be the dominant driver of runoff dynamics in the subcatchment based on the close correspondence between measured runoff and rainfall seen in the measured record.

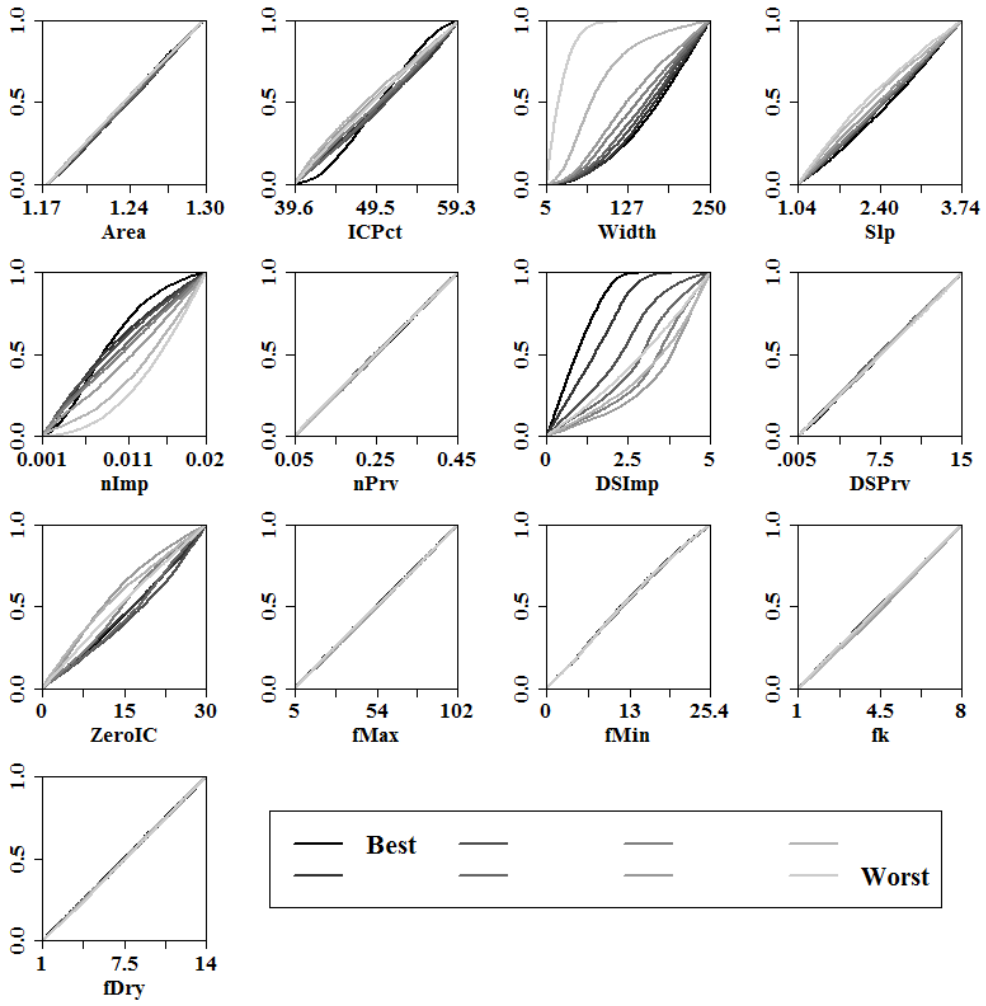

FIGURE 13. RSA plots for 40,000 uniform random simulations of the East Drain subcatchment, using RSR as the goodness of fit measure. The included parameters are selected SWMM surface runoff and Modified Horton infiltration parameters. Model goodness of fit is the RSR of East Drain outflow. SWMM parameter abbreviations as defined in TABLE 9.

3.4.1.2 Surface and Subsurface SW1 Hydrology

Surface and subsurface hydrology RSA plots from 152,000 simulations of the SW1 catchment are plotted in FIGURE 14 and FIGURE 15, assessed on the basis of MAPE. (Note that while the poor calibration results for SW1 precluded predictive

modeling and analysis for this area, this did not diminish its usefulness for assessing the sensitivity of SWMM pervious area flow predictions to pervious area and subsurface parameters.) I considered the surface and subsurface parameters concurrently within this RSA batch due to the role that the surface parameterization plays in delivery of water to the subsurface reservoir. Additionally, to assess meaningfully the parameters used, I calculated several post-hoc SWMM parameters to better illustrate the sensitivity dynamics shown in FIGURE 14 and FIGURE 15. For example, within the SWMM conceptual model the subsurface storage volume is defined by *Por,* multiplied over a subsurface depth calculated as *SElev-BElev*, which can then be positioned at different absolute elevations relative to the receiving node. Thus, the model might be more sensitive to *SElev-BElev* than to either parameter individually given the physical properties they represent. By varying the parameters independently, I was able to assess sensitivity of interdependent parameters independently and combine parameters for joint consideration where it was potentially useful to do so.

There was a wide range in sensitivity, from a high degree of separation between best and worst bins (e.g., *DSPrv*, *Por*), to no discernable separation among bins (e.g., *KSat*, *KSlope*). Among the surface parameters, *DSPrv* was found to be most sensitive. Since this parameter specifies the depth of ponding that occurs prior to initiation of overland flow, it can affect both the water available for infiltration and for surface evaporation. Sensitivity was also evident in the infiltration parameters, in contrast to what was found in the East Drain subcatchment RSA. Given the absence of impervious surface, the partitioning of incident rainfall between pervious surface and subsurface

pathways appeared to play a strong role in defining the runoff dynamics of the system, highlighting the degree to which even a global SA is dependent on the model structure being assessed.

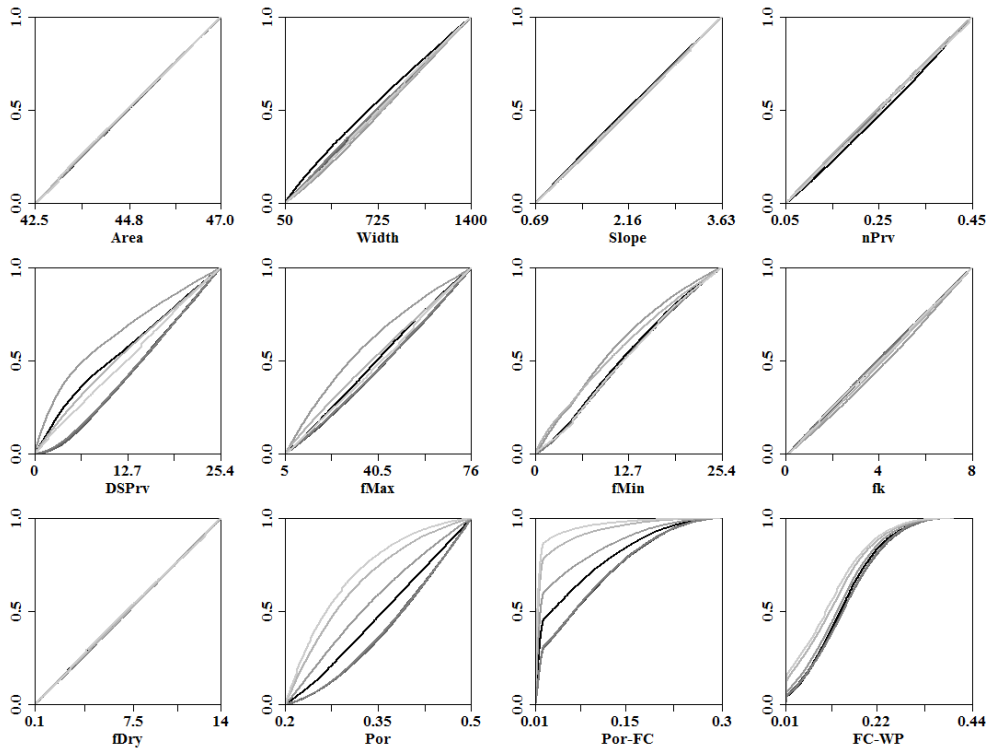

FIGURE 14. RSA plots for 152,000 uniform random simulations of the SW1 subcatchment, using MAPE as the goodness of fit measure. The included parameters are selected SWMM surface runoff, Modified Horton infiltration, and subsurface flow parameters. SWMM parameter abbreviations as defined in TABLE 10.

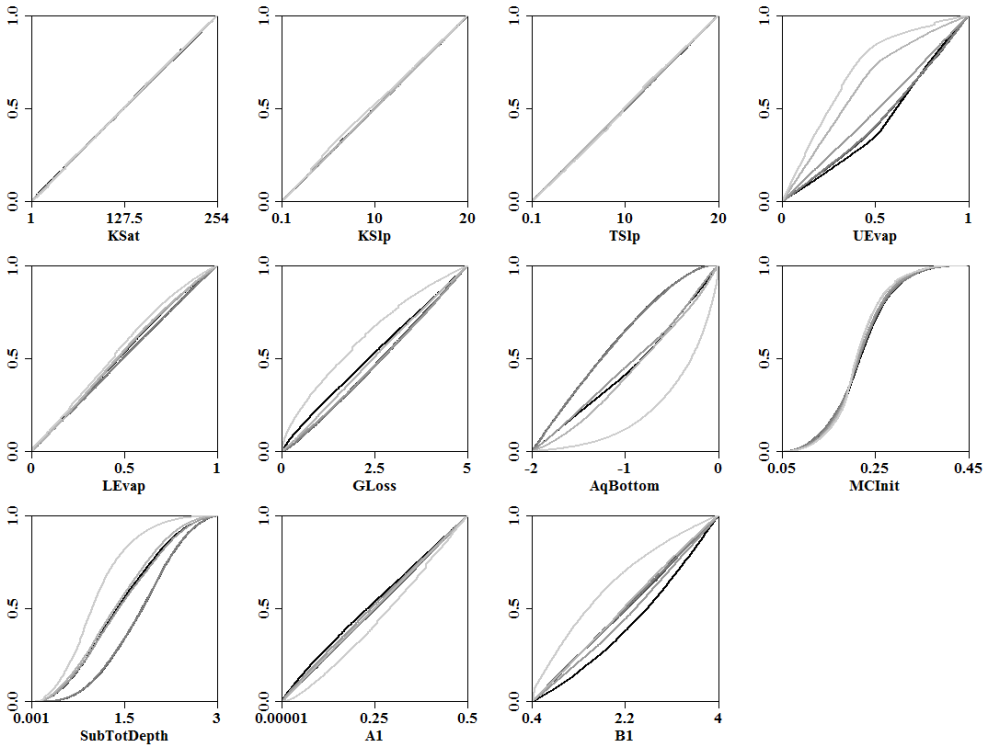

FIGURE 15. RSA plots for 152,000 uniform random simulations of the SW1 subcatchment, using MAPE as the goodness of fit measure. The included parameters are selected SWMM subsurface flow parameters. SWMM parameter abbreviations as defined in TABLE 10.

Among subsurface parameters, sensitivity was generally greatest for those directly affecting evapotranspiration potential and total subsurface storage. For example, *UEvap* showed a moderate degree of sensitivity, which controls the fraction of ET demand not met by surface ponded water that can be met from available water within the unsaturated zone. Given the primarily vegetated nature of the subcatchment, parameters controlling ET would be expected to influence runoff dynamics. Additionally, numerous parameters affecting total subsurface storage available were found to be sensitive within this RSA, including several of the post-hoc calculated parameters. For example, *Por* was found to be sensitive, as was the post-hoc variable *Por-FC*, which represents the storage volume

that must fill before flow to lower groundwater and thus outflow can occur. In contrast, *FC-WP*, or the fraction of subsurface storage space subject only to ET demand was not, suggesting that the free draining void space affects runoff dynamics more strongly than the fraction emptying via ET alone. Similarly, both *SubTotDepth* (calculated as *SElev* - *BElev*) and *AqDepth* (representing the elevation of the bottom of the aquifer relative to the receiving node) were found to be sensitive, defining the total storage depth and depth of storage below the receiving node (interpretable as the channel bed), respectively. Combined, these parameters indicate the total volume of storage and its positioning relative to the receiving node play a strong role in simulated runoff dynamics of this system.

In contrast, parameters controlling the percolation rate between upper (unsaturated) and lower (saturated) groundwater, *KSat*, *KSlp*, and *TSlp* were not sensitive, suggesting that the magnitude of this internal flux was not affecting model performance. *KSlp*, it should be noted, was observed in manual calibration to affect the shape (peakiness) of the groundwater flow hydrograph. I hypothesize that given an isolated groundwater flow record to calibrate against, this parameter could be very sensitive. However, for our tributary flow record, a mixture of response modes makes is such that exact shape of groundwater flow hydrograph, when superimposed on surface hydrographs, would not necessarily strongly affect the MAPE of streamflow. Thus, to the extent that *KSlp* does alter flow dynamics, the small adjustments to hydrograph shape are likely dwarfed in comparison with the effects of other parameters controlling whether or not flow occurs.

Other parameters found to be sensitive in this RSA were the groundwater flow equation parameters *B1*, and to a lesser extent *A1*. These parameters directly control the groundwater flow rate that occurs as a function of the depth of groundwater above the node elevation. Additional groundwater parameters can be specified to simulate other groundwater flow dynamics (e.g., streamflow from the channel into groundwater), however I chose not to include this additional complexity in our model. Given the conceptual application of SWMM groundwater (i.e., based on a need for a slow reservoir to supplement SWMM surface hydrology), these parameters did not appear to be justified. However, they could be important to the structure of other SWMM models and their degree of sensitivity remains unassessed.

3.4.1.3 Buildup and Washoff Models

Buildup and washoff model sensitivity is shown for TSS in FIGURE 16 on the basis of aggregated load error; similar results were found for TN and TP (not shown). For all three of the analytes considered, sensitivity was greatest for buildup parameters, in particular *BuCo*. *BuCo* controls the maximum surface buildup of pollutant between storm events, and as such can uniquely constrain simulated loads among the set of Bu/Wo parameters. The buildup model exponent (*BuEx*), which controls the rate at which surface loads available for washoff return to their maximum level following a washoff event, showed a lesser degree of sensitivity. It is worth noting that among the set of seven sampled storms used in the RSA, only one storm had an antecedent dry period of less than 3 days, with an average of ~6 days. Thus, a wide range of *BuEx* values would

allow inter event surface loads to reach maximum levels (i.e., at or near *BuCo*) under these conditions. Nonetheless, aggregate load error was sensitive to this parameter, and I suspect that it could be even more sensitive given an observed data set with greater inclusion of back to back storm dynamics. In comparison, washoff parameters were relatively insensitive. Given that the RSA objective function was aggregate load error over a series of events, a higher degree of sensitivity in the buildup model is not surprising. Washoff parameters would likely show greater sensitivity at a more granular level (e.g., modeling instantaneous peak concentrations) than the data from this study (i.e., composite samples) would support due to the fact that a wide range of washoff-driven pollutographs could produce the same total load. Thus, the lack of washoff parameter sensitivity identified in this study is unlikely to transfer to other modeling applications using higher resolution data.

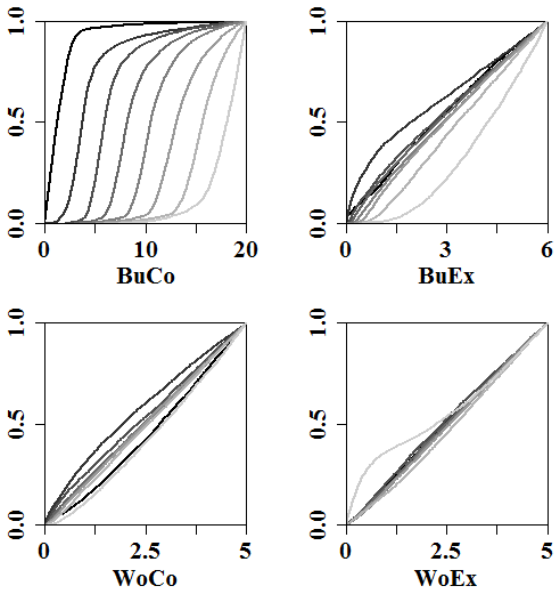

FIGURE 16. RSA plots for 30,000 uniform random simulations of the East Drain subcatchment's buildup and washoff TSS models, using aggregated load error over seven storm events as the goodness of fit measure. SWMM parameter abbreviations as defined in TABLE 11.

In related work, Shaw et al. (2010) have found particulate washoff loads to be insensitive to time between events (i.e., buildup rates), and have shown that a constant available mass (CAM) model can provide similar predictive capability with fewer parameters. The dominance of *BuCo* in our own work is generally in agreement with the findings of Shaw et al. (2010), although the moderate sensitivity of *BuEx* and specifically the tendency of best performance to be associated with small values of *BuEx* are not. However, as already discussed the storms used in our RSA generally had multi day antecedent periods such that a conclusive picture of the role of *BuEx* cannot be drawn.

3.4.1.4 Classification

Based on the results of the RSA, the included SWMM parameters were categorized for the subsequent modeling (TABLE 12; TABLE 17). Parameters which were not sensitive were fixed at base parameterization values for all subsequent modeling (i.e., calibration and prediction). Physical parameters that were sensitive included *ICPct, Width, Slp, fMax* and *fMin*. These parameters were handled via Monte Carlo simulation throughout, whereby parameter values for any simulation were randomly sampled from assigned distributions. A large number of parameters that were found to be sensitive are either inherently or functionally conceptual. These parameters were interpreted to be calibration parameters, such that any values that maximize model fit from within predefined feasible ranges were acceptable. This grouping included the impervious surface characteristics (i.e., *nImp*, *DSImp*, *ZeroIC*), all of the buildup and washoff parameters, and many of the subsurface parameters (TABLE 17). Combined, this classification framework, using the RSA results presented in FIGURE 13, FIGURE 14, FIGURE 15, and FIGURE 16 as well as a degree of subjective judgement, was used to determine how parameters were handled throughout subsequent modeling.

3.4.2 *Calibration and Validation*

I applied a combined Evolution Strategies (ES) and simple Monte Carlo (MC) sampling approach to the calibration of selected SWMM parameters, based on the classification results of the sensitivity analysis work. The spatial, temporal and conceptual discretization of this calibration work is described in the following sections.

3.4.2.1 East Drain and West Drain Hydrology

The classification of parameters for calibration (i.e., fixed, MC sampled, or ES calibrated) is summarized in TABLE 12 with two aspects worth noting. The classification of individual parameters was not solely informed by the East Drain RSA results (FIGURE 13). For example, pervious area parameters were not found to be sensitive in the East Drain RSA, however they were found to be sensitive in the SW1 RSA (FIGURE 14). This suggests that these parameters are important to how incident rainfall is partitioned between surface storage, ET, and subsurface storage, but not in a way that affected outflow to the East Drain over the period of record used in the RSA. Thus, despite the lack of sensitivity in East Drain RSA, parameters *DSPrv*, *fMax*, and *fMin* were subject to Monte Carlo sampling during the calibration. This ensures that calibrated values from the East Drain and West Drain are not only 'best' given particular fixed values of *DSPrv*, *fMax*, and *fMin*, but can be considered 'best' given the broader feasible parameter space. The approach used for the other parameters listed in TABLE 12 follows directly from the results of the East Drain RSA.

TABLE 12. Treatment of parameters within the East Drain and West Drain calibrations. Parameter units and ranges are listed in TABLE 9. All sampling distributions defined as uniform.

Parameter	Treatment
Area	Fixed
ICPct	MC
Width	Calib/MC
Slp	MC
nImp	Calibrate
nPrv	Fixed
DSImp	Calibrate
DSPrv	MC
ZeroIC	Calibrate
fMax	MC
fMin	MC
fK	Fixed
fDry	Fixed

The East Drain and West Drain subcatchments were calibrated independently, with 2009 used for the West Drain, and 2008 for the East Drain, and with alternate years used for model validation (FIGURE 17). I used a population size of two times the number of varying parameters (ES and MC), and ran four 30 generation ES runs for each calibration site / year. During each ES calibration run, the parameters identified as 'MC' in TABLE 12 were randomly sampled from uniform distributions, while model performance was optimized through evolution of the selected 'ES' parameters on the basis of the RSR fitness function. The results of the four calibration runs from each site were then pooled and resampled for a 'blowout' over the feasible MC parameter space, as described in Section 3.3.5.1. The 95th percentiles of simulated flow resulting from this

approach are plotted with the observed flow series' for both calibration and validation periods at the East Drain and West Drain sites in FIGURE 17.

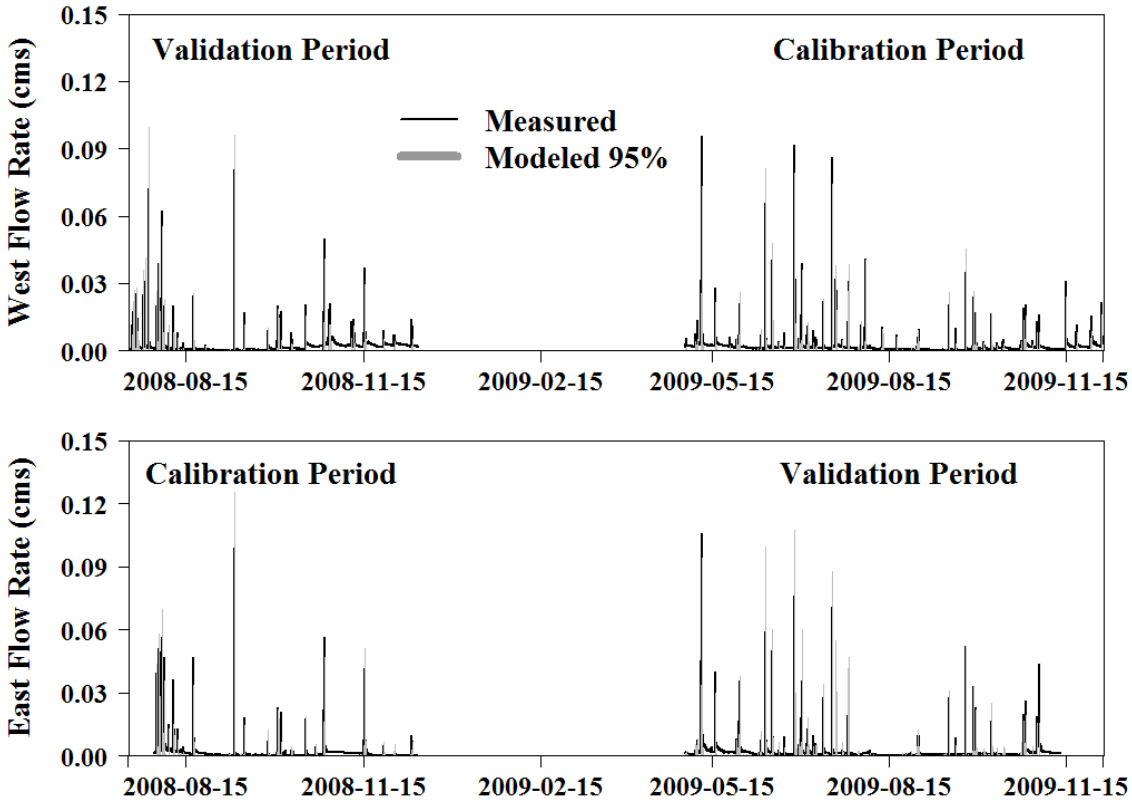

FIGURE 17. Measured and modeled flow series from the East and West Drains for the calibration and validation periods. Measured data (black lines) overlay grey shading of the 95% range of modeled flows from 15,000 samples of the calibrated parameterization.

Calculated RSR fitness values from the drain calibration and validation years are summarized in TABLE 13. Fitness was better in the calibration period than in validation period for both sites, though the difference was greater at the East Drain site. Overall, the best fitness values were somewhat poor, with RSRs indicating that the RMSE was relatively high as compared to standard deviation of the measured data. Moriasi et al.

(2007), in a review of acceptable model performance thresholds, reported that RSR values greater than 0.7 can be considered unsatisfactory based on monthly flow data. However, they also state that relaxed criteria may be used for higher resolution data such as daily flow. The measured and model data for the drain subcatchments are at a 1-minute time step, which is a finer temporal resolution than was used or explicitly considered in either the case study presented or the prior studies reviewed by Moriasi et al. (2007). Thus, we conclude that our relatively high (poor) fitness values are acceptable.

TABLE 13. RSR fitness measure for the best 10% of ES calibration runs and from 15,000 samples of the calibrated parameterization.

		ES Calibration Runs RSR		15,000 MC RSR	
	Type	**Best**	**Mean**	**Best**	**Mean**
East 2008	Calibration	0.708	0.732	0.695	0.753
East 2009	Validation	----	----	0.946	0.986
West 2008	Validation	----	----	0.975	1.017
West 2009	Calibration	0.963	0.994	0.962	1.010

Despite the relatively poor RSR values in calibration and validation, I concluded that the visual hydrograph fit was generally acceptable. The model successfully reproduced the flashy dynamic in the storm drain record, and over and under predicted peaks through the years of record without a substantial bias. Based on visual assessment, the worst model performance occurred in the event recessions and inter-event periods, which I attribute in part to measurement error rather than prediction error. The sensors were 18.3 mm cylindrical pressure transducers, affixed parallel to the direction of flow in

ISCO scissor rings, at the invert of the pipes. Thus, flows at shallow depth were likely subject to greater error due to lack of complete submergence of the sensor and water surface anomalies created by flow over the sensor. A second issue introducing measurement error in the drain records was the tendency for organic matter (e.g., leaves) to catch on the sensor. This material had to be cleared from the sensor periodically. Since flow was calculated from measured depth using Manning's equation, any localized ponding near the sensor would lead to an overestimation of instantaneous flow rates. Lastly, rainfall data were collected primarily at a 5-minute interval via tipping bucket, while drain flow data were collected at a 1-mintue interval. The contributing drainage areas were very flashy with event time-to-peak values of less than 10 minutes in some cases. Thus, small errors in timing, or artefacts of rainfall binning likely introduced additional error into calculations of goodness of fit of the models. Ultimately, we determined that the drain models provided sufficient accuracy for modeling the remainder of the storm drained neighborhood, given the measurement error just discussed.

Results of the parameter evolution as it occurred in the calibration of the East Drain and West Drains subcatchments are shown in FIGURE 18 for one ES run of the West Drain in 2009. For all four variables included in the ES, parameter values converged to similar values over the course of the runs, concurrent with improved model fitness. In the plotted run, it can be seen that both *Width* and *nImp* evolved to best values within the middle of their defined parameter ranges. *DSImp* and *ZeroIC*, in contrast evolved to their ranges lower and upper bounds, respectively, suggesting improved

fitness could be found outside of that range. The *DSImp* and *ZeroIC* parameters are empirical / conceptual, and the parameter limits they evolved to are not physically bounded but rather user specified for this application. However, given that these constraints were based on what was judged to be realistic for the model representation a priori, I did not adjust allowable parameter ranges based the ES results. Loosening the constraints on *DSImp* and *ZeroIC* would have the effect of lowering the impervious surface storage potential, thereby creating additional runoff. It can therefore be inferred that improved model performance, as measured by RSR over the periods of calibration, can be achieved by generating more runoff. However, I suspect this is more attributable to measurement error and model structural errors, as opposed to misspecification in the parameter space boundaries.

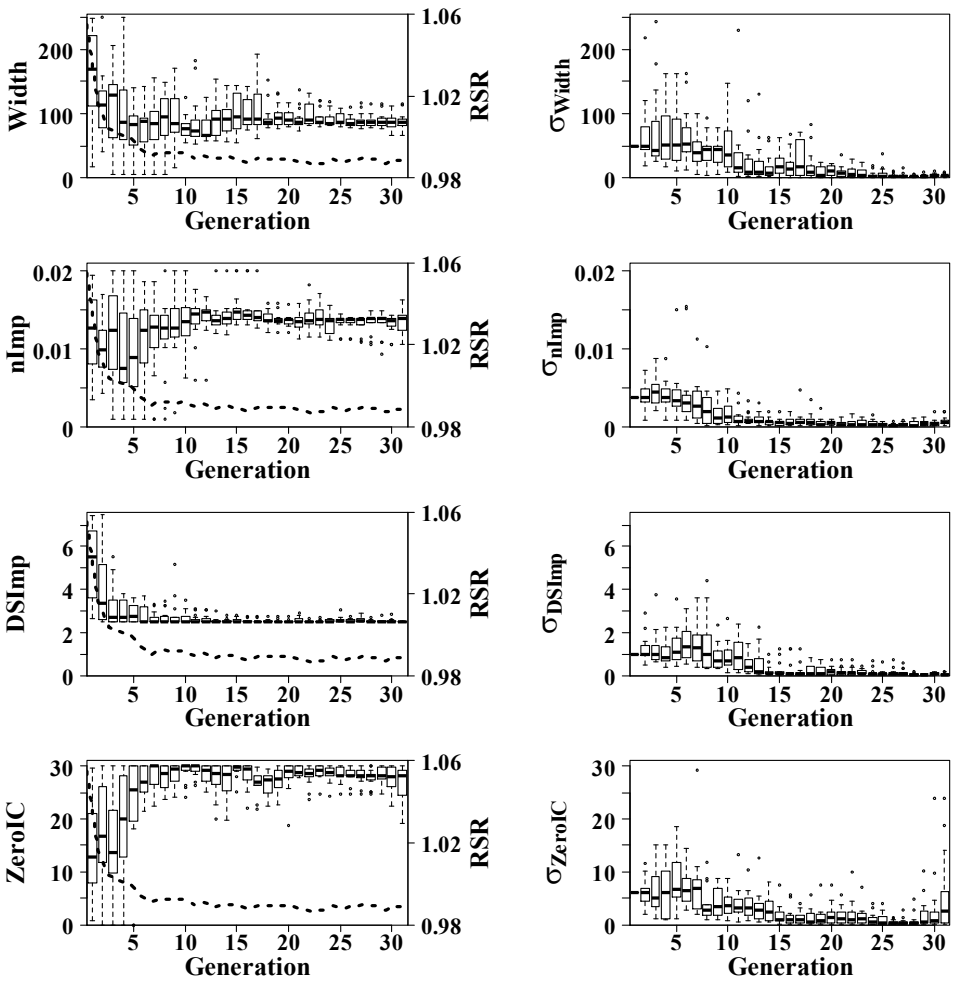

FIGURE 18. Evolution of surface hydrology parameters from a single run of the West Drain 2009 site. Box plots show inter-quartile ranges (IQR) with a median center line, whiskers denoting the largest value less than the 75th percentile value plus 1.5 times the IQR and less than the 25th percentile minus 1.5 times the IQR, and points to show values outside of those ranges. Evolution of SWMM parameters is shown at left, with corresponding mutation step size evolution (σ) at right.

Collectively, from FIGURE 18 it can be seen that over successive generations, the model performance improved, the SWMM parameter values converged, and the search strategy parameters converged toward minimum acceptable levels, even given the noise introduced by the MC sampling of other parameters. The evolution of the mutation step

sizes (i.e., the sigmas) is notable, as it has the functional impact of narrowing the parameter search space over the run as fitness improved. This shows that the self-adaptive aspect of the ES algorithm functioned as designed in the case shown, as it generally did in the East Drain and West Drain calibration runs.

3.4.2.2 Width

After evolving values for the parameter *Width*, I compared the resulting values with those computed by applying Monte Carlo (MC) sampling to the Guo Method approach. The range of *Width* and *Slp* estimates, and the ranges of parameters used in computing those estimates are given in TABLE 14. The SWMM Documentation approach uses *Area* and uncertain estimates of overland flow distances for a subcatchment to define *Width*, with *Slp* estimated independently via available topographic data. Guo's Method, in contrast, estimates *Width* and *Slp* as functions of drainage area and uncertain estimates of topographic slope, collector channel length, and area skewness coefficients. TABLE 14 shows that the range of calculated *Width* values under the Guo Method is considerably narrower than the range of values calculated with the SWMM Methodology. Thus, on this condition alone the use of Guo's Method could be justified due to the a priori reduction in the feasible parameter space for this sensitive parameter.

TABLE 14. Upper and lower values for East Drain and West Drain *Width*, calculated using the SWMM Documentation and Guo Methods. Also included are the subcatchment parameters used in computing those estimates. 'Z' is the area skewness coefficient, following the work of Guo and Urbonas (2009).

	Units	East	West
Subcatchment Area	ha	1.24	1.61
SWMM Guidance			
Overland Flow Dist. (Upper)	m	304	274
Overland Flow Dist. (Lower)	m	58	51
Width 1 (Upper)	m	214	316
Width 2 (Lower)	m	41	59
Mean Slope	%	2.39	2.77
Slope St. Dev.	%	1.35	1.81
Guo Method			
'Z' (Upper)	frac.	0.8	0.85
'Z' (Lower)	frac.	0.6	0.65
Collecter Length (Upper)	m	85	230
Collecter Length (Lower)	m	55	155
Guo Width (Upper)	m	222	184
Guo Width (Lower)	m	163	100
Guo Slope (Upper)	%	0.845	4.03
Guo Slope (Lower)	%	0.104	0.58

To specifically assess the Guo Method's suitability in our case, I next compared ES calibrated *Width* values, optimized from the SWMM Methodology search space, with the ranges of values produced by the Guo Method. As shown in FIGURE 19, the ES calibrated values from both the East Drain and West Drain subcatchments correspond closely with the distributions of values computed via Guo's Method. This suggests that for these two cases, simply applying a MC implementation of Guo's Method produces

the same values as would be achieved through the ES calibration. This result, combined with previous validation work by Guo et al. (2012), provides a defensible basis for the use of this approach to estimate *Width* in subsequent modeling where discrete subcatchment calibration data do not exist.

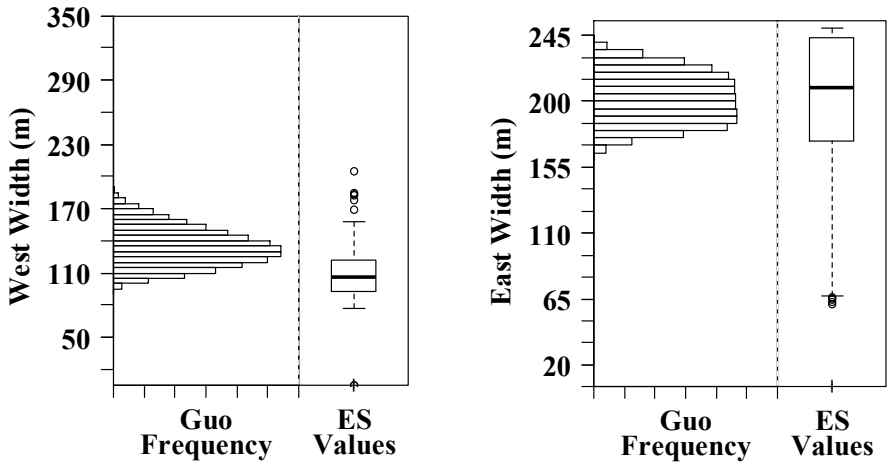

FIGURE 19. A comparison of ES calibrated and Guo Method MC sampled values of Width for the East Drain and West Drain subcatchments.

As a final assessment on the Guo Method, I reran the ES calibration for the East Drain and West Drain areas using a MC Guo Method approach for both *Width* and *Slp*, rather than evolutionary calibration approach to *Width* previously employed. The goal was to determine whether or not other evolving parameters (e.g., *nImp*, *DSImp*) would evolve the same values regardless of the approach used for *Width*. The results of this exercise were indistinguishable (not shown) between the two approaches, further confirming that the Guo Method can be employed as an alternative to the SWMM User's

Manual approach, with the advantage of being calibration independent, with little if any adverse consequence.

3.4.2.3 East Drain and West Drain Water Quality

The storm sampling data available for East Drain and West Drain water quality calibration and validation are summarized in TABLE 15. The storm sampling data set was split for calibration and validation yielding six and nine calibration storms for the East and West Drains, respectively, and seven and nine storms for validation in the East and West Drains, respectively. Using this approach, I ran four batches of 30 generation ES runs, with a population size of eight for each of three analytes to be assessed. All Bu/Wo parameters consistently converged during the ES runs, in many cases while evolving to fractions of the feasible parameter space. The results of all four batches of TN Bu/Wo parameters evolution in the West Drain are plotted in FIGURE 20.

For all sites and analytes, the parameter *BuCo* most consistently evolved to nearly identical values across sets of four runs for a given site and analyte. Consistent with the RSA showing this to be the most sensitive parameter, it also was most consistently associated with a small portion of the parameter space. Parameters *BuEx*, *WoCo*, and *WoEx* converged to narrow ranges in most batches, and often converged on relatively narrow fractions of the feasible parameter space. However, in some cases these parameters converged to different final values across the four runs for each site and analyte, with different resulting fitness values. This suggests that for these less sensitive parameters, the ES algorithm was not able to escape from sub-optimal solutions, due to

either the relatively small fitness gains associated with optimizing those parameters, the additional noise incorporated into the search via the concurrent MC hydrologic model, or both.

TABLE 15. Drain sampled storms by sampling method. Dates shown indicate the date on which the first aliquot was sampled. '---' indicates a sample was not collected and / or analyzed.

Date	East Drain	West Drain
2008-07-23	----	Composite
2008-07-31	Composite	Composite
2008-08-08	Composite	Composite
2008-08-18	Composite	Composite
2008-09-09	Discrete	Discrete
2008-10-25	----	Composite
2008-11-25	----	Composite
2009-06-11	----	Composite
2009-06-26	Composite	Composite
2009-06-29	----	Composite
2009-06-30	Composite	Composite
2009-07-11	Composite	Composite
2009-07-18	----	Composite
2009-08-29	Composite	Composite
2009-11-14	Composite	Composite
2009-11-20	Composite	Composite
2009-11-27	Composite	Composite
2009-12-03	Composite	Composite

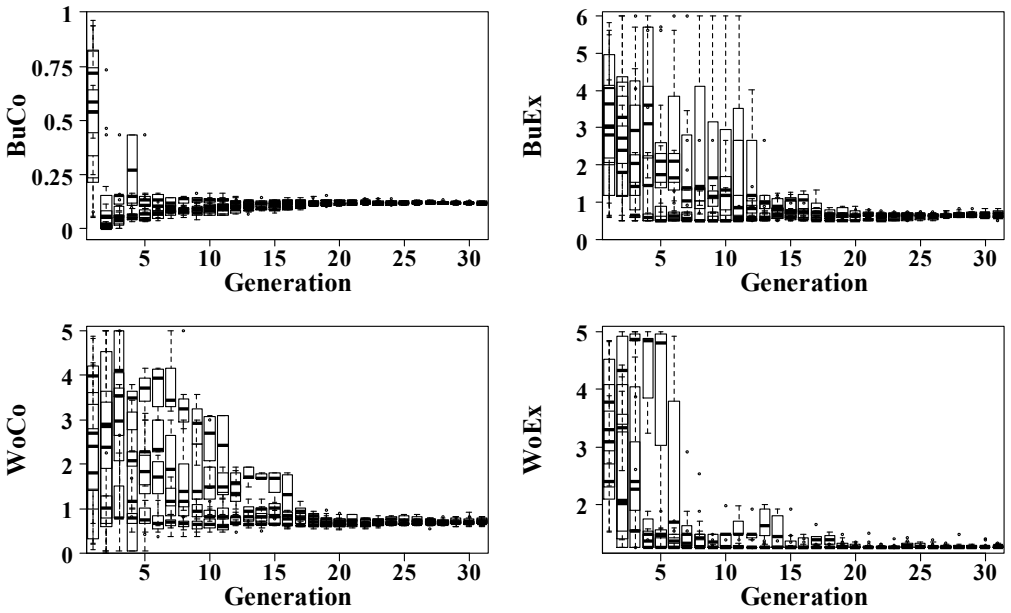

FIGURE 20. West Drain TN Bu/Wo parameter evolution. All four runs are concurrently plotted per parameter.

Thus, for subsequent modeling I considered all Bu/Wo parameters to be calibrated for both sites. Where the parameters had converged to different parts of their feasible ranges across the different runs, I simply considered these to be acceptable suboptimal solutions across the parameter space. Given that the Bu/Wo model is an empirical simplification of the processes controlling sediment and nutrient deposition, generation, and transport, it seems allowable to have different configurations of those models represent the data. To proceed, I extracted the best Bu/Wo parameter sets from last generation of each ES run, and from the best parameter sets among all other runs for each site and analyte, and ran an additional 10,000 simulations by sampling from the retained Bu/Wo sets while currently varying the underlying hydrologic model. Results from

running those simulations and computing load errors for the calibration and validation storms are summarized in TABLE 16.

TABLE 16. Buildup and washoff calibration and validation fitness values. Sum of the absolute load errors was used in calibration and is presented here as the percent of the total measured load over the calibration events (SAE%). PBIAS is essentially the same quantity, but not calculated on an absolute basis, such that over- and under-estimation errors can cancel out.

		TSS				TP				TN			
		Calibration		Validation		Calibration		Validation		Calibration		Validation	
Site	Fitness	Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean
East	SAE (%)	33.4	40.8	25.6	60.2	29.4	32.7	23.5	33.6	52.2	54.2	75.2	76.1
East	PBIAS (%)	0.0	-6.6	0.0	-40.1	-0.4	18.8	0.0	15.8	29.0	30.7	62.9	70.6
West	SAE (%)	23.1	46.0	29.9	60.0	22.8	27.6	74.8	81.7	33.5	39.2	40.1	44.8
West	PBIAS (%)	0.0	22.7	0.0	43.1	0.0	3.1	50.8	58.2	0.0	-7.8	18.0	32.7

TABLE 16 presents the sum of the absolute errors (SAE), as a percentage of the total load over the calibration or validation storms, as well as percent bias (PBIAS), for the same simulations. PBIAS is a more commonly used water quality model objective function, calculated as:

$$PBIAS = \left[\frac{\sum_{i=1}^{N}(meas_i - sim_i)}{\sum_{i=1}^{N}(meas_i)}\right] * 100$$

where N is the number of storm events (in calibration or validation), $meas_i$ is analyte load calculated from measured data, and sim_i is the analyte load calculated from SWMM outputs. While PBIAS is more commonly reported in the literature and is thus reported

here, SAE exerts stronger selective pressure due to the fact that over- and under-estimation errors cannot cancel, as they can in PBIAS.

All mean calibration SAE values were below the PBIAS threshold recommendations of +/- 55% TSS and +/- 70% for TN and TP threshold recommendations of Moriasi et al. (2007). This is notable as SAE is conservative relative to PBIAS, and similarly the PBIAS fitnesses from the calibration runs were all well below the previously described thresholds. Validation fitnesses tended to be a bit worse, but in most cases were still within calibration guidelines. Again, it is worth noting that I used event data at the temporal scale of minutes to hours, while the previously referenced calibration guidelines deal with monthly or daily aggregated data.

Lastly, for validation purposes I compared simulated pollutographs from the 10,000 calibration / validation simulations with the measured concentrations for the reserved portion of the discretely sampled storm event. FIGURE 21 shows the 95% prediction intervals on the modeled flow and analyte concentrations, overlain by measured flow and analyte concentrations for the East Drain. The earlier peak (7:30) was reserved for validation, while the latter peak (13:00) was used in calibration. Hydrologic model performance for the period shown in FIGURE 21 was very good (mean NSE = 0.88, mean RSR = 0.35), with the simulated flow closely matching peak rates and event dynamics over the three rainfall pulses. For water quality, model performance was relatively good for the calibration event, with sampled concentrations generally falling within the 95% range of model predictions. Results for the earlier validation event, in contrast, were quite poor with modeled values greatly over predicting measurements. I

attribute this poor performance to limitations in the model, and to the following unusual circumstance in the data record.

The Bu/Wo models are conceptually simple, but explicitly ignore key factors contributing to the variability in urban runoff concentrations. Factors such as seasonal vegetation dynamics (e.g., leaf fall, pollen deposition), temporary construction activity, and accumulation of different loads over winter periods are not easily incorporated in to this framework and thereby limit the accuracy that can be expected. Within the monitoring record, the discrete validation event performance was also limited by an unusually long dry period (20 days) preceding the event. No calibration events were preceded by similarly long dry periods, and as result, it was not possible for parameter sets producing this dynamic (i.e., unrealistically high concentrations following extended dry periods) to be penalized within the ES calibration framework. For the calibration events, the mean antecedent non-storm lengths were 6 and 6.5 days for the East and West Drains, with maximum dry antecedent periods of 14 and 11 days, respectively. The 20 day dry antecedent period affecting the validation event was also an outlier in the context of the storm events considered in the subsequent load analysis (Section 3.4.3).

It can also been from FIGURE 21 that for the smaller period of event flow between the calibration and validation peaks, the model predicted relatively high concentrations, for example 0.2 mg L^{-1} TP. While I do not have measured data for that event to either confirm or refute the predictions, I am of the opinion that this is a residually high load from the dry period preceding the validation period, and is therefore likely to be erroneous as well. Thus, while an improved water quality model is clearly

desirable, based on both limitations in the general water quality modeling framework and the inherent variability in urban water quality concentrations, I deemed these results to be of sufficient quality such that I could proceed with a qualified loading analysis for the larger SW2 drainage area.

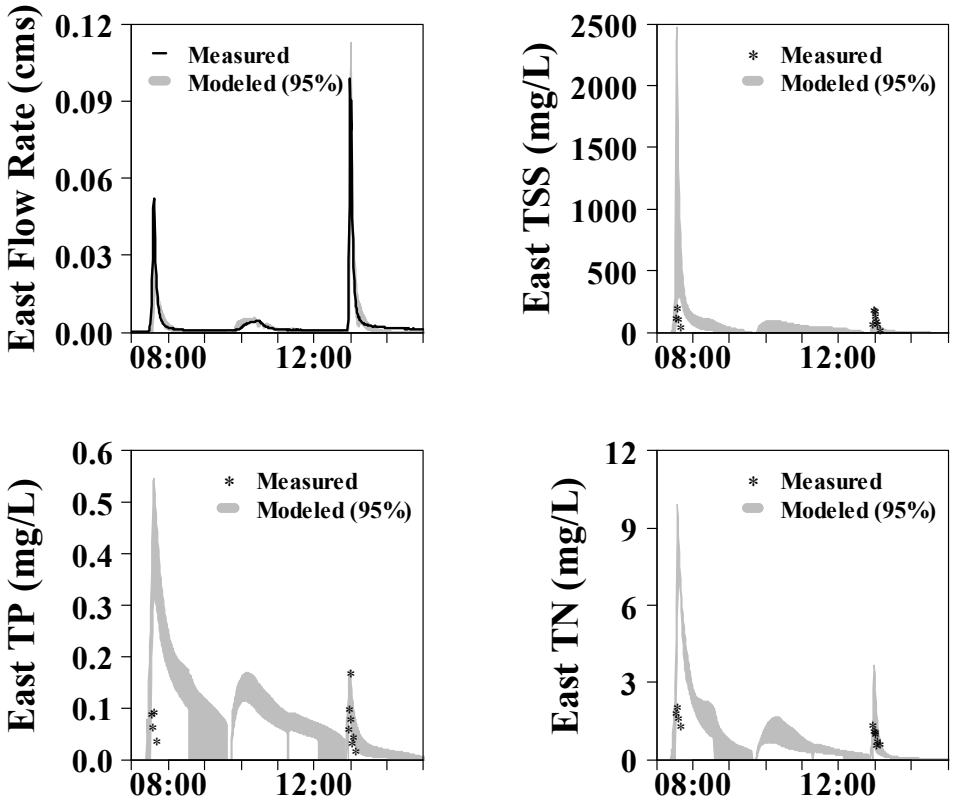

FIGURE 21. The 95% prediction intervals (grey envelope) resulting from the evolved hydrologic and water quality parameter sets for the East Drain, overlain with the measured data (lines or dots). The first peak with measured concentration data was used for validation, while the second was used in the calibration. The poor water quality performance for the validation event is partially attributable to an extended dry period preceding that event, outside the range of conditions included in the calibration data set.

3.4.2.4 Full Watershed Hydrologic Model (SW2)

The full watershed model ('SW2') made use of the RSA results, subwatershed calibration data, and manual calibration experience to define the parameterization approach for the 20 subcatchments within the model (the previously referenced 21 subcatchments included SW1 which was removed from the model based on discussion in Section 3.3.5.3). A subset of 12 parameters were determined to meet my criteria for calibration (TABLE 17), in that they are uncertain, the model was demonstrably sensitive to them, and narrowing of their feasible parameter space to improve model performance was deemed acceptable. For the parameters to be calibrated here, I was limited by the available flow records in my ability to conduct a more spatially discretized framework, for example calibrating neighborhood lawn and offsite meadow pervious areas separately. Thus, most subcatchments parameters were lumped over the SW2 watershed model, the exception being impervious surface characteristics, which were discretely identifiable based on the drain flow gauging. In summary, the parameters over the 20 subcatchments were either (1) fixed at central values, (2) Monte Carlo sampled from assumed distributions, (3) sampled with replacement from the drain calibration parameter sets, or (4) evolved to best value using a further implementation of the ES algorithm.

TABLE 17. Treatment of parameters in the Neighborhood and Pervious subcatchments contributing flow to the SW2 station, and subsequent results of ES calibration process where 'N' indicates the parameter did not consistently evolve to best value, while 'Y' indicates that it did.

Parameter	Neighborhood Subcatchments	Pervious Subcatchments	SW2 ES Result
Subcatchment			
nPrv	Calibrate	Calibrate	N
DSPrv	Calibrate	Calibrate	N
fMax	Calibrate	Calibrate	N
fMin	Calibrate	Calibrate	N
ICPct	MC	Fixed	---
Width	MC	MC	---
Slp	MC	MC	---
nImp	Drain Calib.	Fixed	---
DSImp	Drain Calib.	Fixed	---
ZeroIC	Drain Calib.	Fixed	---
Area	Fixed	Fixed	---
fk	Fixed	Fixed	---
fDry	Fixed	Fixed	---
Groundwater			
Por-FC	Calibrate	Calibrate	Y
FC-WP	Calibrate	Calibrate	Y
KSlp	Calibrate	Calibrate	Y
UEvap	Calibrate	Calibrate	Y
SElev	Calibrate	Calibrate	Y
BElev	Calibrate	Calibrate	Y
A1	Calibrate	Calibrate	Y
B1	Calibrate	Calibrate	Y
KSat	Fixed	Fixed	---
LEvap	Fixed	Fixed	---
GLoss	Fixed	Fixed	---

As discussed in Section 3.3.5.3, the upland agricultural area draining to the SW1 monitoring station was not well represented by any of the SWMM model structures or parametrizations explored in this work. Thus, I opted instead to route the measured SW1 flow record into the upper most node of the SW2 SWMM model, while using SWMM subcatchments and conveyance infrastructure to model the lower 65 hectares of

neighborhood and other watershed contributing area via 20 SWMM subcatchments. Using this framework, I ran four ES runs, following the approach outlined in TABLE 17.

The first objective in assessing the results was to further classify the parameters between those that 'calibrated' and those that did not. Parameters that consistently converged to the same portion of their feasible parameter space were deemed calibrated, resulting in a reduction in the feasible parameter space in subsequent modeling. Parameters that did not converge, or that did so inconsistently, were not considered to be calibrated and retained their full feasible ranges for Monte Carlo sampling in all subsequent modeling. The parameter evolution resulting from a single ES run are plotted in FIGURE 22, FIGURE 23, and FIGURE 24, displaying a range of evolutionary behaviors.

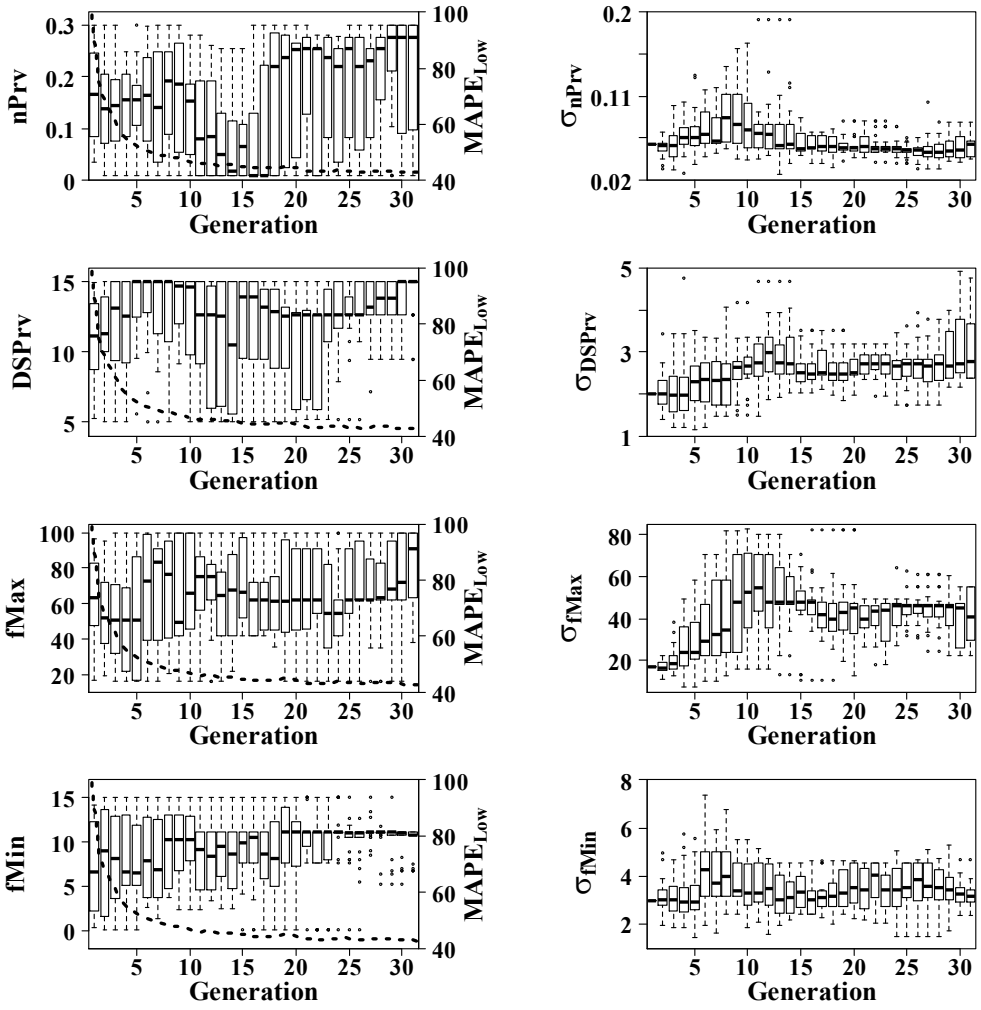

FIGURE 22. Evolution of SWMM parameters (*nPRv*, *DSPrv*, *fMax*, and *fMin*) and ES strategy parameters during an ES run for the Full Watershed model.

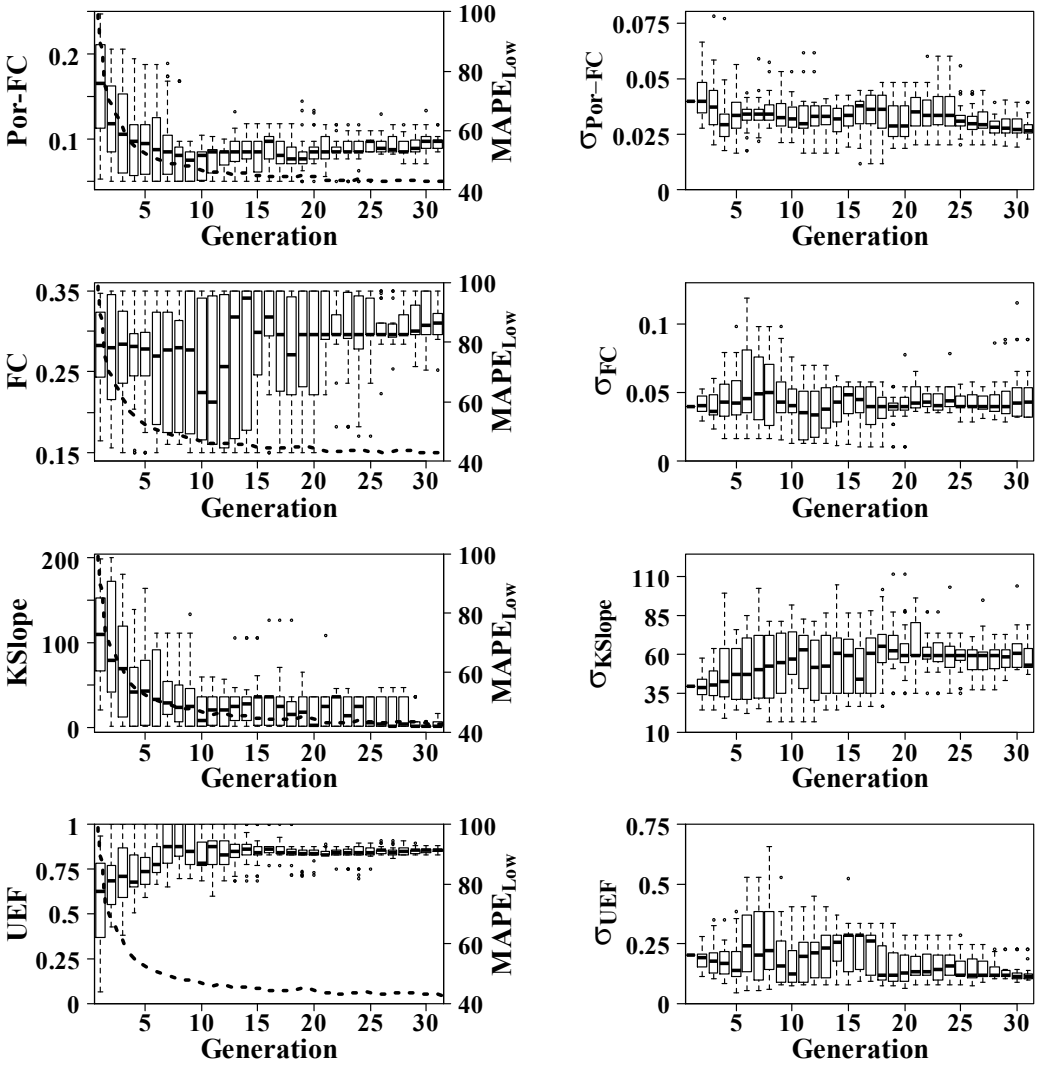

FIGURE 23. Evolution of SWMM parameters (*Por-FC*, *FC*, *KSlope*, and *UEF*) and ES strategy parameters during an ES run for the Full Watershed model.

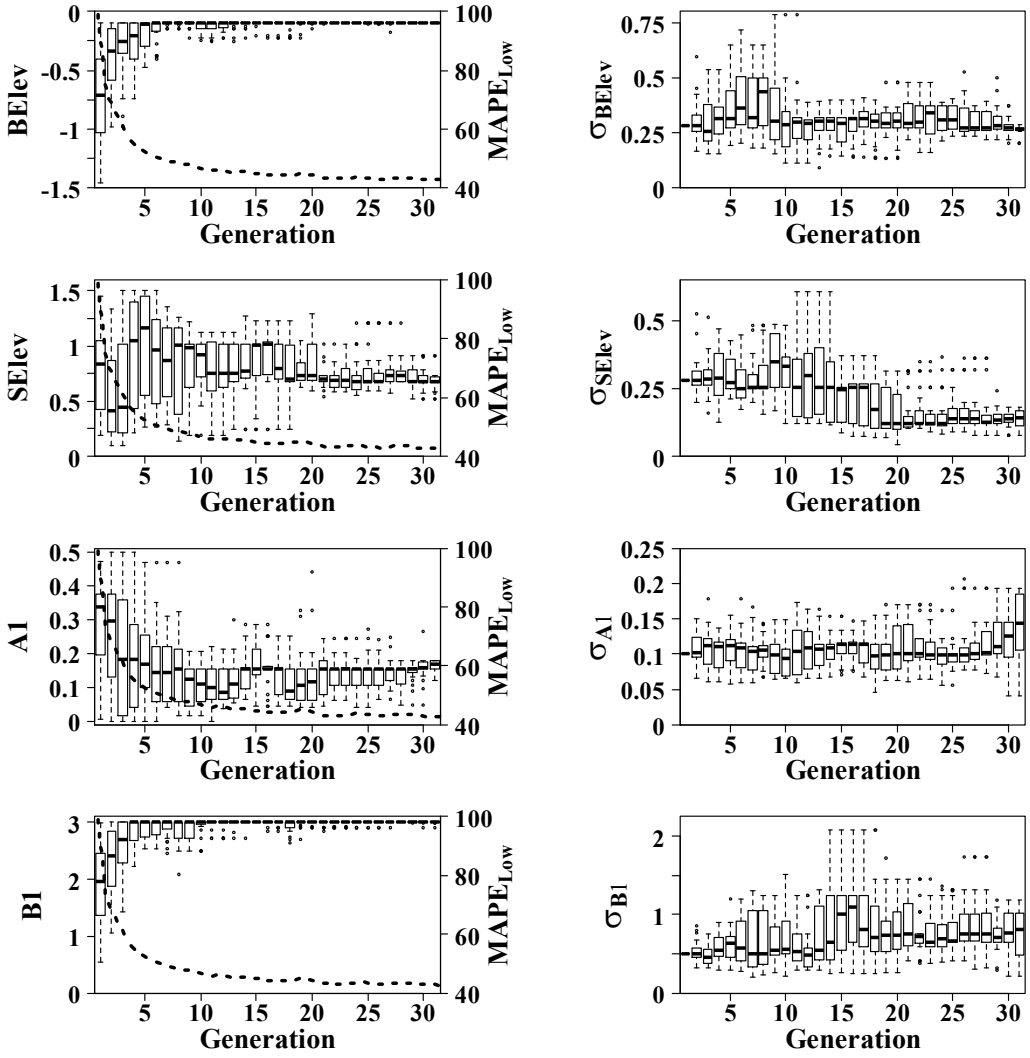

FIGURE 24. Evolution of SWMM parameters (*BElev*, *SElev*, *A1*, and *B1*) and ES strategy parameters during an ES run for the Full Watershed model.

Parameters *nPrv*, *DSPrv*, *fMax*, and *fMin* either did not consistently evolve or did not consistently evolve to the same region of the parameter space. Based on this, I was unable to identify best values and so treated these as Monte Carlo parameters in subsequent modeling.

Several parameters consistently evolved against upper or lower bounds through the ES runs. In some cases these were physical constraints (i.e., evolved toward zero where negative values are not possible) though in other cases these boundaries were imposed. For example, parameters *B1* and *Por-FC* tended to evolve to upper and lower bounds, respectively (though *Por-FC* did not in the run shown in FIGURE 23). In each of these cases, I chose to maintain these boundary limits, based on my a priori definition of the allowable parameter space.

For other parameters, values consistently evolved to the same region of their parameter space, which is strongly suggestive of a best value. For example, parameter *UEF* evolved most consistently of all parameters, converging to a very narrow range in the middle of the feasible parameter range during all four ES runs. *UEF* governs the availability of shallow subsurface water to meet ET demand, and thus can be an important control on water balance. Considering the evolved behavior of these parameters collectively, it is clear that shallow subsurface dynamics were key to the improvement in $MAPE_{Low}$ that occurred during the ES runs. In addition to *UEF* and *Por-FC* (discussed above), *SElev* was among those that consistently evolved to a narrow range (between 0.5 and 1.0 m). Combined, these parameters govern the subsurface volume over which storage, percolation and subsurface ET losses can occur. These storage and flux terms appear to be important model components in maximizing the calibration fitness. The importance of the aggregated parameter *Por-FC* in this study was primarily made possible as a result of our RSA work, and could not be easily detected or calibrated using local or manual methods (although it could hypothesized and tested

independently based on knowledge of soil physics and / or from SWMM algorithm review). These results suggest that for other SWMM modeling applications where stream hydrograph recession performance is to be optimized, the aggregated subsurface void volume above *FC* (as a function of *SElev*, *BElev*, *Por*, and *FC*) should be considered an important parameter grouping to be considered concurrently.

For the parameters that did converge over the ES runs, the corresponding evolution step sizes, or sigmas, consistently evolved toward smaller values as well. Compared with the East Drain ES however, the sigmas did not typically evolve to lowest the allowable values. (These minimum allowable sigma values were defined arbitrarily as parameter range divided by 250 in initial algorithm testing.) Rather, sigmas generally tended to not take on large values later in the ES runs, but without full convergence toward the lower bound. I suspect that the cumulative noise of many ES and MC varying parameters, combined with the use of an indirect fitness function (i.e., $MAPE_{Low}$ of SW2 flow record) interfered with the narrowing of the optimal parameter space. It is also acknowledged that given the high dimensional and interdependent parameter space there are potentially a large number of unique parameter sets that can minimize the $MAPE_{Low}$ objective function. Combined, the evolutionary behavior results across the various batches (i.e., Drains, Bu/Wo, Full Model) demonstrates that the self-adaptivity of our ES implementation works as intended despite the incorporation of noise, but that the ES' performance appears to decline as the amount of noise is increased.

Based on the evolutionary behavior of the SWMM parameters over the four ES runs, I proceeded to determine which parameters had evolved sufficiently that calibrated

values could be extracted, thereby limiting those parameters' feasible space in subsequent modeling. There was a large degree of variability in the evolution behavior, both for the same parameters across ES runs and among the different parameters evolved, requiring some subjective judgement in classification of the parameters. In general, if a parameter repeatedly converged to a similar fraction of the allowable range I considered it calibrated, while allowing that one of the four ES may not have converged. Using this approach, I deemed all but the four surface parameters (*nPrv*, *DSPrv*, *fMax*, and *fMin*) to be calibrated, with the results indicated in the last column of TABLE 17. For those that were deemed calibrated, I extracted complete parameter sets from the final generations and the best runs in prior generations as previously described. These parameter sets were then taken to be the best unique realizations of the eight dimensional calibration parameter space, without intermixing of parameters values from within the sets. For those parameters not determined to have been successfully calibrated by this approach, I subjected them to simple MC sampling from their full parameter ranges in subsequent modeling.

To assess the calibration and validation fitness for the Full Watershed Model, I next ran 40,000 blowout simulations, where for each simulation I sampled 1) an impervious surface parameter set randomly from the pooled sets identified in the East and West Drain calibrations, 2) a subsurface parameter set drawn from the parameter sets retained from the Full Model calibration runs, 3) other sensitive parameters randomly and independently sampled from assumed uniform distributions, and 4) other parameters fixed. I first applied this framework to the 2009 calibration year, in doing so propagating

both the surface and subsurface calibration sets and the full ranges of Monte Carlo parameters through to the hydrograph record. I then applied the same approach to the 2007 and 2008 years of record as model validation.

A last component I needed to assess was the role of routing the upland measured flow record through the SWMM model, as opposed to simulating flow for that area. My goal here was to assess the performance of the Full Watershed model, while controlling to a degree for this flow routing which could be expected to produce acceptable SW2 performance independent of any SWMM subcatchment contributions. To do so, I calculated a number of objective functions both for the Full Watershed model with SW1 measured flow routed through, and for a 'No Rain' scenario in which simulated flow at SW2 consisted solely of SW1 flow routed through the conveyance network (i.e., channel and culverts). In doing so, I attempted to quantify the improvement in fitness that results from modeling the rainfall runoff response of the lower catchment, relative the model performance achieved simply by routing the measured SW1 flow record through the model. Several fitness measures, on the basis of the preceding approaches, are summarized in TABLE 18. In addition to *MAPE*, $MAPE_{Low}$, *RSR*, and *PBIAS*, I include Nash-Sutcliffe Efficiency (*NSE*) here, calculated as:

$$NSE = 1 - \frac{\sum_{i=1}^{N}(meas_i - sim_i)^2}{\sum_{i=1}^{N}(meas_i - meas_{mean})^2}$$

where all variables are as previously defined. Measured and modeled (95%) hydrographs for the three years of record are shown in FIGURE 25.

TABLE 18. Calibration and validation fitness for the Full Watershed model. At top, the fitnesses calculated using the retained simulations from the four Full Watershed model ES runs. The 40,000 blowout runs of the calibrated set are given below, for the calibration (2009) and validation (2007-08) years. At the bottom, the 'No Rain' scenario results are summarized. Objective functions are Mean Absolute Percent Error (MAPE), MAPE of flows below the annual 75% flow ($MAPE_{Low}$), RMSE-observations standard deviation ratio (RSR), and Nash-Sutcliffe Efficiency (NSE).

Year	Type	MAPE (%)		$MAPE_{Low}$ (%)		RSR		NSE		PBIAS (%)	
		Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean
2009	Calibration Set	39.6	48.5	40.0	49.6	0.40	0.54	0.84	0.69	0.2	-13.4
2007	Blowout 40,000	88.7	97.4	76.7	83.0	0.53	0.73	0.72	0.46	-58.0	-70.1
2008	Blowout 40,000	93.8	122.7	108.3	146.5	0.36	0.45	0.87	0.23	0.0001	-3.79
2009	Blowout 40,000	38.9	41.8	39.3	42.5	0.40	0.52	0.84	0.73	2.0	-13.1

		MAPE (%)	$MAPE_{Low}$ (%)	RSR	NSE	PBIAS (%)
2007	Upland Routing	95.2	98.4	0.84	0.29	70.7
2008	Upland Routing	89.5	97.0	0.66	0.57	66.8
2009	Upland Routing	74.3	79.3	0.58	0.66	51.2

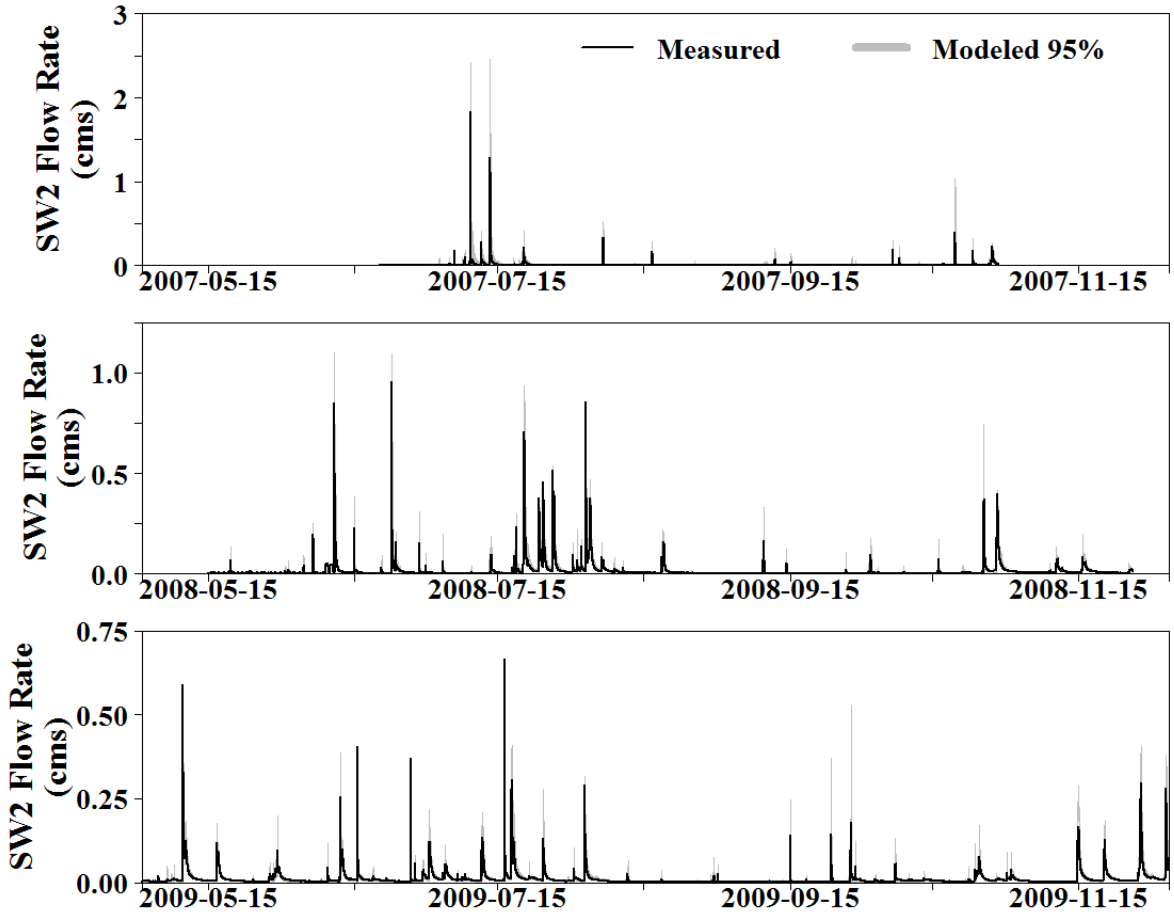

FIGURE 25. 95% prediction intervals for the Full Watershed model calibration (2009) and validation (2007-08) periods. Model predictions (grey bands) are overlain by measured flow (black).

Model performance was generally good in both the calibration and validation periods, based on tabulated objective function values (TABLE 18) and visual assessment (FIGURE 25). Best performance was generally seen in the calibration year 2009, with worse performance in validation years. Given the use of $MAPE_{Low}$ as the optimization criteria during calibration, the parameters appear to have evolved to optimize recession and inter-event low flow performance for the relatively numerous and evenly spaced storms of the 2009 calibration year. Validation years 2007 and 2008 had temporally

tighter clusters of storms interspersed by generally longer inter-event periods where hydrograph recession dynamics have receded. The $MAPE_{Low}$ fitness is highly sensitive to inter-event dynamics since periods of either low / zero simulated or measured flow can compute to very high *MAPE* values. While this is a useful feature during ES algorithm selection, it does not necessarily provide a clear picture of overall model performance, and can partly explain the disparate *MAPE* fitnesses computed between calibration and validation. I also computed *RSR* and *NSE* to more broadly assess the annual records in their entirety. Best values of these objective functions were good for both calibration and validation, and were above minimum performance thresholds identified by Moriasi et al. (2007) for these measures when using daily flows. The mean values reflecting the broader parameter space considered acceptable or plausible simulators of the system, however, were considerably worse in some cases. A key limitation in this work was that I did not collect data that would allow direct calibration of pervious area or subsurface hydrology components of this work (and in doing so, further restrict the parameter space). Instead, I had to rely on low flows in the SW2 streamflow record, which are affected by pervious areas and subsurface, as well as conveyance, stormwater treatment infrastructure, impervious surfaces, climate, and measurement errors.

I also computed the objective function *PBIAS* to assess the average over or underestimation of flow, and particularly to highlight the differences between the Full Model and the 'No Rain' / SW1 routing scenarios. For the calibration (2009) and 2008 validation year, the Full Model best and mean values were within the +/- 25% range recommended by Moriasi et al. (2007) for daily flow data; however the best simulation

for validation year 2007 was -58%. I attribute this 2007 error largely to the two large storms in the beginning of the 2007 monitoring season, which were the largest two storms on record on a peak rate basis. Given that the 2007 flow record was entirely unseen to the calibration algorithm, this poor performance is to be expected. Further, I suspect there were headwater conditions at the culvert inlet ~10 m downstream of the SW1 monitoring station during this event which would have led to overestimation of flow from the upper catchment (SW1). If this were the case, as I suspect, then the actual *PBIAS* for 2007 would be even higher than that reported in TABLE 18 (i.e., if the model were compared against 'true' flows rather than overestimated measured estimates).

Lastly, the routing of just the upland flow record through the model with no additional participation inputs is illustrative of the degree to which good model performance is attributable to accurate SWMM simulation in addition to routing of existing hydrology. The best values of the objective functions *RSR* and *NSE* for example, were both noticeably worse in the upland routing model compared to the full model. The *PBIAS* metric, expectedly, showed significant underestimation of flow by the model when precipitation inputs to the lower catchment are excluded. *MAPE* and $MAPE_{Low}$ were for the most part equally poor in the upland routing model, as compared to the full model validation years. However, since *MAPE* was selected primarily for its strong selective pressure on event recessions and inter-event dynamics during the calibration process, I do not take its lack of differentiation here to negate the performance gains seen in the other fitnesses. Thus, overall I am comfortable attributing the acceptable objective function values from the calibrated and validated model (to the extent they can be

interpreted as such) to good model performance, and not simply the routing of the measured upland flow record through the catchment.

3.4.3 *Full Watershed Water Quality Analysis*

In this section, I applied the calibrated and validated Bu/Wo models to the entire residential area within the SW2 drainage area to estimate the relative contribution of the neighborhoods to the total loads measured at the watershed outlet. To do so, I randomly sampled complete four parameter Bu/Wo sets for each analyte from the retained calibration / validation sets described in Section 3.3.5.2. The TSS, TP, and TN parameter sets were then coupled with a random selection of hydrologic model parameters following the approach used in Section 3.4.2.4 to create a complete hydrologic and water quality parameter set for each model run, consisting of both calibrated and random MC sampled parameters.

The Bu/Wo parameters were only applied to the residential areas within the drainage, with the golf course and meadow areas assigned values that would produce zero load, based on the lack of data for estimating water quality parameters for these areas. This approach resulted in a simulated load at the outlet representing only total residential washoff load. This process was iterated 5,000 times. I then extracted the simulated washoff record and computed loads for 21 storms on record for which SW2 outlet loads were successfully sampled via flow weighting. (Issues that excluded sampled events from this analysis included the autosampler completing a sampling program mid-storm; storms where autosampler aliquots missed rising limbs and/or peaks; and storms where

the autosampler program was not properly configured for rates/volumes that occurred resulting in poor flow weighting.) For many of the SW2 events deemed usable in this context, we also have either sampled load data or a record of minimal or no flow at the SW1 monitoring station, allowing the inclusion of upland contributions of sediment and nutrients in this analysis. The results of this analysis are summarized in FIGURE 26.

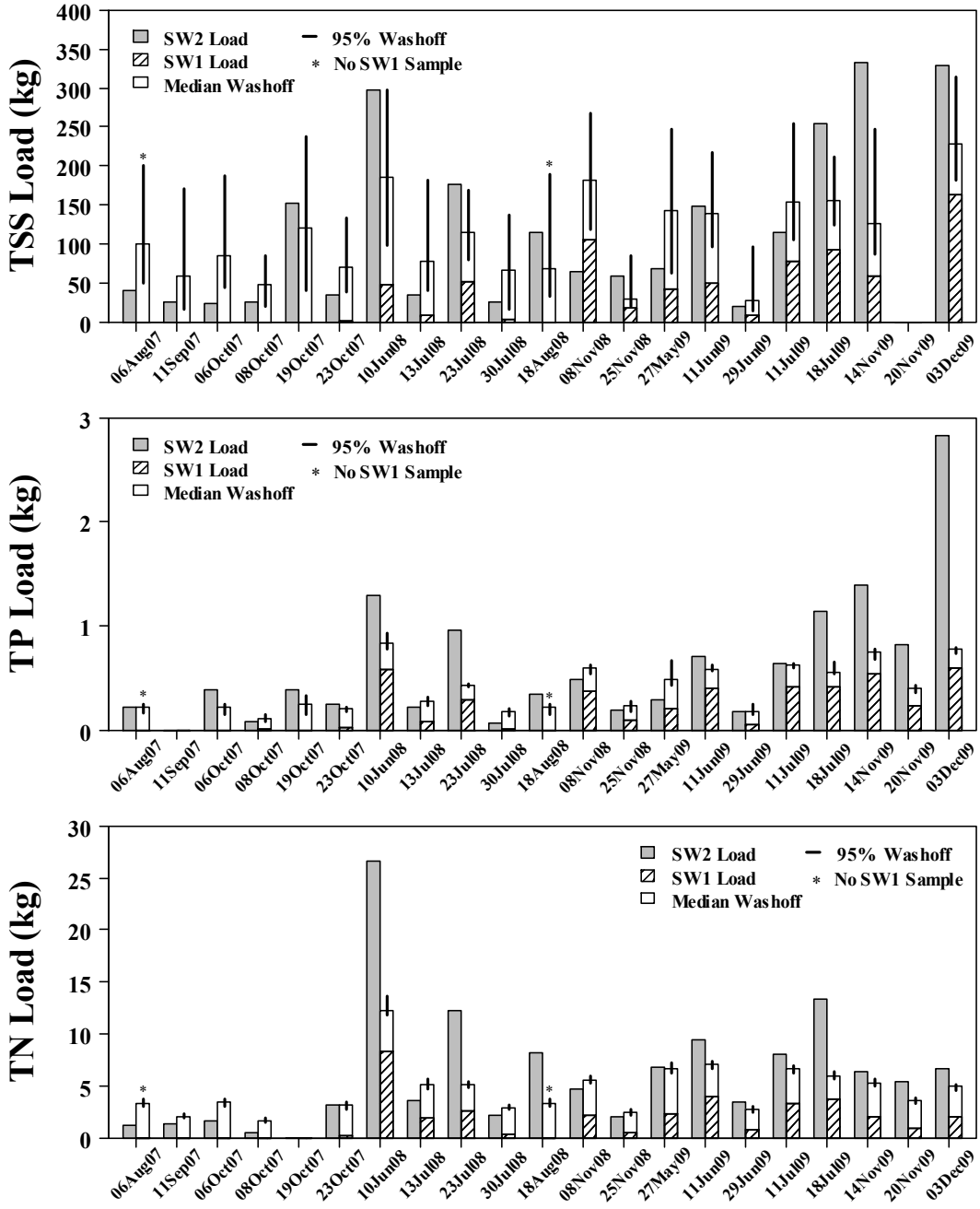

FIGURE 26. Sampling results and model predictions for 21 storm events. Vertical lines indicate 95% prediction range for the neighborhood loads stacked on SW1 loads where available. Asterisks indicate events for which there was measurable flow at SW1, but a valid flow-weighted load was not collected. Other events with missing SW1 loads indicate negligible flow at SW1. Missing events for different analytes are where an analyte was not analyzed, either due to a sample processing issue (e.g., hold time exceeded) or paperwork issue (i.e., failed to indicate that parameter on the chain of custody form).

FIGURE 26 summarizes the modeled and measured storm data by plotting the sampled SW2 loads adjacent to stacked plots of the SW1 loads and modeled neighborhood washoff estimates, with the error bars indicating 95% model predictions on neighborhood loads. For the measured SW1 and SW2 loads, error estimates are not displayed graphically. However, Harmel et al. (2006) provide estimates of uncertainty in various components of small watershed load measurements, including streamflow measurement, sample collection, sample preservation and storage, and laboratory analysis. They compute probable error ranges of 18%, 30%, and 29% for TSS, TP and TN loads, respectively, as average estimates given 'typical scenarios'. These estimates are likely to be either appropriate or conservative for the data collected in this study, given that samples were analyzed by a multi-state certified commercial laboratory.

Cumulatively, this data summary allows for interpretation of the relative magnitude of the loads that were predicted by the model and measured in stream. The modeled neighborhood TSS load, for example, had the most variability out of the three analytes for which this analysis was completed, due in part to the order of magnitude peak TSS error rates seen in the discrete storm validation event (FIGURE 21). For many of the storms, modeled neighborhood load plus the SW1 load (or in many cases, the modeled neighborhood load alone) equaled or exceeded total watershed loads. This suggests that sediment loads conveyed to the lower stream channel by the closed drainage infrastructure and the upland area were not always fully delivered past the lower monitoring station at the event scale. In other storms, notably the second half of 2009, the predicted inputs of sediment to the lower channel were less than the measured load at

SW2, suggesting additional load sources such as remobilization of previously retained loads.

I attribute these discrepancies in mass balance to a combination of retention of sediment within the stream channel, remobilization and erosion of sediment from within the channel, and aggregate error. The tributary to which the neighborhood's closed conveyances discharge is a low gradient first order stream, which has dense vegetation in sections and includes several hydraulically inefficient culverts which provide opportunity for settling of coarser sized particles as entrance losses occur. This dynamic (i.e., loss of neighborhood event loads within the conveyance network) is also anecdotally supported by visual observations made of the collected samples. The storm drain samples appeared to have a greater presence of coarse sediment and coarse organic matter, compared to the instream sampling which tended to have more fines and turbidity in general.

It should also be noted that during the years of sampling there was active bank erosion in sections of the channel through the native clay soil, such that during event flows the channel could conceivably both remove coarse sediment from transport and contribute finer sediment endogenously. Lastly, given the discrete storm validation error (FIGURE 21), the neighborhood washoff loads are more likely biased toward overestimation of sediment export than to underestimation. If so, that would suggest channel retention is being overestimated to some extent in FIGURE 26 and that the system is either more balanced or includes greater unaccounted sediment loads. Further, there was a weak but significant linear relationship (F test *p-value* = 0.035, adjusted R^2 = 0.19) between the event peak flow rates at SW2 and the difference between SW2 TSS

load and the sum of SW1 and neighborhood loads (data not shown). This provides additional circumstantial evidence that upland and neighborhood loads were retained in the channel during smaller events (when the difference between SW2 TSS load and the sum of SW1 and neighborhood loads was negative) and that the lower channel acted as a sediment source during larger events.

For TP in contrast, the sum of the measured SW1 load and the modeled neighborhood load was often less than or close to the measured SW2 load. Further, it should be noted that total neighborhood loads were often small compared to corresponding SW1 loads, suggesting that TP loads from neighborhood washoff were not disproportionate given the contributing drainage area. While the data in FIGURE 26 are presented on the basis of total loads rather than area normalized loads, the total SW1 drainage area is only 35% larger than the neighborhood, while the agricultural area of SW1 is approximately equal to the neighborhood area. Thus, total loads from SW1 and the neighborhood can be compared directly. Since SW1 typically produced more TP than the neighborhood during storms for which SW1 produced flow, it can be inferred that surface washoff loads from the neighborhood were not the dominant source of TP for this tributary.

For many storms the SW2 TP load was more than other estimated sources combined suggesting important roles for unaccounted sources of TP to the system and model error. The modeling framework used does not account for contributions from channel erosion, yard waste disposal within the brook, sanitary sewer leakage, or inputs from the golf course or meadow. Thus, some of the discrepancies between sum of SW1

and neighborhood loads and what was measured at SW2 may be attributable to those sources. However, it should also be noted that two of the largest unaccounted load discrepancies in FIGURE 26 occurred during the late fall 2009. These discrepancies could potentially be attributed to the Bu/Wo model not capturing autumn leaf fall dynamics since the Bu/Wo model was calibrated on a composite basis across storms from spring, summer and fall. The concentration data collected are weakly suggestive of higher fall TP concentrations in the West Drain, however the small number of storms sampled limits further analysis. Thus, it is possible that the fall TP load discrepancies are attributable in part to underestimation of leaf contributions by the Bu/Wo model, and to direct leaf inputs to the upper channel which have begun to decompose prior to reaching the SW2 monitoring station.

The TN analysis showed the best agreement in cumulative loads among the measured and modeled quantities and produced the strongest linear relationship between peak flow rate at SW2 and unaccounted load (the TP regression was not significant). The modeled range plus the measured SW1 load frequently summed to approximately the SW2 measured load, though there were exceptions. For the larger summer storms, there were substantial SW2 loads that were not accounted for. Similar to what was speculated for TP, it is possible that storm flows mobilized additional material via bed or bank erosion, resuspension of material from the main channel or connection with transient storage pools within the drainage network. This is further supported by the linear relationship between SW2 peak flow rates and unaccounted for TN load data, which was significant (F-test *p-value* < 0.001) and explained 62% of the variance in TN load

difference. The TP discussion also considered the role of leaf inputs in the set of three late 2009 storms with large TP load discrepancies. The same discrepancies do not exist for TN, which is consistent with this interpretation.

TN is potentially subject to denitrification losses in addition to transport and retention via bio-assimilation and burial pathways to which TP is subject. For inorganic nitrogen, 50% or more can be retained along a reach (Peterson et al. 2001) with retention in first order streams typically occurring within 101-478 meters (Ensign and Doyle 2006). Among inorganic nitrogen retention and loss pathways, the role of denitrification losses is highly variable, which Mulholland et al. (2009) have estimated as accounting for between 0.5% and 100% of inorganic nitrogen removal from transport, with a median denitrification loss of 16% across a range of study sites including reference, agricultural and suburban-urban streams. Lastly, debris dams in urban streams have been found to be hotspots for denitrification, relative to similar geomorphic features in forested stream reaches, with denitrification potential increasing with ambient stream nitrate concentrations (Groffman et al. 2005).

There is no evidence to suggest that the preceding inorganic nitrogen dynamics would not exist within the headwater Potash Brook tributary studied in this work. The neighborhood had relatively high nitrate concentrations (combined East and West Drain storm interquartile range of 0.41 – 0.87 mg L^{-1}), suggesting that nitrate was available for loss and that levels may have been high enough to stimulate elevated denitrification potential within the channel sediments (Groffman et al. 2005). And although the channel was straightened at one time, likely removing heterogeneous channel features that could

maximize denitrification potential, it has since revegetated along portions of the riparian corridor creating a new supply of terrestrial debris. Further, the many hydraulically inefficient culverts along the channel create areas of slower flow where organic matter can accumulate, potentially leading to conditions that would favor denitrification losses. While nitrogen retention and losses cannot be quantified given the data that were collected in this study, these factors do provide a plausible explanation for the 2009 fall storm differences in TN and TP export dynamics.

Some important caveats limit the interpretation and analysis that can be made from these data. First among them are the inherent limitations in the buildup and washoff algorithms. The model uses time between events to predict accumulation of surface loads, and thus cannot account for autumn shedding of leaves from the deciduous trees which line the streets throughout the neighborhood, or other episodic loads such as pollen deposition or winter detrital accumulation. Thus, to the extent that those processes add variability to the buildup of pollutants at the land surface, the Bu/Wo algorithm cannot be calibrated to match those processes. Rather, the calibration will have identified best parameter sets on average, given the range of conditions contained in the sampling records. This can be noted from FIGURE 26 where modeled loads generally had lower inter-event variability than either the SW1 or SW2 loads. However, the generally acceptable Bu/Wo model performance in model calibration and validation (TABLE 16) suggests that unaccounted for inter-event variability was not so great as to prevent interpretation of these data.

Another limitation stems from the other processes and factors affecting load delivery that are not accounted for. For example, channel processing and meadow and golf course contributions are not accounted for in the SWMM water quality modeling, but are reflected in the SW2 measured loading. Thus, any discrepancies between the measured and modeled loads can be attributed to both error in the modeled and measured loads, and the effects of channel processing and additional loading from the golf and meadow areas. Lastly, the areas of the neighborhood used in water quality calibration and validation include parts of 23 of 245 lots within the neighborhood. This provides an opportunity for any atypical lot dynamics on those parcels to skew this analysis. However, the construction and layout of this neighborhood are quite uniform which was reflected in the fact the correlation between paired East Drain and West Drain EMCs was high (data not shown). Considering the preceding, I judge the sediment and nutrient dynamics discussed in this section as more likely to be correct than incorrect. However, the uncertainties and limitations in methodology discussed herein limit the strength of the conclusions that can be drawn.

3.5 Conclusions

In this work I have applied a global sensitivity analysis and evolutionary calibration approach to EPA SWMM, and conducted a watershed loading analysis for a developed headwater tributary to Potash Brook. Key conclusions from this work include:

- Even using the global RSA method to sensitivity analysis, substantial differences in parameters' sensitivity can emerge as result of differences in model structure

and configuration. Study-specific sensitivity analysis is therefore clearly warranted in cases where a sufficiently similar structure and parameter space have not previously been assessed.

- For the sensitivity analysis of the previously unassessed subsurface flow model, a collection of parameters defining free draining subsurface storage volume and evapotranspiration were of high importance. For other workers attempting manual calibration of the SWMM subsurface model, these RSA plots may provide useful information for efficiently perturbing the high dimensional parameter space.
- I compared the traditional SWMM User's Manual approach to subcatchment *Width* calibration with the more recently proposed Guo Method (Guo and Urbonas 2009). In the two subcatchments that I assessed, a priori Monte Carlo calculations using the Guo Method closely matched the data calibrated results following the SWMM User's Manual approach. That I was able to calculate the same values of this key sensitive parameter without using any calibration data provides further confirmation of the approach.
- A canonical evolution strategies algorithm proved to be a useful approach to calibrate uncertain conceptual and empirical parameters and generally worked well even with the incorporation of noise, as implemented by concurrent MC sampling of non-calibration parameters. The algorithm work less well (i.e., lack of sigma convergence toward lower bounds) as the varied search space size increased, however poorer performance under those circumstances was at least

partially attributable to noncommensurability and low information content in the calibration data. Isolating model processes and collecting commensurate calibration data where possible should be a high priority in future work to allow for further narrowing of the high dimensional parameter space.

- A loading analysis comparing modeled neighborhood loads with measured watershed loads suggested that neighborhood sediment loads were both retained in the channel and derived from unaccounted sources (e.g., the channel), that additional unaccounted for sources (e.g., channel erosion) contributed TP loading to the lower sampling stations, and that TN loads were relatively balanced among the sources explicitly considered. A number of limitations in methodology were discussed which suggest additional measurements and modeling that could be employed to more tightly close watershed mass balance of these analytes.

There are several limitations and caveats on this work that warrant elaboration and summary here. For the RSA work, while the global approach employed provides strong insight into various SWMM component sensitivity, this work does not eliminate the need for additional study specific SA. Differences between parameter sensitivity in the SW1 and East Drain subcatchments highlights the role that model structural configuration and base parameterization play in determining sensitivity, in addition to the magnitude of parameter perturbation used.

In the calibration work, I attempted to employ an objective approach to narrow the feasible parameter space where possible and elsewhere propagate parameter

uncertainty through to model predictions where that parameter uncertainty could not be eliminated. Nonetheless, the approach required a large number of subjective decisions to implement, such that other modelers employing the same approach might generate disparate results. It should also be noted that, while I dealt explicitly with the issues of uncertainty in parameters, uncertainties in forcings / inputs and model structure were not assessed. A more holistic uncertainty assessment would likely generate a wider range of predictions than was presented here.

Lastly, the loading analysis findings are best viewed as exploratory given the various limitations inherent in the approach. The buildup and washoff model calibration and validation produced generally acceptable results, however its simplified empirical form provides at best an average accounting of seasonal or annual conditions. I attempted to interpret the difference between measured and modeled loads on a process basis, while allowing that those differences include both unaccounted processes and cumulative error. Combined, these factors prevent the drawing of more conclusive findings about channel contributions to the watershed TSS, TN and TP loads and dynamics considered.

CHAPTER 4 COMPARISON OF VERMONT TOTAL PHOSPHORUS AND TOTAL NITROGEN EVENT MEAN CONCENTRATION DATA WITH NATIONAL DATASETS

4.1 Abstract

Eutrophication is a leading source of water quality impairment in the U.S., and urbanized nonpoint phosphorus sources are recognized to be high (relative to receiving water targets) and variable. Comprehensive regional and national sampling efforts to characterize nutrient concentrations in urban runoff have excluded Vermont, due in part to its lack of Phase 1 MS4 systems. The pipe outfall total nitrogen (TN) and total phosphorus (TP) data collected in the studies detailed in the two preceding chapters are therefore somewhat unique in their applicability to Vermont urban runoff management. Here, I summarize the previous data compilations and compare the data collected in this dissertation's work with previous findings from elsewhere in the U.S. While TN concentrations sampled in this work were generally commensurate with what has been previously reported elsewhere, TP concentrations were not. Drainage area attributes and an event based rainfall runoff analysis of the study catchments provide circumstantial support for a conclusion that lawns contribute a disproportionate proportion of the high TP concentrations at the Englesby study area. Similar analyses in Butler Farms suggest that many of the sampled storms were driven entirely by directly connected impervious surfaces, which corresponded with relatively low TP concentrations. Combined, this analysis suggests current Vermont State stormwater quality regulations (based solely on impervious cover) could be more precisely targeted at developed land pervious areas if the circumstantial findings of this research can be further verified.

4.2 Introduction

Non-point source pollution is the cause of considerable water quality impairment throughout the U.S. (U.S. EPA 2013). This is not for lack of understanding or management efforts, but rather due to the relative intractability of controlling diffuse sources (Novotny 2003). When considering stormwater (i.e., non-agricultural diffuse source pollution), there are a range of impacts from land surface development which can adversely affect water quality. Impervious surfaces and associated development can alter the storm driven and baseflow hydrologic responses (Leopold 1968; Simmons and Reynolds 1982), the sediment and solute loads (U.S. EPA 1983) and temperatures of runoff (Galli 1990), and can produce geomorphic impacts (Hammer 1972; Booth 1990) and changes in riparian communities (Naiman and Decamps 1997; Lyons et al. 2000), all of which can degrade physical habitat or directly affect organisms in streams (Paul and Meyer 2001).

While each of these impacts can be highly consequential under specific circumstances, increased nutrient loadings and eutrophication are of particular importance for several reasons. First, nutrients are the second and third leading causes of water quality degradation in lakes and ponds and rivers and streams, respectively, with the other leading causes of impairments not shared among these freshwater groupings (U.S. EPA 2013). Additionally, the ubiquity of nutrients in the environment, both at background levels and from a variety of anthropogenic sources, complicates management. The variety of potential sources of nutrients requires detailed knowledge of

the relative contributions of the various sources to formulate effective management strategies such as Total Maximum Daily Loads (TMDLs).

In freshwater ecosystems, phosphorus is often the limiting nutrient and thus key for managing eutrophication (Novotny 2003; Schindler 1977). While there are background sources of phosphorus that affect freshwater ecosystems (e.g., geologic, Abrams and Jarrell 1995; atmospheric deposition, Litke 1999; canopy leaching Waschbusch et al. 1999), land surface development has frequently been found to correlate with elevated phosphorus runoff concentrations, relative to pre-development or background levels (Pitt et al. 2004; U.S. EPA 1983). Some of the sources of phosphorus to which elevated stormwater concentrations have historically been attributed include phosphates in detergents, pet wastes, and turf grass runoff. While these potential contributing factors vary in their manageability, on a combined basis they are important controls on the elevated phosphorus concentrations that have been measured in previous stormwater research (U.S. EPA 1983).

Phosphates (primarily sodium tripolyphosphate) were initially added to soaps and detergents as chelating agents in the 1940s, based on their ability to form complexes with calcium and magnesium thereby enhancing cleaning efficacy of detergents (Baird 1999). Recognition of the role of phosphates in freshwater eutrophication eventually led to their phase-out and replacement with other compounds, beginning in 1967 and progressing through the 1990s (Litke 1999). During the period of their use, detergent based phosphates would primarily enter surface waters through collected wastewater flows or septic return flow, as opposed to entrainment in stormwater. However, their presence in

automobile detergents provided a direct link to surface runoff and stormwater (just as their presence in marine detergents provided a direct link to surface waters). This stormwater source is unlikely persist following the 1990s phase out of phosphates in detergents, however initial research into stormwater pollutant concentrations (e.g., US E.P.A NURP culminating in 1983) was likely affected by this source.

Pet wastes are another source of developed land stormwater pollutants, based on the nutrients, coliform bacteria, and oxygen demanding substances contained therein (U.S EPA 1993). Early pet wastes management efforts included so called "curbing laws", whereby pet wastes were to be concentrated near curbs where they could subsequently be managed through street sweeping. However, later efforts shifted focus to pet owner collection (i.e., so called "pooper-scooper laws") given the questionable efficacy of street sweeping for pet wastes and the opportunities for stormwater transport of curbside pet wastes between sweeping instances. While pet wastes have been linked to bacterial contamination of surface waters (e.g., Long Island Regional Planning Board 1982), little research or documentation exists on the nutrient content or conditions affecting transport of pet wastes under typical stormwater mobilization scenarios. Given the tendency of these wastes to accumulate within the right of way of densely developed areas (i.e., where stormwater transport is plausible) this potential source warrants consideration.

Another source of nutrients to which elevated nutrient runoff concentrations have repeatedly been linked is lawns or managed turf grass. For example, Waschbusch et al. (1999) identified lawns as having the highest total phosphorus concentrations from a set of eight surface types (pervious and impervious) assessed in two developed Wisconsin

basins, with mean total phosphorus runoff concentrations from lawns of 1.03 and 2.34 mg L^{-1}. Bierman et al. (2010) measured mean annual flow-weighted phosphorus concentrations from constructed turf plots in Minnesota ranging from 0.75 to 4.98 mg L^{-1} under different fertilization and clipping management treatments. Clipping management was found to be a non-significant factor for phosphorus losses, however 80% of phosphorus losses over the three-year period of sampling occurred under frozen soil conditions, suggesting that late fall fertilizer application has a high potential for nutrient losses. Steinke et al. (2013) measured surface runoff and phosphorus losses from prairie and turf grass plots in Wisconsin and found turf grass TP concentrations ranging from 0.02 to 7.43 mg L^{-1}, with annual averages of 1.27 and 1.95 mg L^{-1} in the two study basins. Frozen soil conditions were found to dominate the surface runoff regime, with as much as 99% of annual runoff occurring under frozen soil conditions. This hydrologic dynamic was similar to what was reported by Bierman et al. (2010), and has been reported by others for northern climates (Steinke et al. 2007**;** Timmons and Holt 1977). This suggests that hydrologic management of turf grass systems (i.e., lawns) may be a key consideration in managing phosphorus losses from turf grass portions of the developed landscape.

In the remainder of this chapter I provide a summary of the total nutrient sampling data collected during this research project and place these data in context with previously published studies on nutrient concentrations in stormwater. Four previously compiled sources of data on the topic (three national and one regional) are reviewed, which combined, provide a broad estimate of what is known about total nutrient concentrations

in stormwater. Next, the details of the two studies in this work, namely Englesby Brook and Butler Farms / Oak Creek Village, are summarized in this context. Additionally, I conducted a rainfall-runoff analysis for these study areas to further examine potential factors contributing to the notable TP findings from my research. Lastly, potential causes of differences between these local data and national data sources are discussed, which may help to inform local management of eutrophication in cases where urban loads are thought to contribute.

The focus on total nitrogen (TN) and TP concentrations in this chapter is based on several factors. First, nitrogen and phosphorus are key nutrients that support biological production, and thus important in the consideration of eutrophication (Sterner and Elser 2002; Conley et al. 2009). Second, in many cases the total stock of nutrients (i.e., particulate and dissolved, organic and inorganic) is a better descriptor of the state and function of the ecological system than an instantaneous bioavailable fraction given the dynamic fluxes between these states over ecologically relevant temporal and spatial scales (Meybeck 1982; Howarth 1988; Sterner 2008). Further, total nutrient concentrations are often the focus of regulation and eutrophication management and regulation (U.S. EPA 2010; U.S. EPA 2015), giving them further relevance. Lastly, total nutrients or their respective components were sampled for and analyzed in my research, allowing for a direct comparison in this context.

4.3 Previous Data Compilations

Four compiled data sets have been identified, spanning different geographic and temporal ranges, which have sought to characterize baseline piped stormwater quality. These are the U.S. EPA's Nationwide Urban Runoff Program (U.S. EPA 1983), the Nationwide Urban Runoff Quality Database (Smullen et al. 1999) , the National Stormwater Quality Database (Pitt et al. 2004; Pitt 2011), and the Western Washington NPDES Stormwater Data Characterization (Hobbs et al. 2015). While other localized datasets exist and have been incorporated into the larger national datasets (e.g.,Bannerman et al. 1993; Steuer 1997), the Western Washington data are recent and comprehensive, and thus summarized separately here. The following sections review these data sets with a focus on the characteristics of the included data and on total nutrient results. This in turn provides a basis for comparison of total nutrient data collected in this dissertation's original research with other national and regional summaries.

4.3.1 Nationwide Urban Runoff Program

Perhaps the largest concerted effort to characterize the quality of collected stormwater was the National Urban Runoff Program (NURP), a large scale stormwater sampling initiative intended to establish a baseline for stormwater quality that could inform water quality management decision making (U.S. EPA 1983). The work was carried out at 28 sites across the U.S. over five years, with sampling of urban stormwater outfalls and data analysis managed by USGS and in collaboration with state and local

partners. NURP assessed runoff quality from the diverse range of urban sources contributing flow to the outfall points, including single land use (i.e., residential, commercial) and mixed land use areas. At each of the study sites event mean concentration (EMC) samples of the selected analytes were measured during storm events, and the dominant land cover, population density, and percent impervious cover within the contributing drainage areas were recorded.

Among the NURP EMC data collected, some of the key findings included a high degree of variability and lognormal distributions for most analytes. Relatively high concentrations of metals (e.g., copper, lead, zinc) were noted, relative to ecological thresholds, as well as sediment, oxygen demanding substances, and coliform bacteria levels that could be problematic. In general, NURP analyses were not able to attribute differences in pollutant levels to particular urban cover types within the urban landscape, concluding:

> "*geographic location, land use category (residential, commercial, industrial park, or mixed), or other factors (e.g.,, slope, population density, precipitation characteristics) appear to be of little utility in consistently explaining overall site-to-site variability in urban runoff EMCs…*"

However, NURP did summarize EMCs by contributing land type, and these data have been used on that basis regardless of the lack of statistical differentiation in the data.

This speaks to the high demand for generalized urban land cover EMC data for various management applications and scenarios.

4.3.2 Updating the U.S. Nationwide Urban Runoff Quality Database

Numerous urban runoff studies were completed in the decades following the NURP efforts, which spanned different geographic areas, more recent time frames, and different contributing land surface characteristics. The first formal effort to combine the NURP data with subsequent studies was reported on by Smullen et al. (1999). They combined NURP data with data from the USGS National Urban-Storm-Runoff Database, and available sources of data from National Pollutant Discharge Elimination System (NPDES) permit monitoring. These data were pooled and then compared to the original NURP data to investigate the degree to which more recent data suggest changes in the quality of urban runoff. The resulting analysis suggested some differences in specific pollutant concentrations (e.g., median sediment EMCs appear to have declined significantly in the studies conducted after NURP), but in general the additional data confirmed the log-normally distributed character of many parameters and the generally high variability. Future work was planned to assess the importance of land use characteristics in accounting for variation in pollutant distributions, however it is not clear that those analyses were ever completed or published.

4.3.3 National Stormwater Quality Database

More recently, Pitt (2004; 2011) has compiled stormwater outfall sampling data from a variety of sources to characterize urban runoff concentrations in the National Stormwater Quality Database (NSQD). The largest data source has been the more than 200 regulated Phase 1 Municipal Separate Storm Sewer Systems (MS4s) (those serving populations of 100,000 or more), which were required to collect monitoring data as a permit condition. (By summarizing Phase 1 MS4s sampling data, this analysis best reflects the characteristics of stormwater that drains more highly developed areas.) The later Version 3.1 of the NSQD (Pitt 2011) incorporated the MS4 data, NURP data, as well as data from a few other sources including USGS studies, highway runoff studies, and the International Stormwater BMP Database.

Preliminary analysis of Version 1.1 of the NSQD suggested that there were differences in analyte concentrations among the 11 identified contributing land covers present in the data (e.g., residential, mixed commercial, freeways) (Pitt et al. 2004). However, they stated that additional analysis remained to confirm those differences given the confounding factors. The original database (Pitt et al. 2004) was also summarized in tabular form by contributing land cover, with number of observations, detection frequency, median and coefficient of variation values given for the included analytes. The more recent Version 3.1 NSQD (Pitt 2011) is available as a spreadsheet download and does not appear to have been subject to detailed contributing land use analysis at this point in time.

4.3.4 Western Washington

As a part of their Phase 1 MS4 permit, Washington Department of Ecology prescribed monitoring conditions on permittees with the goal to characterize local runoff quality and inform management objectives. The eight permittees, a mix of cities, counties and ports, were required to implement a robust flow weighted sampling program for a suite of contaminants. Study areas were selected to include primarily low-density residential, high-density residential, commercial, or industrial land uses, with allowances for mixed-use areas where suitable single use areas were not available. The resulting data were analyzed for differences among seasons (i.e., wet season vs dry season) and land uses with the goals of identifying a local baseline and informing the most cost effective management strategies for local water quality concerns (Hobbs et al. 2015).

In a certain respect, these are localized western Washington State data which are best viewed when pooled with other MS4 data, as was done in the NSQD. However, the recent nature of the data (collected between 2007 and 2013) and the robust sampling methods across land uses and detailed reporting and analysis warrant individual consideration here. Key findings from this work included that total nutrients and metals were higher among commercial and industrial land uses than residential land uses, although residential lands tended to have higher dissolved nutrient concentrations. This finding in particular was in contrast to NURP, where differences among contributing land covers were not detected (U.S. EPA 1983). Additionally, Western Washington concentrations tended to be higher in the dry season than in the wet season, suggesting

that dry season management (e.g., street sweeping) could be employed to effectively reduce total annual loads.

4.4 Englesby

As described in Chapter 2 of this dissertation, we gauged flow and collected samples of influent into an extended wet detention pond in the Englesby Brook watershed, in Burlington, Vermont. While our primary goal was to assess the performance of the detention pond, the data collected at the pond inlet provides an opportunity to assess piped stormwater quality from the contributing drainage area. Thus, further aspects of the Englesby study are reviewed and reported here to characterize the quality of stormwater that drains from a mixed land use developed area in Burlington, Vermont.

4.4.1 Study Site Characteristics

The contributing area draining to the discharge point of the study area (i.e., the detention pond inlet) is 47.4 hectares, or 19.3% of the total Englesby Brook watershed area (though water that infiltrates in pervious areas within those 47.4 ha does not necessarily route to the pond). The land cover attributes within the pond's drainage area are predominantly developed, including single and multi-family residential, commercial, and educational / institutional uses (TABLE 1; FIGURE 27). The majority of the lots are single family residential, with a median lot size of 790 m^2 (0.20 acres). Development within the area dates to the late 1800s, but with the majority of the residential development having occurred in the early to mid-1900s. Underlying soils are diverse,

including Hydrologic Soils Groups A, B, C, and D (each at 15% extent or more), while 14% of the area soils are classified as hydric, and 41% of the areas soils are classified as Potentially Highly Erodible or Highly Erodible.

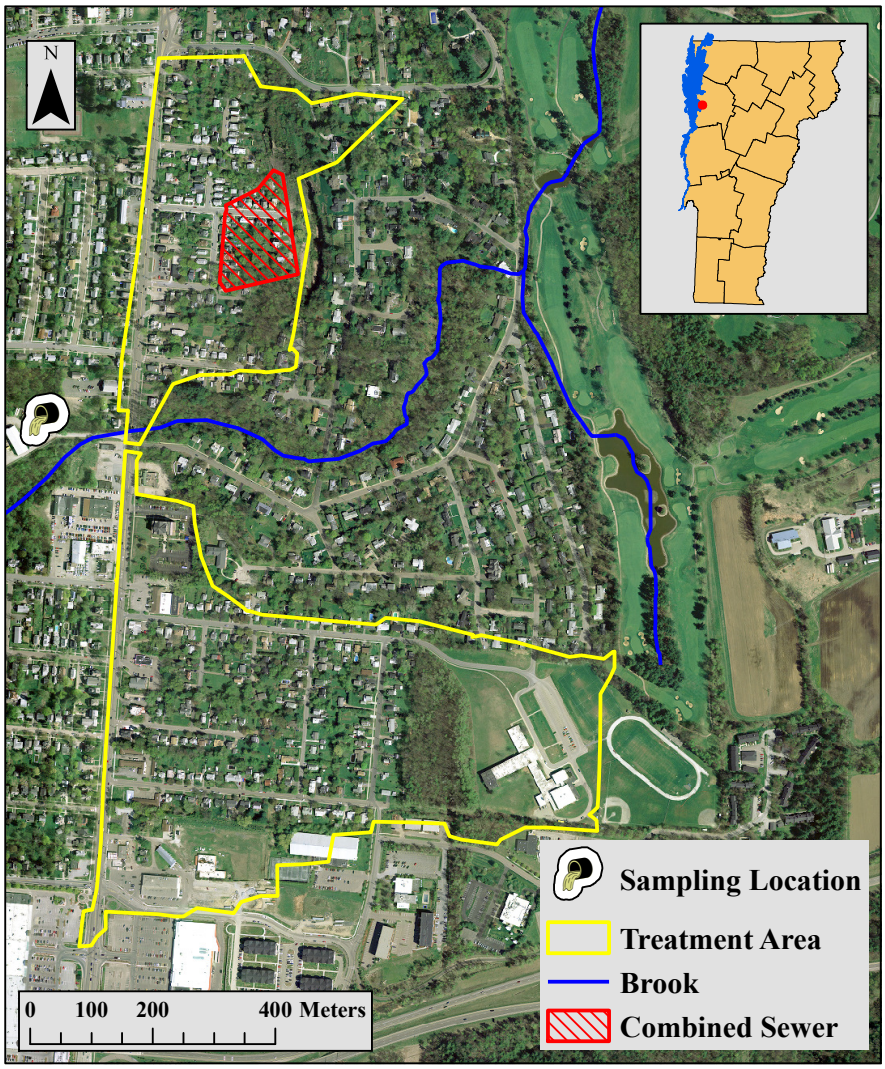

FIGURE 27. Study area contributing flow to the Englesby Brook Detention Pond. Surface runoff from the red cross hatched area is reported to drain to combined sewers, and thus not to our sampling location. (Imagery date is May 2004, downloaded from Vermont Center for Geographic Information.)

Closed conveyance stormwater drainage is believed to have been installed concurrent with development, with discharges direct to the nearby Englesby Brook. In areas with hydric soils or otherwise poor drainage, catch basins are installed in pervious areas providing a direct connection from lawns to the brook (and later, following rerouting, to the study detention pond). Many areas of the drainage also lack curbing or have sunken / deteriorated curbing such that pervious areas extend to the pavement interface (FIGURE 28). This provides an opportunity following extensive winter plowing for road runoff to collect along the disturbed interface of pervious and pavement, and could conceivably result in larger sediment and nutrient loads in spring conditions as compared to a more typical curbing scenario. Alternately, depending on grading, the lack of curbing could allow road runoff to flow on to pervious areas rather than to storm drains, resulting in lower unit area loads.

FIGURE 28. Yard drains installed within the study area, mapped as flowing into the detention pond. Note the lack of curbing. (Photos taken in July 2015.)

While the treatment area falls entirely within the Englesby Brook watershed, it is partially located within the Burlington (35%) and South Burlington (65%) municipalities. Both Burlington and South Burlington are regulated small MS4s with active management programs in place over the period of sampling. Burlington reported sweeping all city streets at least twice per year (but as many as six times) over the period of our sampling (City of Burlington 2008; City of Burlington 2009; City of Burlington 2010). A catch basin cleaning, assessment and repair program was also concurrent with sampling, which would have resulted in catch basins with the drainage area being cleaned (most likely repeatedly) over the period of our sampling. Similarly, the City of South Burlington had an active MS4 management program over this period. Street sweeping was reported to have occurred on all South Burlington streets at least twice a year during the period of sampling (City of South Burlington 2010). Catch basin cleaning occurred throughout the City since 2005, with parts of the study area cleaned in 2009, and other parts cleaned in years prior. Thus, dependent on the timing of the cleaning in 2009, many of the catch basins with the South Burlington portion of the study area may not have been cleaned concurrent or prior to our sampling, or in fact at any time preceding our sampling.

A last drainage area factor worth mentioning is the Lake Champlain phosphorus impairment, which over the past decades has likely increased awareness of water quality issues in ways that could translate to better parcel scale management. For example, the Lake Champlain Basin Program initiated the 'Lawn to Lake' outreach program in 2006, which included a highly visible campaign connecting pet wastes, car washing, and lawn fertilizers to eutrophication and impairment (Lake Champlain Basin Program 2014).

Combined with the outreach work of the Burlington and South Burlington Stormwater Programs (including catch basin stenciling), as well as media coverage of these issues, general awareness of nutrient runoff issues may have been elevated among residents within the drainage area. However, the extent to which awareness of these issues has been increased, and more importantly, what if any changes in behavior are attributable to increased awareness, is not known.

4.4.2 Sampling Setup and Collection

USGS personnel installed monitoring equipment at the inlet of the detention pond in the summer of 2007. Flow gauging equipment was installed inside the detention pond inlet riser, which receives flow from the contributing drainage area via a 0.91 m collector pipe. A Sutron bubble gage was installed within a PVC stilling well, inside the inlet riser. Stage and flow records were recorded at a 5-minute interval. The stage-discharge ratings were developed using a combination of design drawings, site surveys, and temporary weir plates.

Storm event samples were collected using an ISCO 3700 autosampler positioned on top of the inlet diversion structure, linked to and triggered by the flow gaging through a Campbell Scientific Datalogger. Continuous temperature and conductivity readings were also collected at the inlet location. Water sample collection was flow proportional, into a single composite jug for individual storm events. The start of event sampling was either triggered by exceedance of a predefined stage threshold or was triggered manually during a site visit in advance of a storm. The auto-sampler program ran until either the

composite jug was filled, or a site visit occurred for collection of the sample. We made an effort to sample as many storms as possible during the monitoring period, resulting in 43 sampled and analyzed storms over the period of record.

Composite and grab dry weather samples were also collected at the pond inlet. To collect the composite dry weather samples, the autosampler was reprogramed to fill the composite jug via flow proportional sampling over a period of approximately 24 hours. The resulting composite samples included between 45-75 ISCO aliquots per daily sample and were collected on days when it had not rained more than 2 mm in the forty-eight hours preceding the onset of sampling. These 1-day dry weather composite samples were collected once during spring, summer, and fall seasons. Several single grab samples were also collected during summer months. These were collected by positioning a sample container where free discharge entered the forebay at the inlet.

After collection, all storm and dry weather samples were either transported directly to the State of Vermont's National Environmental Laboratory Accreditation Program (NELAP) accredited analytical laboratory for analysis, or were preserved and stored at a University of Vermont laboratory until subsequent transport to the State lab. The collected samples were analyzed for total nitrogen (SM-4500 N C persulfate digestion) and total phosphorus (EPA-4500-P F), both on an unfiltered basis.

4.4.3 Englesby Brook Results

Total nitrogen (TN) and total phosphorus (TP) storm sampling results from the inlet of the detention pond are summarized in TABLE 19. As discussed in Chapter 2, I

attempted to ascribe the variation in inlet concentrations to various potential explanatory variables using best subsets regressions. A large set of predictor variables related to flow, precipitation, and time of year were investigated using linear regression and best subsets multiple regression. The best predictors identified (i.e., month of year, days since storm with peak flow of at least X) were generally of low predictive value for TN and TP event mean concentrations. Many of the sampled events in our data set included what could be identified post-hoc as multiple discrete events. Due to the relatively small area and high connected impervious cover of the treatment area, discrete events at the inlet can occur within periods of a few hours, so that during a day's sampling more than one discretely identifiable inlet hydrograph was often sampled. Thus, a single composite sampled event mean concentration often included multiple discretely identifiable inlet events, each of which would differ in event characteristics (e.g., peak q, time since previous event). This makes it difficult to detect the effect of these predictors.

TABLE 19. Sampling results from pond inlet. Difference in n between Q Peak and nutrient sampling is due to a non-finalized portion of the hydrologic record as of the time of writing.

	ISCO Aliquots	TP ($mg\ L^{-1}$)	TN ($mg\ L^{-1}$)	Q Peak ($m^3\ s^{-1}$)
n	---	43	43	40
Mean	25	0.77	1.55	0.230
Median	21	0.50	1.45	0.140
Geometric Mean	17	0.53	1.38	0.141
Range	2 - 75	0.10 - 3.69	0.53 - 5.20	0.022 - 1.264

Concentrations of TP and TN at the pond inlet included large values that did not group well with the rest of the data set, which is not unexpected given the lognormal distribution of the datasets. For example, the highest two TP concentrations (3.69 and 3.60 mg L^{-1}) were more than double the third highest value of 1.67 mg L^{-1}. Similarly, the highest two TN values (5.20 and 3.63 mg L^{-1}) suggest the characteristics of the tail of the inlet TN distribution, with all other values more closely grouped in the range of 0.53-2.83 mg/L (FIGURE 6; FIGURE 7). The maximum EMCs for TN and TP occurred during the same storm in which 7.8 mm of rain that fell on 2009-03-26, with the composite sample comprised of seven aliquots. As an early spring storm, it is plausible that accumulated winter pet wastes, late fall deciduous organics, and senescent turf grass residuals were mobilized in that event resulting in the high nutrient concentrations. It should be noted it had been a relatively mild month (i.e., warm), with sustained above freezing temperatures and 24 mm of rainfall fell over 4 days earlier in the month (i.e., it was not the first flow event of the year). However, as a relatively small event, the constituent load would have been disproportionately composed of "first-flush" contaminants, which could partially explain the high EMCs.

The second highest TN concentration, 3.63 mg L^{-1}, was measured during a rainfall event on 2009-02-11 (the TP value for the same storm was above median at 1.33 mg L^{-1}, but not among the highest recorded). This was a 25 mm rainfall event that was not closely preceded by a period of warming or rainfall and is typical of what might be expected given a first spring rain on snow event. Finally, the second highest TP value (3.6 mg L^{-1}) was recorded about 6 weeks after the highest value, on 2009-05-07. Two

other storms were sampled in the intervening period with TP concentrations that were above the median but not exceptionally high. Like the 2009-02-11 event, the 2009-05-07 storm was small, while the two events sampled in between had peak flow rates that were roughly double those of the smaller, bracketing events. This pattern is consistent with a first flush producing the high values measured during the smaller storms. However, because we composited sample aliquots within storm events we cannot confirm this assumption.

4.5 Butler Farms

As described in Chapter 3 of this dissertation, we gauged flow and collected samples from two residential storm drain outfalls in South Burlington, Vermont. These data were collected to characterize runoff quality of the Butler Farms and Oak Creek Village neighborhoods and to aid in a pollutant source assessment for Potash Brook Tributary 7. However, the study subcatchments are in many ways typical of single family residential development in Vermont and thus can provide some insight into likely runoff concentrations for similar types of development. Here I review the data we collected in greater detail to provide additional information to the relative paucity of localized urban runoff data sets.

4.5.1 Study Site Characteristics

The two instrumented drainage areas are referred to herein as the 'East Drain' and 'West Drain', based on their position relative to the Potash Brook channel. These

drainage areas are shown in FIGURE 29, with summary attributes provided in TABLE 20. Both drainages are small, encompassing parts of 16 or fewer residential lots each. Lots within the instrumented portions of the neighborhood were constructed between 1987-1999, atop underlying soils mapped as Vergennes Clay and Covington Silt Clay (hydric), both of which are classified as poorly drained with a high runoff potential. The residential lots include fill of unknown origin to create positive drainage and to allow for residential construction in low-lying hydric soils. Lot drainage to the street flows to catch basins connected to a closed drainage system discharging directly to Potash Brook (FIGURE 30).

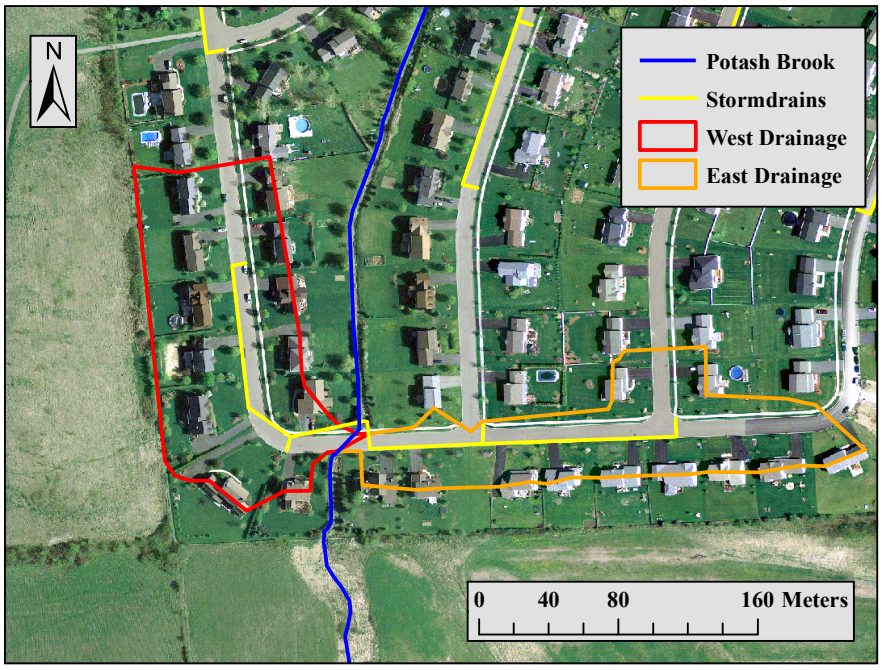

FIGURE 29. East Drain and West Drain drage areas. (Imagery date is May 2004, downloaded from Vermont Center for Geographic Information.)

TABLE 20. East Drain and West Drain drainage area characteristics.

	East Drain	West Drain
Area (ha)	1.24	1.61
Partial Lots	16	12
Median Lot Size (m^2)	1,674	1,396
TIA (%)	49.4	41.4
DCIA (%)	46.6	33.8
Canopy Cover 2004 (%)*	1.2	1.3

*** Hand digitized from 2004 color orthophotos.**

Over the period of sampling (2008-2009), the storm drain system for the study area was managed by the South Burlington Stormwater Utility as a part of its regulated MS4 system. This resulted in street sweeping in the spring and fall of both years of sampling (as well as the year preceding sampling), and catch basins in the study area were reported to have been cleaned in 2008. Other MS4 management (e.g., outreach and education) described in Section 4.4.1 (Englesby Brook) occurred similarly in this area.

An additional unique set of circumstances within the study area was the combination of 'permit' issues for the neighborhood, and the involvement of the Redesigning the American Neighborhood (RAN) Research Program. The permitting issues stem from the neighborhoods having had a stormwater conveyance / management system and an associated stormwater discharge permit at the time of initial construction, but neither the system nor the permit were maintained over the following decades. By 2004, regulatory changes and the 303(d) listing of Potash Brook for stormwater impairment had made it impossible to bring the permit into compliance within the existing regulatory framework, creating issues for real estate transactions in the

neighborhood that may have incentivized residents to engage with the issue of stormwater management.

Over the same period, a collaborative effort among University of Vermont researchers, City of South Burlington officials, and other stakeholders was initiated with a focus on the Butler Farms and Oak Creek Village neighborhoods (McIntosh et al. 2006). Full research goals and rationale are contained in grant work plans, however several aspects are directly relevant to land management within these study neighborhoods. For example, RAN outreach efforts included the creation of a stormwater workgroup to bring together residents, researchers, municipal officials, and others to discuss stormwater management in the context of the neighborhoods. Survey work was conducted among residents to gage background knowledge and attitudes about stormwater, as well as to assess preferences for various management scenarios (Kofstad 2011). Various information modules about general stormwater effects and treatment options were compiled on the RAN website, as well as stormwater content specific to the neighborhoods. Lastly, two Stormwater Field Days were held at the neighborhoods, giving stakeholders an opportunity to discuss on site their concerns and alternative approaches to stormwater management. Cumulatively, these actions could be expected to have raised both general awareness and knowledge about stormwater issues among residents.

FIGURE 30. Looking east in the East Drainage (left) and northwest in the West Drainage (right). (Photos taken in July 2015.)

4.5.2 Sampling Setup and Collection

As previously described in Chapter 3, flow measurements and water quality samples were collected from within two piped stormwater outfalls (i.e., East Drain and West Drain, FIGURE 29). In each location, vented In-Situ Level Troll 500 pressure transducers were mounted with ISCO mounting rings inside of the existing 47 cm diameter PVC storm pipes, with the pipes discharging via flared concrete aprons into the brook. Flow rates in the pipes were calculated as a function of depth using Manning's equations with a variable roughness coefficient as described by Wong and Zhou (2003). The pressure transducers connected to ISCO 6712 auto-samplers positioned on the stream bank above the outfalls. Storm water quality samples were collected using a flow-weighted sampling program, triggered by stage exceedance and stage rate of change thresholds. Samples were collected by sample intake lines affixed to the cable conduit (protective shielding for the transducer cables), approximately 0.25 m downslope of the sensors, near where the storm pipes terminated at the concrete aprons. In addition to the

composite storm samples, we collected samples during one storm using 24-bottle kits within the auto-samplers, allowing us to sample analyte concentrations throughout an event. All collected samples were transported to a commercial lab for analysis, where they were analyzed for total suspended sediment (TSS) (EPA 160.2), unfiltered total phosphorus (TP) (EPA 365.1), nitrate (EPA 300.0), nitrite (EPA 300.0, discontinued following repeated non-detects), total Kjeldahl nitrogen (EPA 351.3 / 350.1), and chloride (EPA 300.0).

4.5.3 Butler Farms Results

Total nitrogen (computed as total Kjeldahl nitrogen plus nitrate plus ½ nitrite detection limit) and TP storm sampling results from the East Drain and West Drain are summarized in TABLE 21. It can be seen that a greater number of samples were collected in the West Drain, which was a result of the physical configurations of the outfalls. That is, at the West Drain we were able to better collect samples at low flow due to a slight gap between the terminating storm drain PVC and the entrance to the concrete apron, which created a small depth of water downstream of the sensor. The ISCO strainer used to collect samples was 32 mm diameter, though it did not need to be entirely submerged to collect a sample. However, this small depth of water at the West Drain resulted in substantially fewer errors caused by 'No Liquid' at the sampling strainer compared to the East Drain. Given the very flashy timing of these systems, the failure to collect a sample (including an attempted line rinse cycle and second attempt at collection) can result in the rising limb and peak of the first storm pulse passing with no sample

collected. These circumstances led us to discard collected samples for some storms due to the poor flow weighted coverage of the hydrograph and the costs of sample analyses, and is the primary cause of the discrepancy in the number of storms analyzed in the East and West Drains (TABLE 21).

TABLE 21. Sampled storm attributes and nutrient results for East Drain and West Drain storm sampling.

	ISCO Aliquots	TP (mg L^{-1})	TN (mg L^{-1})	Q Peak (m^3 s^{-1})
East Drain				
n	---	11	11	11
Mean	19	0.071	1.33	0.046
Median	11	0.073	1.33	0.042
Geometric Mean	13	0.063	1.27	0.041
Range	3 - 38	0.030 - 0.160	0.73 - 1.99	0.010 - 0.989
West Drain				
n	---	19	19	19
Mean	19	0.102	1.44	0.033
Median	21	0.084	1.09	0.025
Geometric Mean	15	0.087	1.30	0.025
Range	3 - 38	0.034 - 0.240	0.57 - 3.32	0.006 - 0.092

Among the data collected, there was a greater range in EMCs for TP at the West Drain, though this is mostly attributable to a number of storms with high TP EMCs but for which samples were not collected and analyzed in the East Drain. There were four West Drain storms where TP concentrations of 0.200 mg L^{-1} or more were measured, while all other West Drain concentrations were 0.103 mg L^{-1} or lower. The four West Drain storms with 0.200 mg L^{-1} and greater included a storm with a long antecedent dry period (14 days) and the sampled storm with the highest peak flow rate, although not all

of these storms included high flow rates or long antecedent dry periods. For these same storms, only one of the four was sampled and analyzed at the East Drain. That storm at the East Drain had the highest East Drain TP concentration on record (0.160 mg L-1) and had the third highest East Drain peak flow rate on record (0.076 m^3 s^{-1}), although the dry antecedent period was not especially long (~5 days). This all suggests that peak flow rate and antecedent dry time have an effect, although the small size of these data sets limits a more extensive analysis.

For TN, concentrations at both sites appear to come from relatively continuous distributions, without apparent outliers. Values were generally higher in the West Drain, although the highest values were not consistently differentiated by higher peak flow rates or antecedent dry periods. Lastly, it is worth noting that the higher TN values at the East Drain were associated with higher TN values at the West Drain, suggesting that similar factors were important (i.e., a single outlier lot was not responsible). However, for TP the same cannot be said due to a failure to collect and analyze East Drain samples for many of the West Drain storms that produced high concentrations.

4.6 Comparison of Englesby and Butler Farms Data with National and Regional Datasets

The various data sources considered in this chapter have been summarized in TABLE 22. It can be seen that for TN, the data reported here for the Englesby and Butler Farms sites were generally in the range of reported values from studies conducted in other locations. Given the previously reported high variability in urban runoff

concentrations (U.S. EPA 1983; Pitt et al. 2004) and the relatively small number of samples in this work, I don't consider the difference noted in TN in this work versus national data to be meaningful. However, the differences in TP concentrations reported in this study versus previously reported national data are notable and warrant additional analysis. Again, due to the high reported variability in urban runoff data it is not entirely unexpected to encounter this variability among sites, and between local and national data. Nonetheless, any greater understanding of the factors responsible for this variability could inform local management efforts.

TABLE 22. Total nutrient concentration data from previous work, and from the studies reported on in this work.

	Median Urban Runoff EMCs	
Reference	**TN (mg/L)**	**TP (mg/L)**
Previous Studies		
NURP (1983)	2.18*	0.266
Smullen et al. (1999)	2.0*	0.259
Pitt et al. (2004)	2.0*	0.27
Western Washington DOE (2015)	1.18*	0.11
This Research		
Englesby Pond Inlet	1.41*	0.498
West Drain	1.09*	0.084
East Drain	1.33*	0.073

*** TKN + NO_2 + NO_3**

Consideration of the factors potentially driving TP EMCs in this work focused on land cover attributes and drainage characteristics. A number of physical characteristics of the study drainage areas, including the presence of yard drains within the Englesby drainage and the relatively recent grading and engineered drainage within the East and

West Drains, suggest that greater lawn contributions were responsible for the elevated Englesby TP concentrations. Previous research has measured TP in stormwater runoff from lawns as being in excess of 1 mg L^{-1} (Bannerman et al. 1993; Bierman et al. 2010; Garn 2002; Steinke et al. 2007; Steinke et al. 2013; Waschbusch et al. 1999), and thus this dynamic would not be unprecedented.

Because we measured flow and nutrient EMCs at pipe outfalls of mixed land use and mix land surface drainage areas, the role of developed pervious area as a driver of variation in TN and TP concentrations cannot be directly assessed. Instead, I resort to an indirect means of investigation as described by Boyd et al. (1993). This methodology uses a post hoc event based hydrologic analysis to infer the types of watershed area contributing flow. To do so, a period of hydrologic record (runoff and rainfall) is reduced to the series of discretely identifiable storm events. Area normalized runoff and rainfall depths are then plotted in a scatter plot for interpretation. If storm runoff for all events is generated solely by an effective impervious subset of the watershed area, then all points will fall along a straight-line that can be interpreted as having a slope equal to the effective impervious watershed fraction and an intercept equal to initial watershed abstraction. Alternately, points that deviate from the straight line can be attributed to storms occurring under wet antecedent conditions (i.e., initial abstraction met by preceding event), storms that generate runoff from areas other than the effective impervious fraction (e.g., pervious area runoff), and error.

There are a number of potential difficulties in applying these interpretations to watershed runoff data, including hydrologic error in the rainfall and runoff terms,

handling of non-surface runoff, conceptual over simplification of watershed responses (e.g., gravel or earthen drive surfaces which may produce runoff at a frequency between that of hard impervious and previous), and subjectivity in the classification of individual events. However, many of these difficulties are significantly reduced when this analysis framework is applied to relatively small areas where flow is collected and routed through engineered conveyance, as opposed to whole watershed stream response. Thus, it has potential applicability for considering the surface types contributing flow to the sampled storm flow and nutrient loads measured in this research.

A collection of R (statistical programming language) scripts were developed for this task, which allowed for relatively fast analysis of seven years of site / event flow records to be considered (i.e., two years of East Drain, two years of West Drain, three years of Englesby). In brief, the R scripts first use a collection of user defined parameters to identify flow events based on a combination of peak rate detection, followed by beginning and end of event detection based on rate of change. The resulting hydrographs and hyetographs are next visualized and edited using a set of command line scripts which allow the user to quickly add and delete events from consideration, and to precisely edit the start and end times for the discretely identified storm events using arithmetic and log transformed hydrographs. Lastly, using the resulting table of start and end times for runoff and rainfall, event rainfall and runoff totals as well as other event based statistics of interest were computed by an additional script developed for this analysis.

Graphical results from the East Drain event identification analysis are presented in FIGURE 31. From the 2008-09 period of hydrologic record, 97 discrete rainfall runoff

event pairs were identified. It can be seen that many rainfall events less than 5 mm depth produced very little runoff response in the storm drains. Thus, initial abstraction for the area can be interpreted as being approximately 4 mm. It is important to note that small rainfall depths producing no runoff response are not captured by this analysis, and thus it will at best identify the upper bound on initial abstraction. Except for a subset of four storms that occurred at or after mid-November, all East Drain storms we sampled were among those falling at the low end of the scatter plot, with respect to runoff depth. This suggests that these storm flow volumes were produced by an effective impervious subset of the watershed area that produces runoff linearly as a function of incident rainfall. It should be noted that the three highest values of TN and TP EMCs that were measured in the East Drain (not specifically annotated in FIGURE 31) were among the non-winter storms falling at the bottom of the graph. That is, the relatively low maximum nutrient concentrations in the East Drainage set (as compared to national data) are likely to be due to effective impervious surface runoff alone.

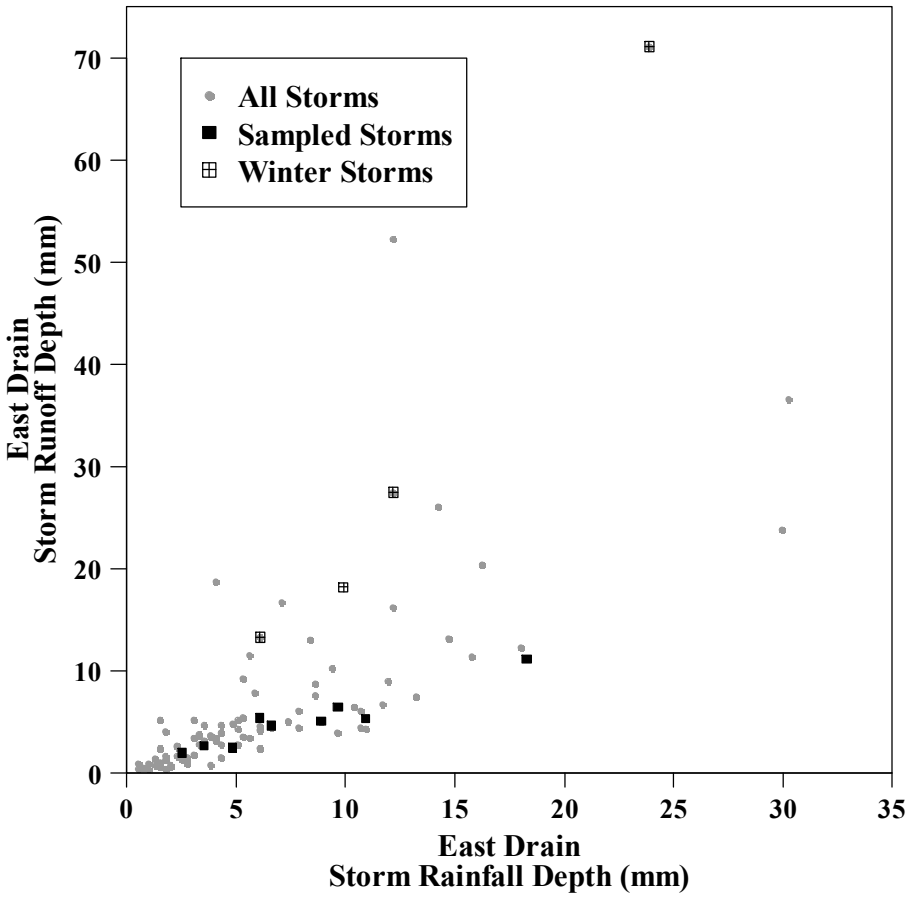

FIGURE 31. East Drain event identification analysis. The three highest TP and TN EMC samples were among the non-winter storm set.

Results from the West Drain event identification analysis are shown in FIGURE 32. A total of 156 discrete rainfall runoff events were identified and plotted, exhibiting a relationship similar to that observed in the East Drain. Many of the rainfall events up to ~3.5 mm of rainfall depth resulted in little to no runoff, which can be interpreted as an estimate of initial abstraction. Many of the rainfall runoff points fall along a relatively well defined bottom edge of the scatter, which can again be interpreted as storm flows resulting from the effective impervious subset of the drainage area. Also highlighted are the four sampled storms with TP concentrations greater than or equal to 0.200 mg L^{-1}, as

compared to the rest of the West Drain data set. In three of four cases, these storms did not fall along the interpreted effective impervious response region. Rather, these storms had close to 1:1 runoff to rainfall ratios, suggesting all areas of the watershed contributed flow. This provides circumstantial evidence that the ≥ 0.200 mg L^{-1} TP EMCs were driven by previous area response (i.e., residential lawns in this case), whereas the primarily impervious driven storms that were sampled were generally of lower TP EMC.

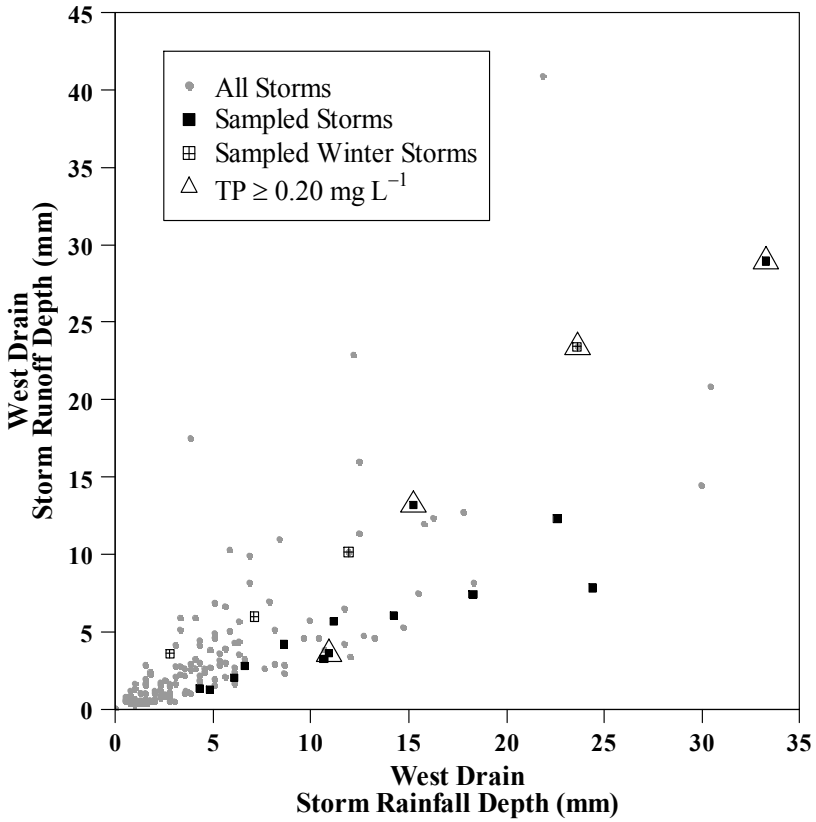

FIGURE 32. West Drain event identification analysis. Sampled storms occurring in winter conditions (i.e., mid- late-November and or December) and storms with high TP concentrations are highlighted.

Lastly, it should be noted that the highest (though not exceptionally high) TN values from the West Drain were not recorded for the same storms with the highest TP concentrations. Rather, they were generally measured during what are interpreted here as impervious flow storms and were not easily attributable to runoff ratios, peak flow rates or antecedent periods.

For the Englesby detention pond area, this analysis was complicated by the availability of commensurate rainfall data. For the East Drain and West Drain sites, no point in the contributing drainage areas was more than 304 m from the tipping bucket rain gage maintained for the period of record. This produced excellent visual correspondence between storm hyetographs and hydrographs, facilitating this analysis. For the Englesby Brook analysis, I relied primarily on data collected for Englesby Brook by the UVM Flow Monitoring Project. This rain gage was maintained within the Englesby Brook watershed to the northeast of the study drainage area and was at most 1.7 km from any point within the study drainage area. For many of the storms in the flow record these hyetograph data corresponded well with the shape of the runoff response. However, for many storms, including most of 2009, these either did not correspond in shape to the runoff record (e.g., a single rainfall tip, but a multi-peak hydrograph response) or were missing. Attempts were made to incorporate precipitation data from Butler Farms, other nearby FMP sites, and data from Burlington airport, which agreed well in some cases. In general, I excluded from this analysis any storms for which there was not a satisfactory, matching event hyetograph. The exception was for storms that were sampled for water quality, which importantly included winter storms and many

early spring storms before the FMP rain gage was operational. For these storms, I relied on less precise hourly Burlington Airport data which was collected ~4 km to the northeast. These data generally lacked the close event shape correspondence that could be seen in the hydrograph and hyetograph data using 5 min rainfall data. However, importantly, the Burlington Airport hourly data allowed this analysis to be extended to the many late fall / winter / early spring storms that were sampled and upon which the following analysis relies.

Using this approach, I was able to include 208 storm events from the period of record in this analysis (FIGURE 33). Similar to the observations from the East Drain and West Drain at the Butler Farms study site, the event based rainfall runoff-analysis for the Englesby site showed a mix of storms that fell along a line that could be inferred as the effective impervious response region and other storms that had higher runoff ratios that could be inferred to have larger contributing runoff areas. The storms sampled during winter months tended to have high runoff to rainfall ratios compared to the larger data set, which is consistent with meltwater runoff and / or frozen ground conditions increasing the land contributing surface flow.

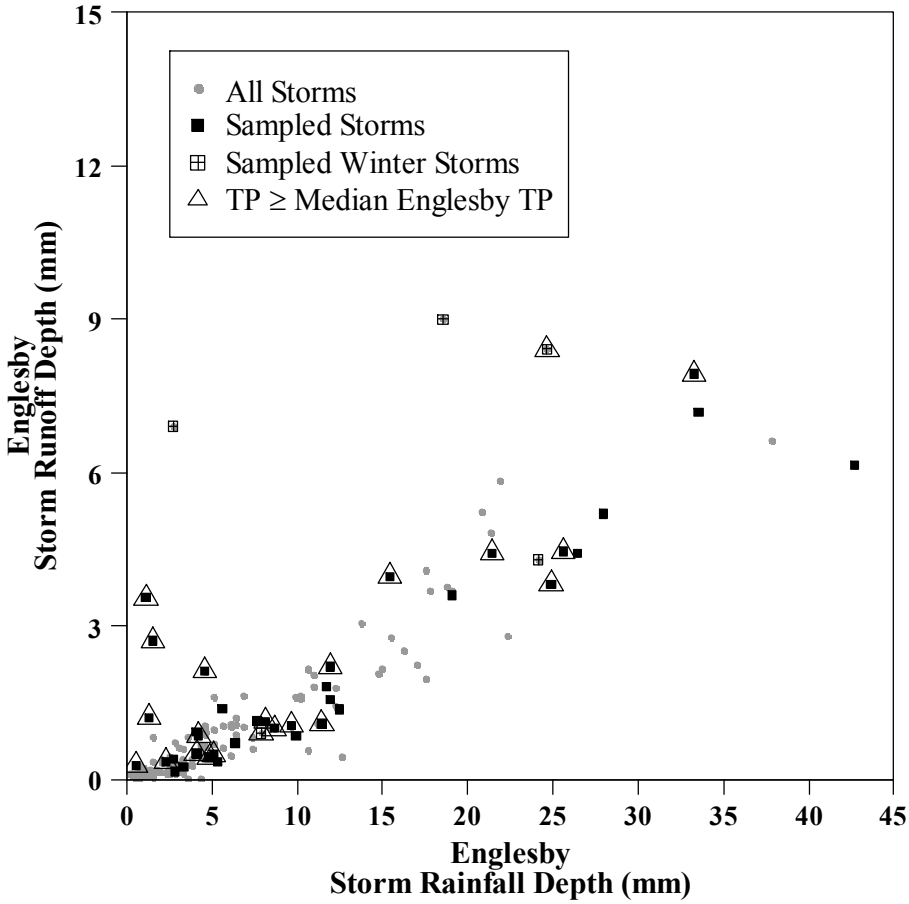

FIGURE 33. Englesby pond inlet event identification analysis. Sampled storms occurring in winter conditions (i.e., mid-November through March) and storms with high TP concentrations are highlighted.

Nutrient results are focused on the TP data collected since those were generally higher than data reported elsewhere. As discussed in Section 4.4.3, the two highest TP concentrations occurred during relatively small storms (less than 10 mm rainfall) in early to mid-spring. Thus, from a rainfall-runoff analysis perspective the storms producing these concentrations do not stand out. FIGURE 33 highlights the sampled storms with TP concentrations greater than or equal to the median value from the Englesby data set. There is no clear pattern in these data, with median or greater TP values measured during both small and large storms, winter and non-winter storms, and high rainfall runoff ratio

and low rainfall runoff ratio storms (though the use of median Englesby TP as a threshold is arbitrary). This suggests there are elevated TP loads within the contributing drainage area that are mobilized during a variety of event conditions over the year.

As previously discussed, the presence of yard drains and lack of curbing in parts of the drainage area provides an opportunity for fertilizers, grass clippings, pet wastes, and other materials to flow off of residential lawns into storm drains. Previous research has documented that lawns can produce TP runoff concentrations that are among the highest in the urban landscape (Bannerman et al. 1993; Bierman et al. 2010; Garn 2002; Steinke et al. 2013), so to the extent that these areas were routinely producing surface runoff within the Englesby study area, it is plausible that hydrologically connected lawns account for the elevated TP concentrations measured in this work.

Canopy cover is another factor that has been identified as having a positive correlation with TP concentration in runoff (Waschbusch et al. 1999). The Englesby study area had canopy cover of approximately 30% (TABLE 1), compared to less than 2% at Butler Farms (TABLE 20), which is in general agreement with the correlation identified by Waschbusch et al. (1999). A final possible source of elevated TP concentrations considered was illicit sewage connections or subsurface inflow of sanitary loads into the stormwater system. However, these sources are not likely to be major factor based on the low TP values sampled at the pond inlet under dry weather conditions when wastewater inflows would be at their maximum proportion of total inflow (TABLE 3).

4.7 Conclusions

Since the first intensive stormwater quality assessment (NURP), stormwater quality has been known to be quite variable. The data collected and reported on in this dissertation are consistent with that understanding. TN and TP concentrations at Butler Farms were generally lower than median estimates from other studies. I attribute this to parcel scale drainage patterns and the subset of storms sampled, such that most sampled storms appeared to represent directly connected imperious runoff response only. Given the relatively well maintained pavement surfaces throughout the Butler Farms drainages and the relatively low traffic levels (i.e., not through streets), the potential sources of nutrients in runoff are limited. A small number of storms from the West Drain sampling appeared to have pervious area hydrologic contributions, and had elevated TP concentrations as well. This circumstantially supports a case that residential lawns are a relatively large source of TP among residential surfaces, consistent with prior research.

That even the high Butler Farms TP concentrations were relatively low is more difficult to explain. It is possible that UVM / RAN educational outreach and other water quality outreach efforts have resulted in lawn care and pet waste habits within Butler Farms that produce low TP concentrations. Another possibility is that the relatively new fill at Butler Farms is well-drained, has high phosphorus sorptive potential, or both. However, these explanations are entirely speculative given the lack of data about the fill characteristics or residential lot management practices.

For the Englesby dataset, the manner in which the samples were collected limits the conclusions that can be drawn about factors responsible for the high TP values that

were measured. I speculate that based on the event based rainfall runoff analysis (and the obvious presence of yard drains), the Englesby dataset includes a high level of lawn runoff with elevated TP runoff concentrations. The deciduous canopy cover, smaller lot sizes (potentially leading to increased pet wastes in the right of way, and also placing potentially fertilized decorative gardening closer to the right of way), lack of or general disrepair of curbing, and age of lawns (and potential exhaustion of phosphorus sorptive capacity) can all potentially explain the greater TP losses. These are, however, entirely post-hoc explanations that were not directly assessed or controlled for in this research.

CHAPTER 5 CONCLUDING REMARKS

Under the overarching theme of stormwater quality, modeling and management, the original research in this dissertation has generated several useful findings and suggested areas for further research. These cumulative findings can generally be categorized under the headings of SWMM Modeling, Loading Analysis, Modeling Framework, Detention Pond Performance, and Nutrient Concentration Data, and are summarized in the following sections.

5.1 SWMM Modeling

There are several findings from the SWMM modeling reported in Chapter 3 that are likely to be of value to others within the SWMM users community. The inclusion of the subsurface hydrology components within the global sensitivity analysis (SA) and the resulting insights are among the findings expected to be most useful for others. While previous SAs have included the surface hydrology parameters (with generally comparable results to those reported here), the subsurface components have not previously been reported on in the published literature. The subsurface model includes many conceptual or practically immeasurable parameters, making it difficult to manually calibrate or assess the sensitivity using the deterministic SWMM graphical user interface. The results of Chapter 3 demonstrate the sensitivity of simulated watershed runoff to the depth of the subsurface reservoir, the free draining void space (computed from porosity and field capacity), and the upper evaporation fraction, which is likely to provide useful

information for other modelers that cannot be easily gleaned from reviews of the algorithms used. Existing guidance on this topic is to my knowledge limited to a series of narrative descriptions of parameter importance that have been posted by experienced modelers to the SWMM user's listserv over time in response to questions. Importantly, that narrative guidance does not necessarily agree with the findings of this SA. Thus, the publication of Chapter 3 in this dissertation and the subsequent peer-reviewed publication of this work is likely to find an audience among the SWMM community.

Another key finding of this work is the additional confirmed usefulness of the Guo Method for transforming the shape of a natural irregularly shaped watershed to a rectangular kinematic plane representation. While Guo and Urbonas (2009) published the initial dimensional analysis and findings, and Guo et al. (2012) demonstrated that the approach worked on several real watersheds, it does not appear to have been adopted and published outside of Dr. Guo's research group. Thus, the independent use and robust confirmation in this work showing that a priori Guo estimates were virtually indistinguishable from computationally expensive ES calibrated values, albeit on two small homogenous subcatchments, may further validate this approach among the SWMM community.

A final SWMM modeling conclusion that should be mentioned from my work relates to the Buildup and Washoff (Bu/Wo) algorithm calibration and validation. Initial results from calibration runs of the Bu/Wo algorithms using composite EMC data were generally unsuccessful given that a wide range of pollutographs (including unrealistically high peaks) can compute to the same EMC. Using the limited discrete sampling data

collected in this study, I was able to constrain these parameterizations to a degree while reserving part of the discrete data set for model validation as modeling best practices dictate, and as described in Chapter 3. However, additional exploratory analysis suggested that ignoring the need for discrete data validation (i.e., simply calibrating against the full available record) can greatly increase the accuracy and precision of modeled water quality concentrations and loads. Composite EMC data have several advantages, including reduced analysis costs and relative simplicity in sampling when using an autosampler. However, it is clear from the results of this work that a single discrete event pulse for calibration (even combined with a larger set of composite sampled loads) may be insufficient to narrow the parameter space.

In hindsight, a greater number of storms should have been sampled by discrete means to improve the usefulness of the sampling data for this task. Nonetheless, calibration against the full data set (albeit without validation data to compare against) shows that good agreement can be found with just two discretely sampled storm pulses, as were collected in this work. This suggests there may be an acceptable compromise between the approach of this work and a complete discrete bottle approach to sampling that maximizes usefulness while also limiting the cost and labor associated with operation of (in this case, four) automated, discrete bottle samplers.

The most useful extension of this work into future research would be an extension of the SA approach used herein (or a similar global approach) to stormwater treatment practices represented in SWMM, particularly the new LID (low impact design) features that have been incorporated. This additional functionality has introduced new parameters

that have not yet been assessed for sensitivity, and thus the SWMM user community would strongly benefit from a close examination. However, given that the LID functionality has been frequently updated with additional functionality and bug fixes, and as of yet lacks comprehensive documentation, it may be premature for such an assessment at this time.

5.2 Loading Analysis

The SWMM modeling just discussed was used to facilitate a measurement and modeling load analysis in the Potash Brook tributary study area. This analysis suggests the stream channel is often a sink for sediment at the event scale and may be the source of unaccounted loads for other events. A comparison of unaccounted TSS loads with event peak flow rates at the lower monitoring station provided further support for this interpretation by linking increased unaccounted TSS loads with increasing peak flow rate at the SW2 (lower) monitoring station. The TP analysis was suggestive of an unaccounted for source of TP which I speculate is derived from channel erosion, however I lack additional circumstantial evidence for this conclusion due in part to the non-significant linear relationship between peak flow rates and unaccounted TP loads. Analysis of the TN data suggested the most closed mass balance of the three analytes considered, although there were exceptions where considerably more TN was measured at the lower monitoring station than could be accounted for by SW1 (upstream) loads and neighborhood washoff.

Overall, there are limits on the conclusiveness of these findings due to uncertainties in the Bu/Wo model and the fact that I did not measure channel effects or golf and meadow inputs directly. However, plausible physical mechanisms exist to explain the dynamics discussed, including channel erosion, channel retention, autumn leaf inputs and denitrification losses. Thus, these findings can be used in localized watershed scale management on that basis, or as the basis for further work into these dynamics.

5.3 Hybrid Monte Carlo and Evolution Strategies Modeling Framework

The hybrid modeling framework I used, whereby uncertain parameters are either fixed at best estimates, subject to Monte Carlo sampling, or calibrated by (in this case) using the Evolution Strategies algorithm, all informed by the SA results, is not limited in its applicability to SWMM. Rather, it draws on previous work on parameter uncertainty analysis and watershed calibration, and could be equally applicable to other modeling contexts. Specifically, the idea that the uncertainty in some parameters must be manifested as noise in the model parameter space, while other uncertain parameters can be calibrated to improve model performance addresses a fundamental issue present in models with uncertain parameters. While this specific approach is presented as new in this work, the volume of modeling literature precludes a definitive conclusion of the same. However, to my knowledge the specific approach I took to this complex problem has not been employed previously and may be of use to others.

An additional observation that should be made regarding this work stems from the relatively precise simulated flow estimates that were produced by this approach. That is, while the hydrologic model objective function values were not excellent in most cases, the range of estimates produced by the model were relatively small given the robust approach to extrapolating parameter uncertainty through the model. A perceived disadvantage to uncertainty analysis in general has been the large uncertainty bands often generated in hydrologic applications, and the resulting difficulty in using highly uncertain predictions for management (Beven 2006). In this work, the high resolution local rainfall data coupled with a relatively simple runoff response mode (i.e., impervious dominated) seems to have effectively constrained the model predictions. (Although it should be noted that I did not explicitly account for uncertainty in input forcings (i.e., rainfall and ET) or model structure, and thus inherently underestimate total prediction uncertainty.) That the impervious surface conceptual model of SWMM closely matched the runoff dynamics of the studied system certainly should have helped to constrain predictions as well. In other watershed modeling contexts, for example using highly parameterized / distributed models, the approaches I employed in this project would likely produce very wide uncertainty bands in many cases, limiting the perceived value of model results for management purposes. While this could be viewed to render a model useless, it can also highlight limitations in input data and process understanding, thereby providing a basis for additional research and model refinement.

5.4 Detention Pond Performance

Detention ponds have been frequently studied, however, as discussed in Chapter 2 much of the research on detention ponds is at this point in time dated and not entirely reflective of current design practices. The recently constructed detention pond that was the focus of this study was found to perform better than average, relative to the many previously studied ponds, despite the relatively high TP concentrations flowing in. This suggests that modern design features, such as forebays, large permanent pools, and long flow paths may add to a pond's ability to reduce total nutrient loads. However, this single pond is insufficient to isolate those factors or make generalized conclusions about the efficacy of various pond features.

Sediment is frequently a stormwater contaminant of concern, both as an impairing pollutant and as a surrogate for other pollutants. An effort was made in this work to collect discrete suspended sediment concentration (SSC) samples and to record continuous turbidity measurements with our USGS partners towards development of a turbidity-SSC rating curve which could estimate sediment fluxes on event and seasonal scales. However, maintenance of the turbidity probe was problematic given the condition of the influent, and thus the resulting data were not usable in this manner. It is not known how reflective these circumstances are of other influent and turbidity meter pairings. Given the conditions encountered in this study, it may be that others should not rely on measuring sediment via turbidity and either dedicate an entire second sampler for parallel SSC sampling or simply use TSS subsampling with the reduced accuracy that entails.

This is not the first study to report elevated stormwater runoff temperatures or the ability of stormwater treatment permanent pools to accumulate heat during warmer months. However, it does highlight the need to be attentive to this issue in Vermont and elsewhere. Under Vermont's current stormwater regulations, permanent pool detention ponds are quite common as a multi-purpose treatment strategy that can be configured to contribute toward or meet the Water Quality, Channel Protection (1-year), Overbank Flooding (10-year), and Extreme Flood Protection (100-year) standards (i.e., all regulatory standards but the Groundwater Recharge Standard). This has led to the construction of hundreds if not thousands of these throughout the state. There is a thermal mitigation design feature, in that the length of time the 1-year storm volume must be detained on average is reduced for sites draining to cold water fisheries from 24 hours to 12 hours. It is worth noting that the design of the pond studied in this work provided only 4.6 hours of detention, and that this pond is somewhat of an outlier in many respects as a large municipal retrofit. Nonetheless, it seems possible that for cold water systems a moderate development density with extensive use of stormwater management ponds could produce thermal alteration that would affect biota. Current state efforts to rewrite the Stormwater Management Manual governing treatment standards for regulated sites is attentive to this dynamic, but would likely not require substantial retrofitting of the many existing ponds.

Lastly, the analysis of the TMDL flow metrics using different temporal resolution data highlighted the sensitivity of these metrics, and thus stormwater management, to the resolution used. It is of course obvious that highly aggregated flow records (e.g., daily

flow) might be used in a large river basin whereas smaller drainage areas (e.g., storm drain subcatchments) would warrant higher resolution data to capture flow dynamics. Intuitively, this can be linked to drainage area, time of concentration, watershed lag, or other hydrologic concepts (Dingman 2002). However, operationally this creates a challenge in Vermont's stormwater impaired watersheds. For TMDL development, watershed flows were modeled at an hourly resolution and then aggregated to mean daily flows for calculation of the TMDL metrics and inter-watershed comparison. Management practices, however, are most often designed at the parcel scale where a much finer temporal resolution may be warranted.

For instance, the design of a treatment practice at different temporal precisions for a 'typical' 0.40 ha parking lot with a time of concentration of 4.4 minutes is illustrative. Using the USDA NRCS TR-20 runoff approach for the Chittenden County regulatory design storm depths (Soils Conservation Service type II distribution) at 1-hour time step produces a peak runoff rate of 11.6 L $s^{-1}$for this parking lot. Routing this flow through a typical regulatory grass channel for water quality treatment, a designer would need 107 m of grass channel to meet the 10 minutes of residence time that would be required under current regulations. When remodeled at a 1-minute time step the resulting peak flow rate from the parking lot increases to 38.2 L s^{-1}, which is only detained for 7 minutes when routed through the same grass channel. At 1-minute temporal resolution the grass channel would need to be extended to 137 m to meet regulatory requirements. A similar analysis extended to water quality filters, which are often used at sites with space constraints, would scale similarly. Other detention based treatment practices are less

easily generalized, however could be affected by this dynamic as well. The Englesby Brook watershed, among other stormwater impaired watersheds, will require numerous stormwater management installations to achieve the flow remediation targets and water quality goals as specified in the stormwater TMDLs. Given that many of the needed installations may be for smaller sites where peak flow rates are sensitive to modeling time step, lack of attention to temporal precision could result in unforeseen difficulties in meeting TMDL targets.

5.5 Concentration Data

The concentration data summary in Chapter 4 provided a combination of analysis and speculation about the causes of differences between the Butler Farms, Englesby, and national TN and TP datasets, with the high TP concentrations measured at the Englesby site being the most notable aspect of this analysis. FIGURE 33 showed that relatively high Englesby TP concentrations occurred along the inferable effective impervious front of the scatter plot, and while not displayed graphically, the points along that front were distributed throughout the year (March through October). Lawns are certain to contribute within the Englesby drainage area given that portions of lawn are directly connected to the piped drainage system and previous research has documented lawn runoff to be high in TP under a variety of circumstances with respect to fertilization, soil phosphorus levels, and clippings management (Bannerman et al. 1993; Bierman et al. 2010; Garn 2002; Steinke et al. 2007; Steinke et al. 2013; Waschbusch et al. 1999). However, if lawns were the dominant factor in the study area I would expect the highest TP

concentrations to have occurred for large storms and winter storms during which runoff from lawns would be greatest. Instead, high concentrations measured for small storms including during warmer months is more consistent with a combination of typical turf grass runoff and other TP sources within hydrologically connected areas which I speculate are dominated by pet wastes.

While pet wastes are suspected of playing a key role in elevated Englesby TP loads, neither the rate at which pet waste nutrient loads accumulate in the landscape nor their transport potential cannot be directly estimated. This is both because we did not assess pet wastes in our sampling, and previous research has not assessed this dynamic in sufficient detail to generalize to our site. For example, many municipal, local, and international government sources cite pet wastes as a source of nutrients and bacteria, but without explicit quantification. An extensive Web of Science and Google search did not produce useful quantitative estimates for pet waste nutrient loads. Despite the lack of quantification and accounting, the small residential lots, potentially high pet density, and lack of park areas within the Englesby sampling area circumstantially supports the hypothesis that pet wastes contribute to year round loads in connected road shoulder areas. Additional data on pet waste nutrient loads from the right way is needed to confirm or refute this hypothesis, which could, if further supported, inform management approaches to more precisely target pet waste nutrient loads.

The other potential sources of TP in Englesby runoff include deciduous leaf fall and lawn management exclusive of pet wastes (e.g., fertilizer, grass clippings) and sewage inputs to the stormwater system. Deciduous leaf fall is unlikely to be the

dominant factor given the timing of effective impervious TP loads throughout the year. However, the moderate deciduous canopy cover throughout the Englesby study area may contribute to higher pervious area runoff and phosphorus loss due to shading and thus lower quality turf. Similarly, the composite and grab water samples collected on days without wet weather flow did not have elevated TN or TP concentrations, which suggests wastewater inputs were not a dominant factor. Finally, while I discount turf grass loads alone as the dominant factor based on the relatively high concentrations measured during small summer storms, it cannot be ruled out that aesthetic concerns coupled with the difficulty of maintaining high quality turf in partially shaded conditions has led to heavy phosphorus fertilization of front lawns and road shoulder areas throughout the study area. These are the same areas that I have speculated would be 'hot spots' for pet wastes, and could plausibly produce elevated nutrient loads based on disproportionate management (e.g., fertilization). Ultimately the data collected in this study do not allow for more conclusive separation of contributing sources of TP.

As a final consideration, current Vermont Stormwater rules regulate water quality using a site-specific water quality volume calculation, which is strictly a function of regulated impervious surfaces and total site area and therefore ignores turf grass contributions. Under this framework, it is common for impervious surfaces to be graded to drain to pervious areas (i.e., 'disconnection'), which on moderately steep slopes or relatively poorly drained soils can result in mobilization of high loads of nutrients from lawns that receive runoff from relatively clean (from a nutrients perspective) roofs, driveways, and other surfaces. It is also common for stormwater catch basins to be set

back within pervious areas, with impervious surfaces draining to those areas. This can provide water quality benefits in some circumstances by diverting the runoff of many smaller storms to an area where volumes can infiltrate prior to reaching the engineering conveyance. However, during storms for which surface runoff flows to the catch basins via pervious areas (i.e., large or intense storms, or those influenced by frozen soils or meltwater), lawn loads and impervious run-on loads can then be jointly mobilized. Simple educational outreach stressing the importance of management of fertilizers and pet wastes in areas near catch basins or other discharge points could help to reduce these loads, if confirmed by further research. However, the inherently high TP concentrations presumably emanating from many existing turf grass surfaces irrespective of fertilizer and pet waste management will be difficult to manage if needed.

5.6 Summary

In summary, this work provides a mix of actionable findings, incremental progress and confirmation of present understanding that may be of use for stormwater modelers, researchers and managers. To maximize visibility of these insights to the relevant communities, publication of this dissertation will be followed by peer-reviewed publication of key findings, as well as submittal of the detention pond data to the International Stormwater BMP Database and summary findings to the SWMM users listserv.

REFERENCES

Aad, M. P. A., M. T. Suidan and W. D. Shuster (2010). Modeling techniques of best management practices: rain barrels and rain gardens using EPA SWMM-5. Journal of Hydrologic Engineering 15(6): 434-443.

Aguilar, C. and M. J. Polo (2011). Generating reference evapotranspiration surfaces from the Hargreaves equation at watershed scale. Hydrology and Earth System Sciences 15(8): 2495-2508.

Arnold, D. V. and H. G. Beyer (2003). A comparison of evolution strategies with other direct search methods in the presence of noise. Computational Optimization and Applications 24(1): 135-159.

Aronica, G., G. Freni and E. Oliveri (2005). Uncertainty analysis of the influence of rainfall time resolution in the modelling of urban drainage systems. Hydrological Processes 19(5): 1055-1071.

AVMA (2012). U.S. Pet Ownership and Demographic Sourcebook. Schaumburg, Il, American Veterinary Medical Association.

Bäck, T. and H. P. Schwefel (1993). An Overview of evolutionary algorithms for parameter optimization. Evolutionary Computation 1(1): 1-23.

Baird, C. (1999). Environmental Chemistry. 2nd Edition. New York, NY, W.H. Freeman and Company.

Balascio, C. C., D. J. Palmeri and H. Gao (1998). Use of a genetic algorithm and multi-objective programming for calibration of a hydrologic model. Transactions of the ASAE 41(3): 615-619.

Bannerman, R. T., D. W. Owens, R. B. Dodds and N. J. Hornewer (1993). Sources of pollutants in Wisconsin stormwater. Water Science and Technology 28(3-5): 241-259.

Barco, J., K. M. Wong and M. K. Stenstrom (2008). Automatic calibration of the US EPA SWMM model for a large urban catchment. Journal of Hydraulic Engineering-ASCE 134(4): 466-474.

Barrett, M. E. (2008). Comparison of BMP performance using the International BMP Database. Journal of Irrigation and Drainage Engineering-ASCE 134(5): 556-561.

Bayer, P. and M. Finkel (2004). Evolutionary algorithms for the optimization of advective control of contaminated aquifer zones. Water Resources Research 40(6): W06506.

Bayer, P. and M. Finkel (2007). Optimization of concentration control by evolution strategies: formulation, application, and assessment of remedial solutions. Water Resources Research 43(2): 402410.

Benaman, J., C. A. Shoemaker and D. A. Haith (2005). Calibration and validation of soil and water assessment tool on an agricultural watershed in upstate New York. Journal of Hydrologic Engineering 10(5): 363-374.

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources 16(1): 41-51.

Beven, K. (2006). On undermining the science? Hydrological Processes 20(14): 3141-3146.

Beven, K. and A. Binley (1992). The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6: 279-298.

Beven, K. and A. Binley (2014). GLUE: 20 years on. Hydrological Processes 28(24): 5897-5918.

Beven, K. J., P. J. Smith and J. E. Freer (2008). So just why would a modeller choose to be incoherent? Journal of Hydrology 354(1-4): 15-32.

Beyer, H. G. and D. V. Arnold (2003). Qualms regarding the optimality of cumulative path length control in CSA/CMA-evolution strategies. Evolutionary Computation 11(1): 19-28.

Bicknell, B. R., J. C. Imhoff, J. L. Kittle, T. H. Jobes and A. S. Donigan (2001). HSPF user's manual. Mountain View, California.

Bierman, P. M., B. P. Horgan, C. J. Rosen, A. B. Hollman and P. H. Pagliari (2010). Phosphorus runoff from turfgrass as affected by phosphorus fertilization and clipping management. Journal of Environmental Quality 39(1): 282-292.

Blasone, R. S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson and G. A. Zyvoloski (2008). Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. Advances in Water Resources 31(4): 630-648.

Booth, D. B. (1990). Stream-channel incision following drainage-basin urbanization. Water Resources Bulletin 26(3): 407-417.

Booth, D. B. and C. R. Jackson (1997). Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation. Journal of the American Water Resources Association 33(5): 1077-1090.

Borah, D. K. and M. Bera (2003). Watershed-scale hydrologic and nonpoint-source pollution models: review of mathematical bases. Transactions of the ASAE 46(6): 1553-1566.

Box, M. J. (1965). A new method of constrained optimization and comparison with other methods. Computer 8(1): 42-52.

Boyd, M. J., M. C. Bufill and R. M. Knee (1993). Pervious and impervious runoff in urban catchments. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 38(6): 463-478.

Brazier, R. E., K. J. Beven, J. Freer and J. S. Rowan (2000). Equifinality and uncertainty in physically based soil erosion models: Application of the glue methodology to WEPP-the water erosion prediction project-for sites in the UK and USA. Earth Surface Processes and Landforms 25(8): 825-845.

Burszta-Adamiak, E. and M. Mrowiec (2013). Modelling of green roofs' hydrologic performance using EPA's SWMM. Water Science and Technology 68(1): 36-42.

Cantone, J. P. and A. R. Schmidt (2009). Potential dangers of simplifying combined sewer hydrologic/hydraulic models. Journal of Hydrologic Engineering 14(6): 596-605.

City of Burlington. (2008). City of Burlington phase II stormwater annual report 2007.. Burlington Dept. of Public Works. Burlington, VT

City of Burlington. (2009). City of Burlington phase II stormwater annual report 2008. Burlington Dept. of Public Works. Burlington, VT

City of Burlington. (2010). City of Burlington phase II stormwater annual report 2009. Burlington Dept. of Public Works. Burlington, VT

City of South Burlington. (2010). City of South Burlington 2009 annual stormwater report. South Burlington Stormwater Services Division. South Burlington, VT

Comings, K. J., D. B. Booth and R. R. Horner (2000). Storm water pollutant removal by two wet ponds in Bellevue, Washington. Journal of Environmental Engineering-ASCE 126(4): 321-330.

Comprehensive Environmental and NH DES (2008). New Hampshire stormwater manual, vol. 2. NH Dept. of Env. Services. Concord, NH.

Conley, D. J., H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C. Lancelot and G. E. Likens (2009). ECOLOGY cntrolling eutrophication: nitrogen and phosphorus. Science 323(5917): 1014-1015.

Cook, S. E. K. (1976). Quest for an index of community structure sensitive to water pollution. Environmental Pollution 11(4): 269-288.

Center for Watershed Protection (2000). Englesby brook watershed restoration project draft final report**.** Center for Watershed Protection. Ellicott City, MD.

Deb, K., A. Pratap, S. Agarwal and T. Meyarivan (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6(2): 182-197.

Dietz, M. E. and J. C. Clausen (2008). Stormwater runoff and export changes with development in a traditional and low impact subdivision. Journal of Environmental Management 87(4): 560-566.

Dingman, S. L. (2002). Physical Hydrology. Upper Saddle River, N.J., Prentice Hall.

Doherty, J. and J. M. Johnston (2003). Methodologies for calibration and predictive analysis of a watershed model. Journal of the American Water Resources Association 39(2): 251-265.

Dow, C. L. and D. R. DeWalle (2000). Trends in evaporation and Bowen ratio on urbanizing watersheds in eastern United States. Water Resources Research 36(7): 1835-1843.

Downing, D. J., R. H. Gardner and F. O. Hoffman (1985). An examination of response-surface methodologies for uncertainty analysis in assessment models. Technometrics 27(2): 151-163.

Duan, Q. Y., S. Sorooshian and V. Gupta (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28(4): 1015-1031.

Eiben, A. E., and J.E. Smith (2003). Introduction to Evolutionary Computing. New York, NY, Springer.

Elliott, A. H. and S. A. Trowsdale (2007). A review of models for low impact urban stormwater drainage. Environmental Modelling & Software 22(3): 394-405.

Ensign, S. H. and M. W. Doyle (2006). Nutrient spiraling in streams and river networks. Journal of Geophysical Research-Biogeosciences 111: G04009.

Fang, T. J. and J. B. Ball (2007). Evaluation of spatially variable control parameters in a complex catchment modelling system: a genetic algorithm application. Journal of Hydroinformatics 9(3): 163-173.

Fennessey, L. A. J., A. C. Miller and J. M. Hamlett (2001). Accuracy and precision of NRCS models for small watersheds. Journal of the American Water Resources Association 37(4): 899-912.

Ferrey, S. (2004). Environmental Law. New York, NY, Aspen Publishers.

Fitzgerald, E. P., W. B. Bowden, S. P. Parker and M. L. Kline (2012). Urban impacts on streams are scale-dependent with nonlinear influences on their physical and biotic recovery in Vermont, United States. Journal of the American Water Resources Association 48(4): 679-697.

Foley, J. and W. B. Bowden (2005). University of Vermont stormwater project: statistical analysis of watershed variables. University of Vermont. Burlington, VT.

Fraley-McNeal, L. (2007). National pollutant removal performance database version 3. Center for Watershed Proteciton. Ellicott City, MD.

Freer, J., K. Beven and B. Ambroise (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. Water Resources Research 32(7): 2161-2173.

Galli, J. (1990). Thermal impacts associated with urbanization and stormwater management best management practices. Final report to: Sediment and Stormwater Administation of Maryland Department of the Environment, Metropolitian Washington Council of Governments.

Garn, H. S. (2002). Effects of lawn fertilizer on nutrient concentration in runoff from lakeshore lawns, Lauderdale Lakes, Wisconsin. Middleton, WI. USGS WRI 02-4130.

Gaume, E., J.-P. Villeneuve and M. Desbordes (1998). Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model. Journal of Hydrology 210(1-4): 38-50.

Geotech Consultants and Wright Water Engineers (2012). International stormwater BMP database pollutant catergory summary statistical addendum: TSS, bacteria, nutrients and metals. Feb. 2014 from http://www.bmpdatabase.org.

Gilbert, J. K. and J. C. Clausen (2006). Stormwater runoff quality and quantity from asphalt, paver, and crushed stone driveways in Connecticut. Water Research 40(4): 826-832.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA, Addison-Wesley Publishing Company.

Gray, J. R., G. D. Glysson, L. M. Turcios and G. E. Schwarz (2000). Comparability of suspended-sediment concentration and total suspended solids data. Reston, VA. USGS WRI 00-4191.

Groffman, P. M., A. M. Dorsey and P. M. Mayer (2005). N processing within geomorphic structures in urban streams. Journal of the North American Benthological Society 24(3): 613-625.

Guo, J. C. Y., J. C. Y. Cheng and L. Wright (2012). Field test on conversion of natural watershed into kinematic wave rectangular plane. Journal of Hydrologic Engineering 17(8): 944-951.

Guo, J. C. Y. and B. Urbonas (2009). Conversion of natural watershed to kinematic wave cascading plane. Journal of Hydrologic Engineering 14(8): 839-846.

Gupta, H. V., S. Sorooshian and P. O. Yapo (1998). Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resources Research 34(4): 751-763.

Gupta, V. K. and S. Sorooshian (1985). The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. Water Resources Research 21(4): 473-485.

Hammer, T. R. (1972). Stream channel enlargement due to urbanization. Water Resources Research 8(6): 1530-1540.

Hansen, N. and A. Ostermeier (1996). Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. Proceedings of the 1996 IEEE International Conference on Evolutionary Computation, Piscataway, New Jersey, IEEE Press.

Hargreaves, G. H. and Z. A. Samani (1982). Estimating potential evapo-transpiration. Journal of the Irrigation and Drainage Division-ASCE 108(3): 225-230.

Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney and J. G. Arnold (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASABE 49(3): 689-701.

Hendrickson, J. D., S. Sorooshian and L. E. Brazil (1988). Comparison of newton-type and direct search algorithms for calibration of conceptual rainfall-runoff models. Water Resources Research 24(5): 691-700.

Herb, W. R., B. Janke, O. Mohseni and H. G. Stefan (2008). Thermal pollution of streams by runoff from paved surfaces. Hydrological Processes 22(7): 987-999.

Hjelmfelt, A. T. (1991). Investigation of curve number procedure. Journal of Hydraulic Engineering-ASCE 117(6): 725-737.

Hobbs, W., B. Lubliner, N. Kale and E. Newell (2015). Western Washington NPDES phase 1 stormwater permit: final data characterization 2009-2013. Olympia, WA.

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems. Ann Arbor, Michigan, University of Michigan Press.

Hossain, M. A., M. Alam, D. R. Yonge and P. Dutta (2005). Efficiency and flow regime of a highway stormwater detention pond in Washington, USA. Water Air and Soil Pollution 164(1-4): 79-89.

Houle, J. J., R. M. Roseen, T. P. Ballestero, T. A. Puls and J. Sherrard (2013). Comparison of maintenance cost, labor demands, and system performance for LID and conventional stormwater management. Journal of Environmental Engineering 139(7): 932-938.

Howarth, R. W. (1988). Nutrient limitation of net primary production in marine ecosystems. Annual Review of Ecology and Systematics 19: 89-110.

Huff, F. A. and S. A. Changnon, Jr. (1973). Precipitation modification by major urban areas. Bulletin of the American Meteorological Society 54(12): 1220-1232.

Ibbitt, R. P. and T. O'Donnell (1971). Fitting methods for conceptual catchment models. Journal of the Hydraulics Division-ASCE 97(9): 1331-1342.

International Stormwater BMP Database. (2012). Retrieved Jan. 7, 2014, from www.bmpdatabase.org.

Jang, S., M. Cho, J. Yoon, Y. Yoon, S. Kim, G. Kim, L. Kim and H. Aksoy (2007). Using SWMM as a tool for hydrologic impact assessment. Desalination 212(1-3): 344-356.

Janke, B. D., O. Mohseni, W. R. Herb and H. G. Stefan (2011). Heat release from rooftops during rainstorms in the Minneapolis/St. Paul Metropolitan Area, USA. Hydrological Processes 25(13): 2018-2031.

Jeon, J. H., C. G. Park and B. A. Engel (2014). Comparison of performance between genetic algorithm and SCE-UA for calibration of SCS-CN surface runoff simulation. Water 6(11): 3433-3456.

Kertesz, R. and J. Sansalone (2014). Hydrologic transport of thermal energy from pavement. Journal of Environmental Engineering 140(8): 04014028.

Kim, B., B. F. Sanders, K. Han, Y. Kim and J. S. Famiglietti (2014). Calibration of stormwater management model using flood extent data. Proceedings of the Institution of Civil Engineers-Water Management 167(1): 17-29.

Knighton, J., E. White, E. Lennon and R. Rajan (2014). Development of probability distributions for urban hydrologic model parameters and a Monte Carlo analysis of model sensitivity. Hydrological Processes 28(19): 5131-5139.

Kofstad, A. (2011). Stormwater management and the American neighborhood: a survey of New England residents. Rubenstein School of the Environment and Natural Resouces. Burlington, VT, University of Vermont. MS.

Kollat, J. B. and P. M. Reed (2006). Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. Advances in Water Resources 29(6): 792-807.

Krebs, G., T. Kokkonen, M. Valtanen, H. Koivusalo and H. Setala (2013). A high resolution application of a stormwater management model (SWMM) using genetic parameter optimization. Urban Water Journal 10(6): 394-410.

Kumar, S. and S. C. Jain (1982). Application of SCS infiltration-model. Water Resources Bulletin 18(3): 503-507.

Lake Champlain Basin Program (2014). Lawn to lake: lawn care tips for green lawns not green lakes! . Retrieved Jun. 16, 2015, from http://www.lawntolake.org/index.htm.

Lee, S. C., I. H. Park, J. I. Lee, H. M. Kim and S. R. Ha (2010). Application of SWMM for evaluating NPS reduction performance of BMPs. Desalination and Water Treatment 19(1-3): 173-183.

Lenat, D. R. and J. K. Crawford (1994). Effects of land-use on water-quality and aquatic biota of 3 North Carolina piedmont streams. Hydrobiologia 294(3): 185-199.

Leopold, L. B. (1968). Hydrology for urban land planning- a guidebook on the hydrologic effects of urban land use. Geologic Survey Circular 554.

Liong, S. Y., W. T. Chan and J. Shreeram (1995). Peak-flow forecasting with genetic algorithm and SWMM. Journal of Hydraulic Engineering-ASCE 121(8): 613-617.

Litke, D. (1999). Review of phosphorus control measures in the United States and their effects on water qualtiy. Denver, CO. USGS WRIR 99-4007.

Long, D. L. and R. L. Dymond (2014). Thermal pollution mitigation in cold water stream watersheds using bioretention. Journal of the American Water Resources Association 50(4): 977-987.

Long Island Regional Planning Board 1982. The Long Island segment of the Nationwide Urban Runoff Program. Hauppauge, NY.

Lucas, W. C. (2010). Design of integrated bioinfiltration-detention urban retrofits with design storm and continuous simulation methods. Journal of Hydrologic Engineering 15(6): 486-498.

Lyons, J., S. W. Trimble and L. K. Paine (2000). Grass versus trees: managing riparian areas to benefit streams of central North America. Journal of the American Water Resources Association 36(4): 919-930.

MacRae, C. R. (1993). An alternate design approach for the control of instream erosion potential in urbanizing watersheds. Proceedings of the Sixth International Conference on Urban Storm Drainage. Niagra Falls, Ontario, Canada.

Maier, U., C. DeBiase, O. Baeder-Bederski and P. Bayer (2009). Calibration of hydraulic parameters for large-scale vertical flow constructed wetlands. Journal of Hydrology 369(3-4): 260-273.

Mallin, M. A., S. H. Ensign, T. L. Wheeler and D. B. Mayes (2002). Pollutant removal efficacy of three wet detention ponds. Journal of Environmental Quality 31(2): 654-660.

Mantovan, P. and E. Todini (2006). Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology. Journal of Hydrology 330(1-2): 368-381.

McCuen, R. H. (1979). Downstream effects of stormwater management basins. Journal of the Hydraulics Division-ASCE 105(11): 1343-1356.

McCuen, R. H. and G. E. Moglen (1988). Multicriterion stormwater management methods. Journal of Water Resources Planning & Management 114(4): 414-431.

McIntosh, A., A. Hackman, B. Kirk, B. Bowden, E. Fitzgerald and J. Todd (2006). RAN: working with neighborhoods to manage stormwater. Stormwater May/June: 95-99.

McKay, M. D., R. J. Beckman and W. J. Conover (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2): 239-245.

Md. Code, Env. Art. §4-201.1 and §4-203. Maryland's stormwater managment act of 2007. 2007.

ME DEP (2006). BMP technical design manual. Augusta, ME. Maine Dept.of the Environment.

Medalie, L. (2012). Effects of urban best management practices on streamflow and phosphorus and suspended-sediment transport on Englesby Brook in Burlington, Vermont, 2000-2010. Montpelier, VT. USGS SIR 2012-5103.

Meybeck, M. (1982). Carbon, nitrogen, and phosphorus transport by world rivers. American Journal of Science 282(4): 401-450.

Monteith, J. L. (1965). Evaporation and environment. Symposia of the Society for Experimental Biology 19: 205-224.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel and T. L. Veith (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50(3): 885-900.

Mulholland, P. J., R. O. Hall, D. J. Sobota, W. K. Dodds, S. E. G. Findlay, N. B. Grimm, S. K. Hamilton, W. H. McDowell, J. M. O'Brien, J. L. Tank, L. R. Ashkenas, L. W. Cooper, C. N. Dahm, S. V. Gregory, S. L. Johnson, J. L. Meyer, B. J. Peterson, G. C. Poole, H. M. Valett, J. R. Webster, C. P. Arango, J. J. Beaulieu, M. J. Bernot, A. J. Burgin, C. L. Crenshaw, A. M. Helton, L. T. Johnson, B. R. Niederlehner, J. D. Potter, R. W. Sheibley and S. M. Thomas (2009). Nitrate removal in stream ecosystems measured by N-15 addition experiments: denitrification. Limnology and Oceanography 54(3): 666-680.

Naiman, R. J. and H. Decamps (1997). The ecology of interfaces: riparian zones. Annual Review of Ecology and Systematics 28: 621-658.

Naranjo, R. C., R. G. Niswonger, M. Stone, C. Davis and A. McKay (2012). The use of multiobjective calibration and regional sensitivity analysis in simulating hyporheic exchange. Water Resources Research 48: W01538.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. Computer Journal 7(4): 308-313.

Nielson, L. and C. L. Smith (2005). Influences on residential yard care and water quality: Tualatin watershed, Oregon. Journal of the American Water Resources Association 41(1): 93-106.

Novotny, V. (2003). Water Quality: Diffuse Pollution and Watershed Managament. Hoboken, N.J., Jon Wiley and Sons, Inc.

National Research Council (2009). Urban stormwater management in the United States. Committee on Reducing Stormwater Discharge Contributions to Water Pollution. National Research Council. Washington, D.C.

Obropta, C. C. and J. S. Kardos (2007). Review of urban stormwater quality models: deterministic, stochastic, and hybrid approaches. Journal of the American Water Resources Association 43(6): 1508-1523.

Ostermeier, A., A. Gawelczyk and N. Hansen (1994). A derandomized approach to self-adaptation of evolution strategies. Evolutionary Computation 2(4): 369-380.

Paul, M. J. and J. L. Meyer (2001). Streams in the urban landscape. Annual Review of Ecology and Systematics 32: 333-365.

Peterson, B. J., W. M. Wollheim, P. J. Mulholland, J. R. Webster, J. L. Meyer, J. L. Tank, E. Marti, W. B. Bowden, H. M. Valett, A. E. Hershey, W. H. McDowell, W. K. Dodds, S. K. Hamilton, S. Gregory and D. D. Morrall (2001). Control of nitrogen export from watersheds by headwater streams. Science 292(5514): 86-90.

Peterson, E. W. and C. M. Wicks (2006). Assessing the importance of conduit geometry and physical parameters in karst systems using the storm water management model (SWMM). Journal of Hydrology 329(1-2): 294-305.

Peterson, J. R. and J. M. Hamlett (1998). Hydrologic calibration of the SWAT model in a watershed containing fragipan soils. Journal of the American Water Resources Association 34(3): 531-544.

Pickup, G. (1977). Testing the efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models. Hydrological Sciences Bulletin 22(2): 257-274.

Pitt, R. (2011). The National Stormwater Quality Database, Version 3.1. Tuscaloosa, AL, Department of Civil and Environmental Engineering, University of Alabama.

Pitt, R., A. Maestre and R. Morquecho (2004). The National Stormwater Quality Database, Version 1.1. Tuscaloosa, AL, Department of Civil and Environmental Engineering, University of Alabama.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks and J. C. Stromberg (1997). The natural flow regime. Bioscience 47(11): 769-784.

Ponce, V. M. and R. H. Hawkins (1996). Runoff curve number: has it reached maturity? Journal of Hydrologic Engineering 1(1): 11-19.

Pratt, J. M. and R. A. Coler (1976). A procedure for the routine biological evaluation of urban runoff in small rivers. Water Research 10(11): 1019-1025.

Reed, P. M., D. Hadka, J. D. Herman, J. R. Kasprzyk and J. B. Kollat (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. Advances in Water Resources 51: 438-456.

Rossman, L. A. (2010). Storm water management model user's manual. U.S. EPA Office of Research and Development. Cincinatti, OH.

Rossman, L. A. (2014). Re: [SWMM-USERS] Applications of SWMM in rural/undeveloped areas. SWMM User's Listserve. Posted on: Jan. 7, 2014.

Roux, H. and D. Dartus (2008). Sensitivity analysis and predictive uncertainty using inundation observations for parameter estimation in open-channel inverse problem. Journal of Hydraulic Engineering-ASCE 134(5): 541-549.

Sabouri, F., B. Gharabaghi, A. A. Mahboubi and E. A. McBean (2013). Impervious surfaces and sewer pipe effects on stormwater runoff temperature. Journal of Hydrology 502: 10-17.

Saltelli, A. (2002). Sensitivity analysis for importance assessment. Risk Analysis 22(3): 579-590.

Saltelli, A., S. Tarantola and K. P. S. Chan (1999). A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41(1): 39-56.

Schindler, D. W. (1977). Evolution of phosphorus limitation in lakes. Science 195(4275): 260-262.

Schueler, T. (2000). Irreducible pollutant concentrations discharged from stormwater practices. Watershed Protection Techniques. Center for Watershed Protection. Ellicott City, MD.

Schueler, T. R., L. Fraley-McNeal and K. Cappiella (2009). Is impervious cover still important? review of recent research. Journal of Hydrologic Engineering 14(4): 309-315.

Shaw, S. B., J. R. Stedinger and M. T. Walter (2010). Evaluating urban pollutant buildup/wash-off models using a Madison, Wisconsin catchment. Journal of Environmental Engineering-ASCE 136(2): 194-203.

Shenk, G. W., J. Wu and L. C. Linker (2012). Enhanced HSPF model structure for Chesapeake Bay watershed simulation. Journal of Environmental Engineering-ASCE 138(9): 949-957.

Sieber, A. and S. Uhlenbrook (2005). Sensitivity analyses of a distributed catchment model to verify the model structure. Journal of Hydrology 310(1-4): 216-235.

Simmons, D. L. and R. J. Reynolds (1982). Effects of urbanization on base-flow of selected south-shore streams, Long Island, New York. Water Resources Bulletin 18(5): 797-805.

Singer, M. B., R. Aalto, L. A. James, N. E. Kilham, J. L. Higson and S. Ghoshal (2013). Enduring legacy of a toxic fan via episodic redistribution of California gold mining debris. Proceedings of the National Academy of Sciences of the United States of America 110(46): 18436-18441.

Smullen, J., A. Shallcross and K. Cave (1999). Updating the U.S. Nationwide Urban Runoff Quality Database. Water, Science and Technology 39(12): 9-16.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55(1-3): 271-280.

Song, X. M., J. Y. Zhang, C. S. Zhan, Y. Q. Xuan, M. Ye and C. G. Xu (2015). Global sensitivity analysis in hydrological modeling: review of concepts, methods, theoretical framework, and applications. Journal of Hydrology 523: 739-757.

Spear, R. C. and G. M. Hornberger (1980). Eutrophication in peel inlet--II. identification of critical uncertainties via generalized sensitivity analysis. Water Research 14(1): 43-49.

Spinello, A. G. and D. L. Simmons (1992). Base flow of 10 south-shore streams, Long Island, New York, 1976-85, and the effects of urbanization on base flow and flow duration. Syosset, NY. USGS WRI 90-4205.

Spronken-Smith, R. A. and T. R. Oke (1998). The thermal regime of urban parks in two cities with different summer climates. International Journal of Remote Sensing 19(11): 2085-2104.

Stedinger, J. R., R. M. Vogel, S. U. Lee and R. Batchelder (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44: W00B06.

Steinke, K., W. R. Kussow and J. C. Stier (2013). Potential contributions of mature prairie and turfgrass to phosphorus in urban runoff. Journal of Environmental Quality 42(4): 1176-1184.

Steinke, K., J. C. Stier, W. R. Kussow and A. Thompson (2007). Prairie and turf buffer strips for controlling runoff from paved surfaces. Journal of Environmental Quality 36(2): 426-439.

Sterner, R. W. (2008). On the phosphorus limitation paradigm for lakes. International Review of Hydrobiology 93(4-5): 433-445.

Sterner, R. W. and J. J. Elser (2002). Ecological Stoichiometry: the Biology of Elements from Molecules to the Biosphere. Princeton, NJ, Princeton University Press.

Steuer, J., W. Selbig, N. Hornewer, and J. Prey (1997). Sources of contamination in an urban basin in Marquette, Michigan and an analysis of concentrations, loads, and data quality. Middleton, WI. USGS WRI 97-4242.

Strecker, E. W., M. M. Quigley, B. R. Urbonas, J. E. Jones and J. K. Clary (2001). Determining urban storm water BMP effectiveness. Journal of Water Resources Planning & Management 127(3): 144.

Sun, N., B. G. Hong and M. Hall (2014). Assessment of the SWMM model uncertainties within the generalized likelihood uncertainty estimation (GLUE) framework for a high-resolution urban sewershed. Hydrological Processes 28(6): 3018-3034.

Tang, Y., P. Reed, T. Wagener and K. van Werkhoven (2007). Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrology and Earth System Sciences 11(2): 793-817.

Temprano, J., O. Arango, J. Cagiao, J. Suarez and I. Tejero (2006). Stormwater quality calibration by SWMM: a case study in northern Spain. Water SA 32(1): 55-63.

Tetra Tech Inc. (2005). Stormwater modeling for flow duration curve development in Vermont: final report. Fairfax, VA. Prepared for: U.S. EPA Region 1 and VT DEC.

Thompson, A. M., K. Kim and A. J. Vandermuss (2008). Thermal characteristics of stormwater runoff from asphalt and sod surfaces. Journal of the American Water Resources Association 44(5): 1325-1336.

Thompson, A. M., A. J. Vandermuss, J. M. Norman and A. Roa-Espinosa (2008). Modeling the effect of a rock crib on reducing stormwater runoff temperature. Transactions of the ASABE 51(3): 947-960.

Timmons, D. R. and R. F. Holt (1977). Nutrient lossess in surface runoff from a native pararie. Journal of Environmental Quality 6(4): 369-373.

Trimble, S. W. (1997). Contribution of stream channel erosion to sediment yield from an urbanizing watershed. Science 278(5342): 1442-1444.

Tsihrintzis, V. A. and R. Hamid (1998). Runoff quality prediction from small urban catchments using SWMM. Hydrological Processes 12(2): 311-329.

UNH Stormwater Center (2009). University of New Hampshire stormwater center 2009 biannual report. University of New Hampshire Stormwater Center. Durham, NH.

UNH Stormwater Center (2014). Urban watershed renewal in Berry Brook. Retrieved Dec. 14, 2014, from http://www.unh.edu/unhsc/berrybrook.

U.S. EPA. (1983). Results of the nationwide urban runoff program- fnal report. Water Planning Division. Washington, DC.

U.S. EPA (1993). Guidance specifying management measures for sources of nonpoint pollution in coastal waters. Chapter 4: management measures for urban areas. U.S. EPA Office of Water. Washington, D.C.

U.S. EPA (2010). Chesapeake Bay total maximum daily load for nitrogen, phosphorus and sediment. U. S. EPA, Regions 2 and 3. Philladelphia, PA, Annapolis, MD, and New York, NY.

U.S. EPA (2013). EPA water quality assessment and TMDL information: national summary of state information. Retrieved Dec. 3, 2013, from http://ofmpub.epa.gov/waters10/attains_nation_cy.control.

U.S. EPA (2015). Phosphorus TMDLs for Vermont segments of Lake Champlain. U.S. EPA Region 1. Boston, MA.

U.S. SCS (1965). National engineering handbook section 4 hydrology. U.S. Soils Conservation Service. Washington, DC.

Van Buren, M. A., W. E. Watt, J. Marsalek and B. C. Anderson (2000). Thermal enhancement of stormwater runoff by paved surfaces. Water Research 34(4): 1359-1371.

van Werkhoven, K., T. Wagener, P. Reed and Y. Tang (2008). Characterization of watershed model behavior across a hydroclimatic gradient. Water Resources Research 44(1): W01429.

van Werkhoven, K., T. Wagener, P. Reed and Y. Tang (2009). Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources 32(8): 1154-1169.

Vrugt, J. A. and B. A. Robinson (2007). Improved evolutionary optimization from genetically adaptive multimethod search. Proceedings of the National Academy of Sciences of the United States of America 104(3): 708-711.

VT ANR (2002). The Vermont sormwater management manual. vol. 1: stormwater treatment standards. VT Agency of Natural Resouces. Waterbury, VT.

VT ANR (2006). Total maximum daily load to address biological impairment in Potash Brook, Chittenden County, Vermont. VT Agency of Natural Resouces. Waterbury, VT.

VT ANR (2007). Total maximum daily load to address biological impairment in Englesby Brook, Chittenden County, Vermont. VT Agency of Natural Resouces. Waterbury, VT.

VT ANR (2007b). Total maximum daily load to address biological impairment in Morehouse Brook, Chittenden County, Vermont. VT Agency of Natural Resouces. Waterbury, VT.

VT ANR (2014). Vermont water quality standards. VT Agency of Natural Resources. Montpelier, VT.

Vt. Env. Pro. Rules, § 18-302. Stormwater management rule. 2011.

Vt. Stat. Ann. tit. 10, § 1266b. Application of phosphorus fertilizer. 2012

Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta and S. Sorooshian (2001). A framework for development and application of hydrological models. Hydrology and Earth System Sciences 5(1): 13-26.

Wagener, T. and J. Kollat (2007). Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. Environmental Modelling & Software 22(7): 1021-1033.

Walker, W. W. (1990). P8 urban catchement model program documentation.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman and R. P. Morgan (2005). The urban stream syndrome: current knowledge and the search for a cure. Journal of the North American Benthological Society 24(3): 706-723.

Wang, Q. J. (1991). The genetic algorithm and its application to calibrating conceptutal rainfall-runoff models. Water Resources Research 27(9): 2467-2471.

Waschbusch, R.J., W.R. Selbig, and R.T. Bannerman (1999). Sources of phosphorus in stormwater and street dirt from two urban residential basins in Madison, Wisconsin, 1194-95. USGS WRIR 99-4021.

Weiss, P. T., J. S. Gulliver and A. J. Erickson (2007). Cost and pollutant removal of storm-water treatment practices. Journal of Water Resources Planning and Management-ASCE 133(3): 218-229.

WERF (2010). Bellevue, Washington- merging stormwater features with parks and recreation. Retrieved Jun. 10, 2015, from http://www.werf.org/liveablecommunities/studies_bell_wa.htm.

Whittemore, R. C. (2004). Discussion "Methodologies for calibration and predictive analysis of a watershed model, by John Doherty and John M. Johnston". Journal of the American Water Resources Association 40(1): 267-267.

Willeke, G.E. (1997). Discussion "Runoff curve number: has it reached maturity?, by V.M. Ponce and R. H. Hawkins." Journal of Hydrologic Engineering 2(3): 145-148.

Wilson, C. E., W. F. Hunt, R. J. Winston and P. Smith (2015). Comparison of runoff quality and quantity from a commercial low-impact and conventional development in Raleigh, North Carolina. Journal of Environmental Engineering 141(2).

Winer, R. (2000). National pollutant removal database for stormwater treatment practices. Center for Watershed Protection. Ellicott City, MD.

Wissmar, R. C., R. K. Timm and M. G. Logsdon (2004). Effects of changing forest and impervious land covers on discharge characteristics of watersheds. Environmental Management 34(1): 91-98.

Wolman, M. G. (1967). A cycle of sedimentation and erosion in urban river channels. Geografiska Annaler. Series A, Physical Geography 49(2/4): 385-395.

Wong, T. S. W. and M. C. Zhou (2003). Kinematic wave parameters and time of travel in circular channel revisited. Advances in Water Resources 26(4): 417-425.

Wright Water Engineers and Geosyntec Consultants (2012). International stormwater BMP database: narrative overview of BMP database study characteristics. Retrieved Feb. 2014 from http://www.bmpdatabase.org.

Wu, J. S., R. E. Holman and J. R. Dorney (1996). Systematic evaluation of pollutant removal by urban wet detention ponds. Journal of Environmental Engineering-ASCE 122(11): 983-988.

Yapo, P. O., H. V. Gupta and S. Sorooshian (1998). Multi-objective global optimization for hydrologic models. Journal of Hydrology 204(1-4): 83-97.

Yatheendradas, S., T. Wagener, H. Gupta, C. Unkrich, D. Goodrich, M. Schaffner and A. Stewart (2008). Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resources Research 44(5): W05S19.

Yoon, J. H. and C. A. Shoemaker (1999). Comparison of optimization methods for ground-water bioremediation. Journal of Water Resources Planning and Management-ASCE 125(1): 54-63.

Zhang, C., J. G. Chu and G. T. Fu (2013). Sobol"s sensitivity analysis for a distributed hydrological model of Yichun River Basin, China. Journal of Hydrology 480: 58-68.

Zhang, G., J. M. Hamlett, P. Reed and Y. Tang (2013). Multi-objective optimization of low impact development designs in an urbanizing watershed. Open Journal of Optimization 2(4): 95-108.

Zhang, W. and T. Li (2015). The influence of objective function and acceptability threshold on uncertainty assessment of an urban drainage hydraulic model with generalized likelihood uncertainty estimation methodology. Water Resources Management 29(6): 2059-2072.

Zoppou, C. (2001). Review of urban storm water models. Environmental Modelling & Software 16(3): 195-231.

APPENDIX A

TABLE 23. Englesby detention pond inlet sampling details. Sampled flow computed as volume of flow between time of first sample and time of last sample plus the average volume between sample aliquots.

First Sample	Last Sample	Aliquots	TN (mg L^{-1})	TP (mg L^{-1})	Sampled Flow (m^3)	Peak Flow (m^3 s^{-1})
2007-09-09 09:52	2007-09-09 19:47	2	0.580	0.303	388	0.045
2007-09-11 18:25	2007-09-11 19:09	2	1.455	0.232	360	0.150
2007-10-19 20:30	2007-10-20 02:51	57	0.880	0.101	2,410	0.582
2007-10-23 11:14	2007-10-25 07:22	36	1.010	0.187	1,562	0.262
2007-10-26 21:19	2007-10-28 16:24	51	0.530	0.105	3,099	0.225
2007-11-15 00:09	2007-11-16 08:20	75	0.730	0.135	3,277	0.094
2008-04-01 11:43	2008-04-02 10:17	54	2.010	0.480	2,150	0.151
2008-04-11 11:43	2008-04-12 22:55	75	1.720	0.625	3,132	0.355
2008-04-28 12:19	2008-04-30 03:38	26	1.470	1.120	2,220	0.130
2008-06-03 13:26	2008-06-04 11:05	21	1.970	0.945	527	0.066
2008-06-10 13:28	2008-06-11 07:14	40	2.830	0.875	1,037	0.376
2008-06-14 19:15	2008-06-15 11:11	14	1.360	0.414	712	0.220
2008-06-28 15:09	2008-06-30 01:05	9	1.730	1.010	1,599	0.558
2008-07-09 12:49	2008-07-09 14:33	5	1.320	0.644	1,044	0.548
2008-07-13 14:37	2008-07-13 22:09	8	1.270	1.097	1,647	0.144
2008-07-23 04:17	2008-07-24 10:47	39	0.960	0.212	2,008	0.468
2008-07-30 15:11	2008-07-31 06:36	9	1.960	0.630	444	0.030
2008-08-08 17:49	2008-08-11 05:13	9	1.450	0.494	1,021	0.078
2008-08-18 20:37	2008-08-19 08:25	18	0.820	0.498	1,870	0.288
2008-09-09 06:34	2008-09-09 10:07	4	2.225	1.190	478	0.203
2008-09-26 17:48	2008-09-26 21:48	7	1.770	1.185	476	0.074
2008-10-21 07:47	2008-10-22 01:37	7	2.320	1.350	536	0.022
2008-10-25 15:44	2008-10-27 12:38	63	1.050	1.670	4,122	0.619
2008-11-08 18:26	2008-11-10 23:34	22	1.100	0.281	1,393	0.119
2009-02-11 06:15	2009-02-13 04:05	25	3.630	1.330	4,002	0.134
2009-03-26 18:31	2009-03-26 23:14	7	5.200	3.690	434	0.046
2009-04-03 17:26	2009-04-06 06:50	30	1.790	1.110	2,112	0.079
2009-04-06 12:04	2009-04-08 07:14	27	1.650	0.615	1,810	0.092
2009-05-07 02:08	2009-05-07 23:54	8	2.470	3.600	588	0.028
2009-05-24 05:07	2009-05-24 17:08	3	0.600	0.194	245	0.030
2009-05-27 00:48	2009-05-29 09:15	30	0.800	0.362	2,012	0.044
2009-06-07 17:55	2009-06-10 07:55	15	1.300	1.040	1,111	0.136
2009-06-11 16:11	2009-06-13 10:19	45	0.850	0.387	2,972	0.237
2009-06-18 15:07	2009-06-19 23:25	8	1.450	0.468	598	0.031
2009-06-26 16:34	2009-06-27 05:35	21	2.040	1.150	1,065	0.661
2009-06-29 05:27	2009-06-30 13:19	15	1.130	0.414	824	0.057
2009-06-30 15:18	2009-07-01 05:16	14	1.290	0.590	767	0.267
2009-07-06 14:12	2009-07-07 02:03	3	1.710	0.304	169	0.057
2009-07-07 15:07	2009-07-08 12:05	54	1.455	0.738	2,851	1.264
2009-07-11 14:55	2009-07-13 04:51	33	1.330	0.310	1,950	0.240
2009-11-14 12:05	2009-11-16 08:23	40	1.850	0.488	---	---
2009-11-26 23:40	2009-11-30 08:54	41	0.840	0.138	---	---
2009-12-03 00:31	2009-12-04 05:07	26	0.940	0.204	---	---