A METHODOLOGY FOR SELECTING AND EVALUATING ALTERNATIVE LONG-TERM MONITORING NETWORKS FOR WATER QUALITY IN LARGE WATERSHEDS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Sue Nee Tan

May 2013

ABSTRACT

This paper presents a practical methodology for selecting and evaluating alternative long-term monitoring networks in large watersheds. The methodology takes into account weather variability, optimizes the sensor locations, and provides information on the tradeoff among different monitoring networks. Assessment of a sensor network is based on the accuracy of hydrologic model predictions with inputs from sampling points in the sensor network. Accuracy is assessed by comparison to the "true output," a spatially distibuted time series of the flow and contaminant constituents throughout the watershed. The true output is generated by simulating a hydrologic model with a a set of parameters selected based on the best available information, with complete weather information generated using a spatially distributed weather generator. The methodology provides a platform for testing several sensor network configurations in a matter of hours rather than years, since sensors do not have to be physically placed in the watersheds. Three case studies illustrate the integrated placement of in-stream and rain gauges in the watershed as long-term monitoring networks for hydrologic model calibration. Results show that raingauge networks are more important for flow simulations rather than nutrient simulations. Additional in-stream gauges may not be necessary for flow estimation, but nutrient estimation benefits from information from additional gauges. In addition, statistically optimal placement of rain gauges is found to work better than placing the rain gauges according to geographical criteria. Finally, the decomposition of the Nash Sutcliffe Efficiency measure into its components on the case study shows that the model performance is largely driven by SWAT's ability to

match the timing and shape of the hydrograph, and for high model performaces that there exists a tradeoff between matching the distribution of the flows and nutrient outputs, and correctly approximating the timing and shape of the hydrograph.

BIOGRAPHICAL SKETCH

Sue Nee Tan was born in Kuala Lumpur, Malaysia. She received double Bachelor of Science degrees in Civil and Environmental Engineering (B.Sc. CEE) and Earth and Environmental Science (B.Sc. EES) from Lehigh University in Bethlehem, PA in 2009. Sue Nee is currently continuing her Ph.D. research on wind integration in hydropower systems with her advisor Dr. Christine A. Shoemaker at Cornell University.

To my husband, Kevin for his patience, understanding, late-night counseling and support;

To my parents, for everything

ACKNOWLEDGMENTS

Thank you to my advisor Prof. Christine A. Shoemaker for her continued support, time and guidance.

Thanks to Prof. Daniel S. Wilks from the Cornell Earth and Atmospheric Sciences Department for the use of his weather generator code, an instrumental part of this research.

Also, thank you to my seniors Drs. Taimoor Akhtar and Joshua Woodbury for sharing their experiences and their patience in helping a fresh graduate student find navigate work and life at Cornell.

This research is a culmination of the work of the abovementioned people, as well as previous graduate students of Prof. Shoemaker and was made possible through financial support from NSF grants EAR 0711491 and CBET 075675, and the Cornell Civil and Environmental Engineering Department.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SWAT	Soil and Water Tool
POPT	Precipitation gauge location optimization
NSE	Nash-Sutcliffe Efficiency measure of model performance
BVR-rectangle	Bias variance correlation rectangle plot

LIST OF SYMBOLS

TP	"True" precipitation scenarios
OP(R)	observed precipitation data from the rain gauge network R
SI(R)	simulated output at each potential in-stream gauge site based on observations from rain gauge network R
TI	"True" observations at all the possible in-stream gauge sites
OI	observed in-stream gauge outputs at each potential in-stream gauge site
OI(F)	observed data from the in-stream gauge network F
SI(Rbest,F)	simulated model output using input from the best rain gauge network and the candidate in-stream gauge F
θ(true)	the known hydrologic model parameters
$\theta(R_{best},F)$	a set of calibrated parameters as a function of the best rain gauge configuration R_{best}, and the candidate in-stream gauge configuration, F
$\overline{NSE}$	average Nash-Sutcliffe Effiency measure for the model outputs over all subbasins
$\overline{\alpha}$	average relative variability over all subbasins, or the average ratio of the standard deviation of model outputs to the observed standard deviation.
$\overline{\beta_n}$	average normalized bias over all subbasins, or the average difference of the model output mean and the observed mean normalized by the observed standard deviation.
$\overline{r_k}$	is the average correlation coefficient over all subbasins between model output and observed data

CHAPTER 1
INTRODUCTION

Long-term monitoring networks that are cost-efficient and informational can help hydrologists model and assist watershed managers in making better decisions involving environmental and economic activities within the watershed [*Reed et al.*, 2006; *Mishra and Coulibaly*, 2009]. However, given the specific goals of a monitoring network, there is the question of determining the number, location and types of sensors to be installed. This is a complicated problem due to the spatial and temporal variability of weather inherent in large watersheds, the large number of potential sensor configurations, and the interaction of different hydrological processes such as rainfal-runoff generation and nutrient generation.

Data from long-term monitoring networks are generally used for hydrologic model calibration. The data has to be collected continuously over a sufficient period of time so that many different conditions (high flow, low flow) can be captured in the watershed. Since this time period should be at least a year in length for daily data, evaluating several different candidate sensor network configurations cannot be done by physically placing sensors in a watershed and collecting data over a period of time, as shown by Vandenberghe et al [2007]. Evaluation of monitoring networks by physically placing sensors in the watershed of interest may take either a long time to evaluate (on the order of years), or poor estimates of model performance due to insufficient data capture.

This paper proposes a methodology for selecting and evaluating alternative long-term monitoring networks which take into account weather variability, optimizes the sensor locations, and provides information on the tradeoff among different monitoring configurations. The methodology has the advantage of: 1) Allowing the watershed manager to test various configurations of sensor networks in a relatively fast timeframe without the cost of having to place actual sensors within a watershed; 2) allowing the watershed manager to see how the watershed model that they are using will react given more information through additional sensors; and 3) allowing the watershed manager to see different aspects of model performance.

The methodology allows for the configuration of several different types of sensors. In this paper, the configuration of a long-term monitoring network consisting of in-stream and rain gauges is demonstrated. Studies that discuss configuration of multiple types of gauges [*Bras et al.*, 1988; *Vandenberghe et al.*, 2007] are relatively sparse compared to that of single gauge types [*Mishra and Coulibaly*, 2009]. Generally, the problem of sensor network design in watersheds has been viewed as either a problem of only rain gauge network design [*Bradley et al.*, 2002; *Dong et al.*, 2005; *Bárdossy and Das*, 2008], or only in-stream gauge network design [*Sanders et al.*, 1987; *Strobl et al.*, 2006, 2007; *Telci et al.*, 2009] , with a majority of the focus on rain gauge network design [*Mishra and Coulibaly*, 2009].

In addition to multiple types of gauges, the sensor network's ability to obtain data that will result in a calibrated model with an accurate simulation of multiple

constituents is considered. Previous studies such as the one conducted by Bras et al. [1988] consider a rain and in-stream sampling network for flow prediction for flood monitoring, while Telci et al. [2009] and Strobl et al. [2006] consider an in-stream sampling network to detect phosphorus and other pollutants over a short timeframe. These studies do not help us consider the long-term effects of monitoring.

Furthermore, the methodology uses heuristic combinatorial optimization techniques to optimally locate rain gauges according to statistically based performance measures, rather than geographical criteria, as in Bardossy and Das [2008]. We illustrate this through the use of a Tabu Search. Previous studies for rain gauge configuration methods have used random sampling to determine rain gauge locations. Bradley et al. [2002] assigns random locations and evaluates 40 alternative rain gauge networks for a fixed rain gauge network density. Dong et al. [2005] uses variance of areal rainfall estimation as their criteria selecting good rain gauge networks via random sampling. Other heuristic combinatorial optimization algorithms have been used to optimally locate rain gauges, as in the paper by Bardossy and Das [2008], who use simulated annealing to minimize the mean distance of the station to the whole catchment and to maximize the minimum distance between the stations.

Finally, a huge advantage to the methodology proposed in this paper is that multiple weather scenarios can be generated and used to test the robustness of the sensor network configuration. The weather scenarios generated are by a parameterized weather generator that is tailored to the watershed. The parameters of the weather

generator can be set to model average weather conditions in the watershed, or be set to generate potential weather scenarios under climate change conditions. Most of the methods discussed previously use historical data as inputs into their hydrological models. For example, the "true" precipitation data for rainfall in the study by Bardossy and Das [2008] were generated via interpolation of historical rain gauge data. In the study by Bradley et al [2002], NEXRAD radar data was used as the "true" precipitation over the watershed. Therefore the ability to determine how the sensor network is affected by different weather patterns for previous studies is limited to what has been seen in the past, with no ability to extrapolate to possible hypothetical climate conditions.

To summarize, this methodology expands on what has previously been done by configuring multiple types of gauges (rain and in-stream gauges) for long-term monitoring of multiple watershed processes (flow and rainfall generation). Rain gauge locations are optimized using a heuristic combinatorial optimization method optimizing statistically based performance measures, rather than geographical criteria. Finally, the methodology uses a parameterized weather generator to simulate true weather conditions in the watershed and allows us to test various weather patterns that may occur.

CHAPTER 2
METHODOLOGY

2.1 General Framework

The idea in our approach is to develop a hydrologic model of the watershed based on the best available data that accepts variable weather input from a weather generator ("true weather"). This model is then simulated to generate a spatially distibuted time series of the flow and contaminant constituents. This output is assumed to be the "true" observations at potential in-stream gauge sites. The assessment of the monitoring network is then based on the accuracy of the model predictions (from a model with inputs from sampling points in the sensor network) compared to the true observations.

The framework for configuring and evaluating alternative hydrometric sensor network configurations can be described as a three-step process: 1) develop a model for the "true" conditions in the watershed and generate data ; 2) configure and evaluate the different rain gauge networks; and 3) using the best rain gauge network, configure and evaluate the in-stream gauge network. Detailed steps of the framework are as follows:

Step 1: Develop a model for the "true" conditions in the watershed and generate data

Step 1.1: With known weather generator parameters, a parametric, gridded, multisite weather generator [*Wilks*, 2009] is used to generate several precipitation scenarios (TP)

Step 1.2: TP and θ(true), the known hydrologic model parameters (randomly generated or obtained by previous calibrations) are used in the hydrologic model to obtain TI, the true observations at all the possible in-stream gauge sites

Step 2: Configure and evaluate alternative rain gauge networks

For each candidate rain gauge network R and weather scenario:

Step 2.1: A subset of TP, determined by the rain gauge network R configured using a method described in Section 2.3.1 is fed into an error model (equation A.5) to produce OP(R), the observed precipitation data from the rain gauge network.

Step 2.2: OP(R) is used as input into the hydrologic model with true parameter set θ(true) to produce SI(R), the simulated output at each potential in-stream gauge site

Step 2.3: An error model (equation A.7) is used on TI generated in step 1.2 to produce OI, the observed in-stream gauge outputs at each potential in-stream gauge site.

Step 2.4: SI(R) from step 2.2 and OI from step 2.3 are compared, generating a set of the model performance measures Perf(R) from Section 0.

Output: The best rain gauge network, R_{best}

Step 3: Configure and evaluate alternate in-stream gauge networks

For each in-stream gauge network (F) and weather scenario

Step 3.1: A subset of TP, determined by the best rain gauge network R_{best} is fed into an error model (equation A.5) to produce OP(R_{best}), the observed precipitation data from the rain gauge network.

Step 3.2: An error model (equation A.7) is used on TI generated in step 1.2 to produce OI, the observed in-stream gauge outputs at each potential in-stream gauge site.

Step 3.3: Configure in-stream gauge network *F* (e.g. as described in Section 2.3.2), and get OI(F), the observed data from the in-stream gauge network.

Step 3.4: Perform multisite and multivariable calibration as discussed in Section 2.5. Calibration of the model produces $\theta(R_{best},F)$, a set of calibrated parameters as a function of the best rain gauge configuration R_{best}, and the candidate in-stream gauge configuration, F.

Step 3.5: OP(R_{best}) and the calibrated parameters are used as input into the hydrologic model to produce SI(R_{best},F) the simulated model output using input from the best rain gauge network and the candidate in-stream gauge F.

Step 3.6: SI(R_{best},F) from step 3.4 and OI from step 3.2 are compared, generating a set of the model performance measures Perf(R_{best},F) from Section 2.5.

Output: best sensor gauge network ($R_{best,}$ F_{best})

Figure 1 summarizes the steps, inputs, processes used in the methodology (discussed in later sections), and the output from each of the steps.

	Action	Inputs	Process	Output
Step 1	Generate data	Weather generator parameters	Weather generator (Section 2.2)	TP
		Θ(true)	Hydrologic Model (Section 3.2)	TI
		TP		
Step 2	Configure rain gauge network	TP	Error model (Section 2.3.3)	OP(R)
		Rain gauge network, R (section 2.3.1)		
		θ(true)	Hydrologic Model	SI(R)
		OP(R)		
		TI	Error model	OI
		SI(R)	Model performance measure (Section 2.4)	Perf(R)
		OI		
Step 3	Configure in-stream gauge network	best rain gauge network (Rbest),	Error model	OP(Rbest)
		TP		
		TI	Error model	OI
		OI	In-stream gauge configuration (section 2.3.2)	OI(F)
		OP(Rbest),	Multisite & multivariable calibration (section 2.5)	θ(Rbest,F)
		OI(F)		
		Hydrologic model		
		OP(Rbest),	Hydrologic Model	SI(Rbest,F)
		θ(Rbest,F)		
		SI(Rbest, F),	Model performance measure	Perf(Rbest,F)
		OI		

Figure 1: Overview of the sensor networks methodology. Precipitation and in-stream data are abbreviated in the following manner: T = true, O = observed, P = precipitation, I = in-stream output, R = rain gauge network, F = in-stream gauge network. For example, TP is true precipitation, OI(F) is observed in-stream output from in-stream gauge network F. θ(true) denotes the true hydrologic model parameters

while θ(Rbest,F) denotes the hydrologic model parameters calibrated to the (Rbest, F) sensor network.

A few remarks are necessary regarding the framework. The first step is crucial to the methodology framework in that generally, the distribution of precipitation and temperature within the watershed of interest as well as the true model parameters can never be known with certainty. In addition, having several samples of weather data allows for the testing of the sensor network response to spatial and temporal weather variability. Knowing the true distribution of watershed processes allows for a more precise assessment of the sensor network performance. The sensor network configurations are generated step-wise, so as to observe the individual effects of the different types of gauges on model performance.

In the second step, it is important to emphasize that the objective functions for optimization of rain gauge locations, are based on the ability of the rain gauge network to capture the distribution of the rainfall over the watershed. A rain gauge "data point" is data from a particular point in the grid used in the gridded weather generator from the first step. No calibration of the model is performed in this step, and thus the rain gauge selection can be optimized relatively quickly. The performance measures of each rain gauge network is compared to the performance measures of the base case rain gauge network, to evaluate the relative improvement from additional observations from the new network.

Finally in the third step, in-stream gauges are configured. Here, an in-stream "gauge" refers to the output from specific subbasins within the distributed or semi-distributed model used to model the watershed of interest. One calibration data set is generated using the weather generator, and is then used in the calibration (step 3.4). No optimization scheme was developed for in-stream gauge configurations. Rather, techniques used previously such as stream-ranking techniques, as well as dividing the watershed into independent subbasins were used to configure in-stream networks. This is because a calibration with a large number of iterations and for watershed models that take on the order of minutes per simulation, the resulting computation time can be on the order of days – a computationally expensive endeavor. The performance measures used to evaluate the performance of the in-stream gauge network is the same as used in the second step.

2.2 Weather Generation

Inputs in the form of precipitation and temperature are generated using a parametric gridded, multisite weather generator developed by Wilks [2009]. The Wilks multisite weather generator runs single-site weather generators in parallel by forcing them with spatially correlated random number streams. The precipitation and temperature inputs generated are based on existing distributions of both data points at currently gauged sites then interpolated to ungauged sites taking into account the correlations between each gridpoint. Additional details on how precipitation and non precipitation variables are generated can be seen in Appendix 5A.1.

In our methodology, one weather sample is generated for calibration of the model, used in step 3.4 of the framework. Then, several independent weather samples of longer durations are generated for the validation (evaluation) of the model in steps 2.4 and 3.6. Figure 2 shows an example of the weather grid corresponding to the case study watershed. The color corresponds to the average rainfall over a 40-year sample at a particular gridpoint.

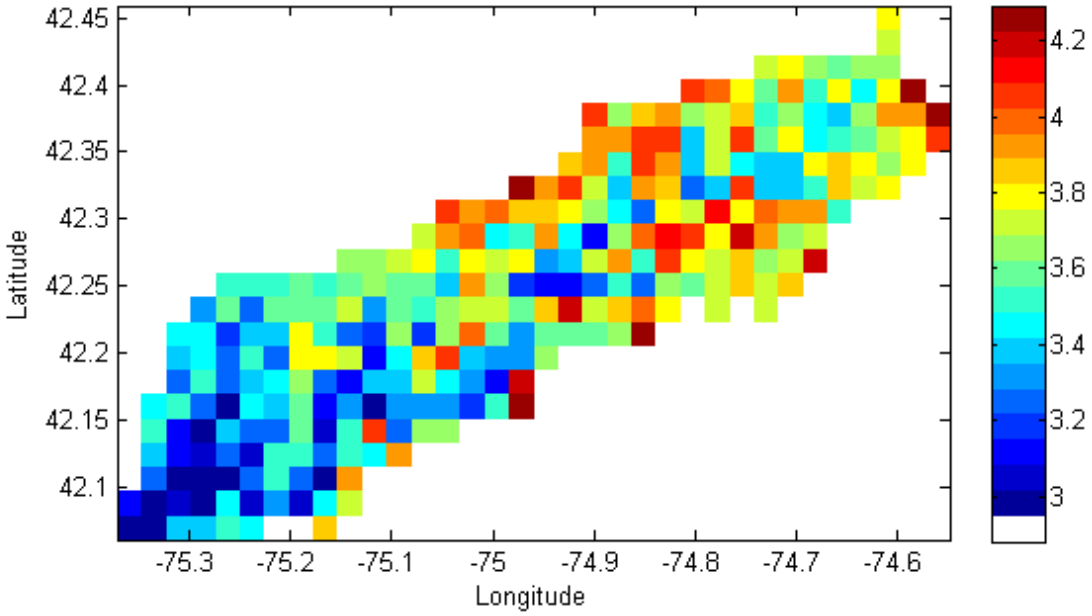


Figure 2: The weather generator grid for the Cannonsville Reservoir. The color for each gridpoint indicates the average daily precipitation (in mm) over a 40-year sample of validation data.

2.3 Sensor Network Configuration

2.3.1 Precipitation Gauge Location Optimization

We employ three different methods in raingauge selection: 1) DET, a deterministic approach based on a weighted distance and elevation score; 2) POPT-VAR, optimal location of raingauges using the mean squared error of areal average rainfall predicted by the gauge configuration; and 3) POPT-MSE, optimal location of raingauges using the variance of areally averaged rainfall.

The first method, DET does not use optimization techniques. Rather, it is a configuration of a raingauge network based on a weighted distance and elevation score that ensures the geographical diversity of the rain gauge locations. Further detail on how this technique is implemented is shown in Appendix 5A.2.

Precipitiation gauge location optimization (POPT) is performed using heuristic methods for solving combinatorial optimization problems. We illustrate the use of Tabu Search [*Sait and Youssef*, 1999]. Tabu Search is a general iterative heuristic with short term memory attributes used for solving combinatorial optimization problems. It is a form of a local neighborhood search. Further detail on the algorithm as well as the pseudocode of our implementation of Tabu search is shown in Appendix 5A.4.

The decision variable for the Tabu search is the n-vector of grid locations corresponding to the locations of the n precipitation gauges in the watershed. After a

user-defined maximum number of iterations have been completed, the Tabu search algorithm outputs the rain gauge configuration network with the best cost. The cost function of the Tabu search is the MSE criterion (gauge networks configured using this criterion are denoted POPT-MSE) or the VAR criterion (POPT-VAR), an extension of research done by Dong et al., [2005]. The VAR criterion is discussed in Appendix 5A.3.

The MSE criterion evaluates how well rain gauge network *i* made up of *Ni* gauges approximate the "true" areal averages by comparing the estimated average predicted by those raingauges, $\overline{p}_{i,t}$ to the "true" areal average at time *t*, μ_t :

$$MSE_i = \frac{1}{N_{obs}} \sum_{t=1}^{N_{obs}} (\overline{p}_{i,t} - \mu_t)^2 \quad (1)$$

where

$$\overline{p}_{i,t} = \frac{1}{N_{sub}} \frac{1}{Npts_k} \sum_{k=1}^{N_{sub}} \sum_{m \in sub_k} \hat{p}_{m,t} \quad (2)$$

is the average areal precipitation predicted by gauge configuration *i*, $\hat{p}_{m,t}$ is the assigned rainfall at gridpoint *m* at time *t*

$$\mu_t = \frac{1}{N_{sub}} \frac{1}{Npts_k} \sum_{k=1}^{N_{sub}} \sum_{m \in sub_k} p_{m,t} \quad (3)$$

is the "true" mean over all gridpoints in the weather generator, $p_{m,t}$ is the "true" rainfall at gridpoint *i* at time *t*, N_{sub} is the number of subbasins delineated in the watershed, $Npts_k$ is the number of points in subbasin *k*, sub_k is the set of points in

subbasin *k*. The best raingauge network i_{best} for each n_i-gauge network should have the lowest MSE.

2.3.2 In-stream Gauge Network Configuration

To select candidate flow gauge sites, two alternative methods were tested: (1) Sharp's procedure [*Sanders et al.*, 1987] to identify and rank stream reaches for data capture, and (2) dividing the watershed into hydrologically independent subbasins and placing gagues at the outlet of these basins [*White and Chaubey*, 2005]. The details of these methods and how they were applied to the Cannonsville watershed are discussed in Appendices 5A.5 and 5A.6 respectively. For semidistributed watershed models (e.g. SWAT) that only have distributive features up to the subbasin level, subbasins containing the selected stream reach report its output for the calibration; therefore a flow gauge configuration is a set of subbasins that contain stream reaches of interest. In this study, we assume flow and water quality measurements are taken at the same location and at the same frequency.

2.3.3 Error Models

Measurement error is inherent in every data capture process. This error is simulated for rain and flow measurement capture using a multiplicative error model, and is discussed in more detail in Appendix 5A.7. A multiplicative error model ensures that no negative observed flows or precipitation will be generated and that zero flows will remain zero.

2.4 Evaluating the Performance of the Sensor Network

The Nash-Sutcliffe efficiency (NSE) is one of the most commonly used criteria for calibration and evaluation of model performance of hydrological models with observed data [*Gupta et al.*, 2009]. The NSE criterion [*Nash and Sutcliffe*, 1970] is a convenient and normalized measure of model performance, in that NSE values of less than or equal to zero indicate that the model is only as good as using the mean of the observed time series as a constant predictor. However, [*Schaefli and Gupta*, 2007] warn that the NSE does not measure how good a model is in absolute terms, and should always be compared to an appropriate reference value for proper interpretation. In this paper, this issue is addressed by comparing the performance of all sensor networks to a base case sensor network configuration.

Since the ground truth (TP and TI) and subsequently the observed data (OP and OI) from the watershed are all known from applying the first step of the methodology, the performance of the model with known parameters and a subset of rain gauge data, or the calibrated model can be evaluated as an average NSE, or $\overline{NSE}$ for the outputs throughout the watershed. This avoids user bias towards sensor networks that work well at arbitrarily chosen key sites, but may not work for the watershed as a whole. This criterion for a constituent *k* is formulated as equation 4:

$$\overline{NSE}_k = \frac{1}{N_{sub}} \sum_{m=1}^{N_{sub}} NSE_{k,m} \tag{4}$$

where $NSE_{k,m}$ is the NSE for constituent *k* at subbasin *m* and is defined as equation 5:

$$NSE_{sub,k} = 1 - \frac{\sum_{t=1}^{Nobs} (x_{sim,k,m,t} - x_{obs,k,m,t})^2}{\sum_{t=1}^{Nobs} (x_{obs,k,m,t} - \mu_{obs,k,m})^2} \quad (5)$$

where $x_{sim,k,m,t} \in \{SI(R)\}$ for evaluating rain gauge network R, $x_{obs,k,m,t} \in \{SI(R_{best}, F)\}$ for evaluating sensor network (R_{best},F), $x_{obs,k,m,t} \in \{OI\}$, and $\mu_{obs,k,m}$ is the mean of the observed constituent k at subbasin m. The $\overline{NSE}$ over the 25 different weather scenarios are shown as a boxplot.

A decomposition of the $\overline{NSE}$ into its components of correlation, relative variability and relative bias provides a diagnostic look at what component of the model has improved and could potentially highlight tradeoffs between the different components [*Gupta et al.*, 2009]. This decomposition is shown in Equation 6:

$$\overline{NSE}_k = 2 * \bar{\alpha}_k * \bar{r}_k - \bar{\alpha}_k^2 - \bar{\beta}_{n,k}^2 \quad (6)$$

where

$$\bar{\alpha}_k = \frac{\sigma_{sim,k}}{\sigma_{obs,k}} \quad (7)$$

$$\bar{\beta}_{n,k} = \frac{\mu_{sim,k} - \mu_{obs,k}}{\sigma_{obs,k}} \quad (8)$$

$\bar{r}_k$ is the average correlation coefficient over all subbasins between x_{sim} and x_{obs}

All the overbars on the components of the NSE represent the average value of these components over all the subbasins delineated in the watershed. This is consistent with the notation in equation 4. The average relative variability, $\bar{\alpha}$ (equation 7)

describes the ability of the model to capture the variability in the flow and has an ideal value of 1. The average normalized bias, $\overline{\beta}_n$ (equation 8) describes the ability of the model to approximate the mean, normalized by the standard deviation of the observed flow time series data. The ideal value for $\overline{\beta}_n$ is 0. The normalization is a convenient way to compare the bias between flow gauges at different locations. Finally, the average correlation coefficient, $\bar{r}$ describes how well the model matches the timing and shape of the hydrographs, with an ideal value of 1.

A novel way of visualizing the three components of NSE, called a Bias Variance coRrelation (BVR)-rectangle plot, is shown in Figure 3. Each combination of performance values are presented as a rectangle, with a line from the center of the rectangle to the top side of the rectangle. The value for correlation is described by the vertical location of the center of the rectangle on the chart. The relative variability is shown as the ratio of the horizontal sides of the rectangle to its vertical sides which are set at unit length. A short squat rectangle indicates low simulated variance relative to the observed variance, or $\overline{\alpha}$ values of less than 1, and a tall skinny rectangle indicating high simulated variance relative to the observed variance, or $\overline{\alpha}$ values of greater than 1. The bias is shown as the tilt of the rectangle relative to the vertical. A negative bias results in a rectangle tilted with the top pointed towards the left, while a positive bias results in a rectangle tilted with the top pointed towards the right. Rather than presenting the values as numbers in a graph, a graphic visualization of the components for the performance provides more insight to the relative performance of

each. In ideal conditions, the relative variability is equal to 1, normalized bias is equal to 0 and the correlation coefficient is equal to one, resulting in a perfect square.

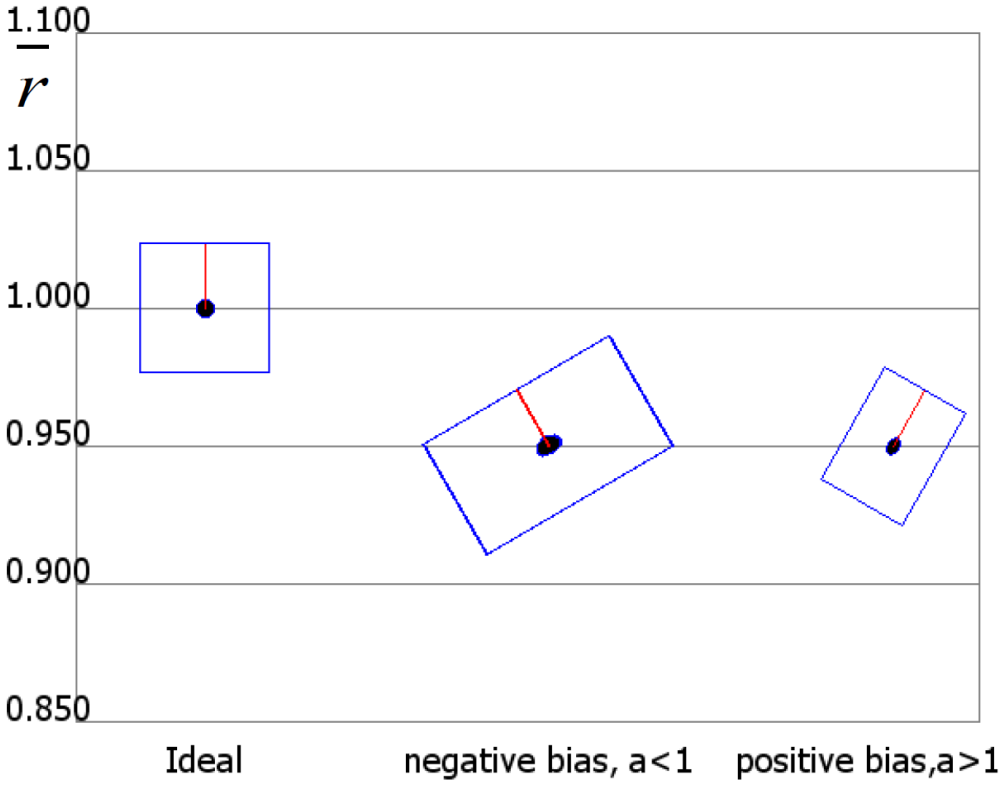


Figure 3: Visualizing the three components of NSE: $\overline{\alpha}$, $\overline{\beta}_n$ and $\overline{r}$ in equation 6. The ideal values of each of the components are $\overline{\alpha} = 1$, $\overline{\beta}_n = 0$, and $\overline{r} = 1$. The positions of the centers of the rectangles show the value of $\overline{r}$. The tilt of the red line relative to the vertical shows $\overline{\beta}_n$. The shape of the rectangle shows $\overline{\alpha}$.

2.5 Multisite and Multivariable calibration

Typically, semi-distributed watershed models (for example, SWAT) require parameters that describe the processes such as baseflow generation and surface runoff within the watershed of interest. Many of these parameters cannot be measured and require calibration in order for the model to properly simulate conditions in the

watershed. Model calibration of multiple parameters, or inverse modeling problems are typically solved via an optimization method that optimizes an objective function which is a function of the error between the simulated constituent and some observed data for a given calibration period.

Automatic calibration of the SWAT model parameters was performed using dynamically dimensioned search (DDS), developed by [*Tolson and Shoemaker*, 2007b] for watershed calibration. DDS is a heuristic, global search algorithm that requires no parameter tuning and automatically scales the search to find good solutions within the maximum number of function evaluations, specified by the user. It has been shown to perform better than other popular automatic calibration methods such as the shuffled complex evolution algorithm for many different types of test functions as well as SWAT model calibrations, finding not only better solutions but converging to a good solution in fewer iterations (function evaluations) [*Tolson and Shoemaker*, 2007b].

Input for calibration is made up of OP(R), weather data (temperature and precipitation) from the rain gauge sensor network, chosen using methods described in Section 2.3.1. The "ground truth" precipitation data TP is transformed into observed data OP(R) using equation A.5.

The objective function for multi-site and multi-variable calibration of the watershed model is a weighted sum over all constituents of the average NSE, taken over the observations from the in-stream gauge configuration at a particular timestep:

$$Objective = \max_{\theta \in \{parameters\}} \left(\sum_{k \in \{constituents\}} \left(w_k * \frac{1}{N_{gauge}} * \sum_{m=1}^{N_{gauge}} NSE_{k,m} \right) \right) \quad (9)$$

where $NSE_{k,m}$ is now defined as the NSE for constituent k at in-stream gauge m (equation 5), with the components in equation 5 now defined as $x_{obs,k,m,t} \in \{SI(R_{best}, F)\}$ for sensor network (R_{best},F), $x_{obs,k,m,t} \in \{OI\}$, and $\mu_{obs,k,m}$ is the mean of the observed constituent k at subbasin m. In equation 9, θ is the parameter set for the hydrologic model, w_k is the weight for the individual constituent, and N_{gauge} is the number of in-stream gauges in the sensor network.

Calibration of the watershed model to observations from multiple sites (multi-site calibration) has been shown to be a pragmatic approach to calibrating models for watersheds with areas similar to the Cannonsville or larger [*Van Griensven and Bauwens*, 2003; *White and Chaubey*, 2005; *Cao et al.*, 2006; *Migliaccio and Chaubey*, 2007; *Zhang et al.*, 2010]. Equation 9 uses an aggregated objective function approach (as in [*White and Chaubey*, 2005; *Zhang et al.*, 2010]).

The calibration problem becomes further complicated when multiple constituents such as flow and water quality (e.g. total dissolved phosphorus) outputs are considered simultaneously. Usually, calibration of flow and sediment parameters

are performed first, then phosphorus parameters are performed. Preliminary investigation for the current SWAT model in our case study showed that this stepwise calibration method provides a good performance for flow, but overpredicts the variability in phosphorus and thus results in poor performance of the phosphorus simulations (shown in Appendix 5A.9). A weighted objective function is proposed to simultaneously calibrate flow, sediment and phosphorus parameters.

One reason for using the NSE as the calibration objective function is that it scales the errors at each gauge by the variance at that gauge, thereby accounting for the different magnitudes of flow between gauge locations that have large and small drainage areas. Moreover, as mentioned in Section 2.4, NSE provides an intuitive sense of whether the model with calibrated parameters simply does not work, i.e. for values of NSE $\leq$ 0, the model performs just as well as a constant predictor function at the mean of the observed values [*Schaefli and Gupta*, 2007].

CHAPTER 3
CASE STUDIES

3.1 Site Description

The Cannonsville watershed is a 1200 km^2 rural watershed located in Catskill region in upstate New York. It is a primarily forested and agricultural watershed, with dairy farming and corn and hay production making up the bulk of the agricultural activity. Less than 1 percent of the watershed is designated as urban. The Cannonsville watershed drains into the Cannonsville Reservoir, which supplies New York City with unfiltered drinking water. Phosphorus contamination of the reservoir from agricultural activities in the watershed has prompted New York City to impose strict agricultural practices in the area [*Tolson and Shoemaker*, 2007a].

The existing gauges in the watershed are used as the "base case" sensor network configuration against which all other sensor networks are compared. Currently there exist five flow gauges, one water quality station, and four rain gauges within this watershed. Only one of the flow and water quality gauges is currently used in calibration of the Soil and Water Assessment Tool (SWAT) model, and is shown in Figure 4. Information from all four rain gauges within the watershed is used in the SWAT model for simulation of the processes in the Cannonsville basin. Together, these four rain gauges and the single flow and water quality gauge are considered as the base case sensor network configuration.

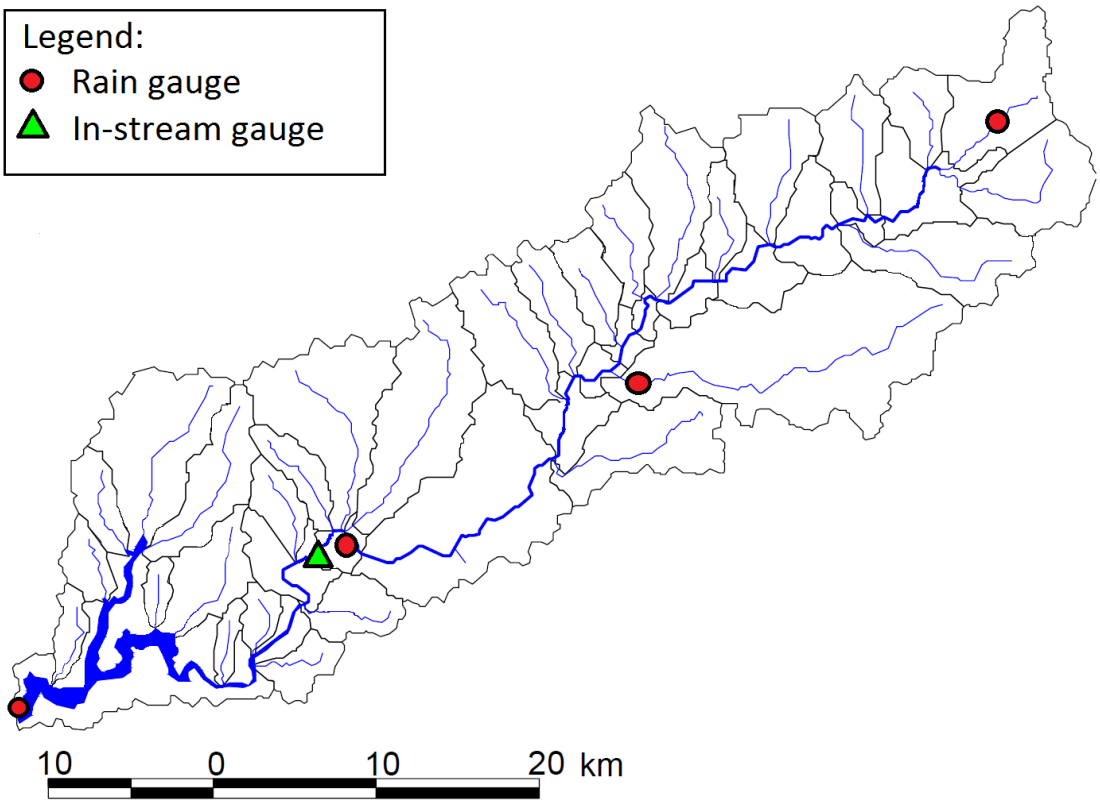


Figure 4: The Cannonsville watershed and locations of the in-stream gauges (triangles) and rain gauges (circles) used for calibration of the SWAT model

SWAT model predictions can possibly be improved with additional gauges. Since the Cannonsville is a relatively large watershed, calibrating watershed parameters to observations at one in-stream gauge location may result in parameter combinations that result in model simulations that do not reflect the true conditions in the watershed. Recall, in this approach, the true conditions refer to TI, the flow and water quality outputs generated from a model with all available input and known parameter values. In addition, more rain gauge data may be needed for this watershed. Figure 5 shows the average rainfall captured by the base case rain gauge network for

the same 40-year validation data set as in Figure 2. This figure shows that the resultant areal average rainfall prediction is much lower than what is actually occuring.

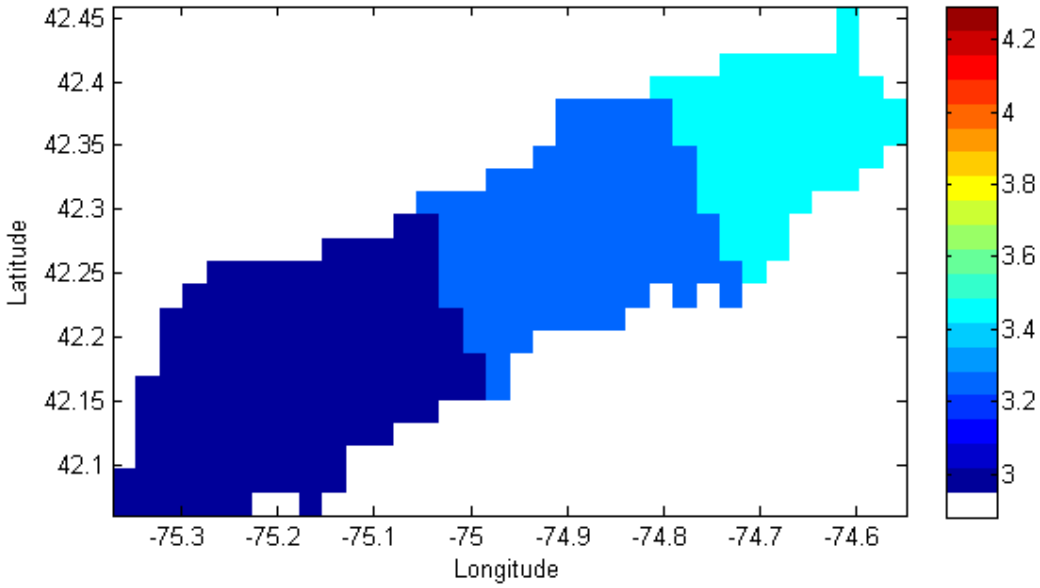


Figure 5 : Average daily precipitation (in mm) over a sample of 40-year generated weather data at each gridpoint of the Cannonsville reservoir as captured by the base case rain gauge configuration.

3.2 SWAT Model Overview

The SWAT model is a distributed-parameter, physically based watershed scale model designed to model processes in largely agricultural and forested watersheds such as the Cannonsville watershed. It is designed to simulate the impact of land management practices in watersheds on the water, sediment, and agricultural chemical yields [*Neitsch et al.*, 2005]. As such, SWAT is designed to model conditions in a watershed over long periods of time, and is not used for single-event simulation such as flood routing or point source pollution detection and warning. Hence SWAT benefits from datasets that span a long timeframe (at least 5 years). Initial development and calibration of the earlier SWAT2000 version of the model was performed by

[*Benaman et al.*, 2005; *Tolson and Shoemaker*, 2007a], and a later version (SWAT2005) for the Cannonsville watershed was developed by [*Easton et al.*, 2008]. This research uses the later version of the model.

SWAT divides the model into subbasins that have a geographic location in the watershed and are spatially related to one another (e.g. outflow from one subbasin number 1 enters subbasin number 3). The subbasins are further subdivided into hydrological response units (HRUs). Each HRU within a subbasin describes the total area within the subbasin that has a unique landuse, management, and soil type combination. HRUs are not spatially related to one another, and do not interact; loadings from each HRU are calculated separately and then summed together to determine the total loadings from the subbasin.

The structure of the SWAT model has a significant impact on how input data is used. Inputs into the SWAT model are weather data in the form of precipitation and daily max and min temperatures, soil type, and land use. Weather data is assigned for the entire subbasin; soil type and land use along with land use management practices are assigned to the HRU level. For weather inputs, the Cannonsville watershed is discretized into 296 2 km * 2 km grids, the centers of which are the potential location of a rain gauge. The grid is consistent with the gridded multisite generator used to generate weather scenarios for calibration and simulation [*Wilks*, 2009] described in Section 2.1. Estimation of areal rainfall for use in the SWAT model is then performed: each gridpoint is assigned to a subbasin in the SWAT model. Then, for each ungauged

gridpoint k the observed rainfall from the closest gauge gridpoint j is assigned to it. The arithmetic average is taken for each subbasin delineated in the SWAT model, and this arithmetic average is used as input into the SWAT model for each subbasin. The location of a raingauge within the subbasin could potentially affect the simulation of the SWAT model, thus the data from a raingauge should be representative of the rainfall within the subbasin. For relatively flat terrain the effects may not be very apparent, but one can imagine that for subbasins with high orographic effects and hence a lot of spatial variability in rainfall this could cause potential problems for runoff calculations within the watershed. For the purposes of this research, land use and soil type is assumed fixed for the entire simulation period.

Thirty-two SWAT parameters were calibrated: 15 for flow, 11 for sediment and 6 for phosphorus. The details for the calibrated parameters are discussed in Appendix 5A.8. The rain gauges are assumed to be unbiased, i.e. $\mu_{\beta,j} = 1$, and $\frac{\sigma_{obs,m}}{\sigma_{true,m}} = 1.05$ in equation A.5, i.e. an inflation of 5 percent over the true standard deviation due to random gauge errors. The value of 1.05 was based on a study performed by [*Bradley et al.*, 2002], who performed a study in the same geographical area. The observed output (OI for flow, sediment, etc.) is obtained from the "ground truth" using equation A.7, with $\sigma^2_{\varepsilon,k,m} = 0.1$. This implies that the gauge captures within 20 percent of the actual flow about 95 percent of the time. This is consistent with the precision of such measurements found in available literature, which is typically 5 to 10 percent accurate [*Hirsch and Costa*, 2004].

3.3 Land Use Cases

As a further verification process the methodology was tested on three different land use cases within the watershed: 1) the existing land use case (referred to as Cville1), 2) all-forested watershed (referred to as Cville2), and 3) the watershed with all agricultural activity moved upstream (referred to as Cville3).

The different landuse cases affect the runoff volume and nutrient loading of the watershed. Figures 6-8 illustrate how the land use changes affect the nutrient loading in the watershed. In the existing land use case (Cville1) shown in Figure 6, the high phosphorus yield areas are located at the bottom of the watershed.

The two hypothetical land use cases Cville2 and Cville3 can be viewed as completely different watersheds compared to the existing land use case. When all agricultural activity (pasture, hay, and corn land uses) are converted to forest in the Cville2 land use case, the total amount of phosphorus contribution decreases relative to the Cville1 land use case, and the high yield phosphorus areas (shown in Figure 7) are more evenly distributed throughout the watershed. Finally, shifting all the agricultural activity upstream as in the Cville3 land use case results in the concentration of high yield phosphorus areas at the top of the watershed, shown in Figure 8. The soil type remains unchanged for all three land use scenarios.

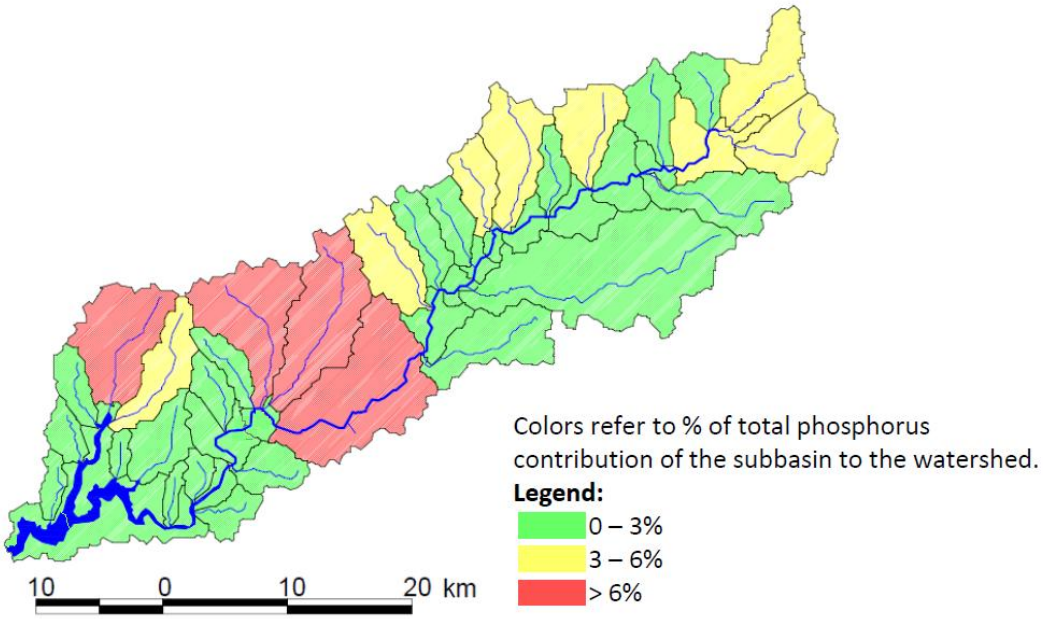


Figure 6: Distribution of phosphorus contribution from each subbasin for the existing land use case, Cville1

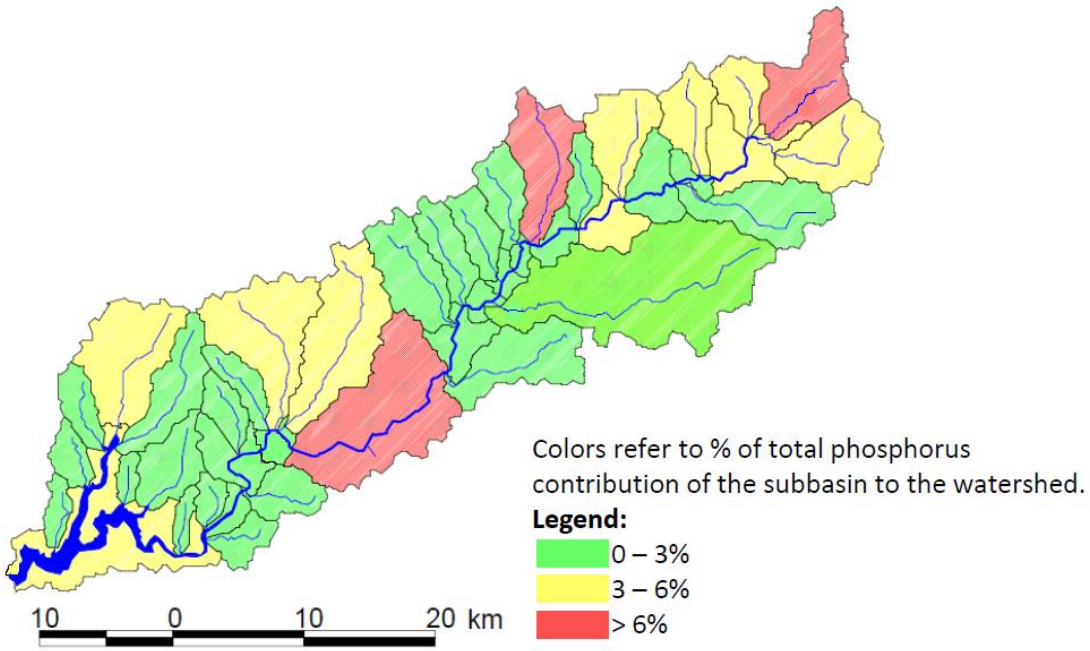


Figure 7: Distribution of phosphorus contribution from each subbasin for an all forested land use case, Cville2

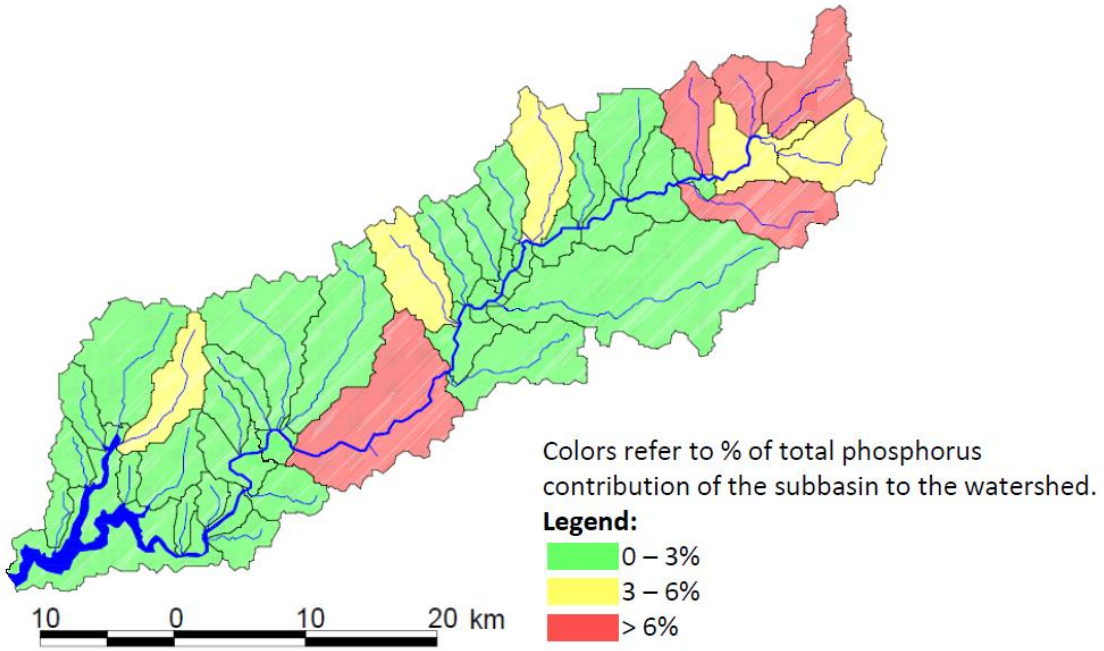


Figure 8: Distribution of phosphorus contribution from each subbasin for the land use case with a majority of agricultural activity shifted upstream, Cville3

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Selecting the Best Rain Gauge Network

Results are presented for the configuration of 5, 10, 15, and 20 rain gauge network using the three methods described in Section 2.3.1. First, a discussion of the results for the existing land use case Cville1 is presented in Section 4.1.1. In particular, this section discusses performance of the rain gauge network as a function of how well the SWAT model with inputs from the rain gauge network simulates flow and total dissolved phosphorus in the watershed, shown by the $\overline{NSE}$ performance criterion (equation 4) and its components (equations 6-8). The effect of different gauge configuration techniques and increasing the number of rain gauges in the configuration is discussed.

Then, the applicability of the methodology in different land use cases is demonstrated in Section 0. In particular, the network configuration method for each of the rain gauge networks is fixed to be the POPT-MSE method.

4.1.1 Existing Land Use Case (Cville1)

A visual representation of the performances of rain gauge networks configured using the three different configuration methods DET, POPT-MSE and POPT-VAR is shown in Figure 9. We show the results for the selection of a 10-rain gauge network for total dissolved phosphorus simulations compared to the performance of the base case rain gauge network shown in Figure 4.

The three methodologies for configuring rain gauge networks consistently results in rain gauge networks that perform better than the base case rain gauge network. Figure 9(a) shows that the $\overline{NSE}$ values for the 25 different weather scenarios are consistently lower for the base case than for the three 10-rain gauge networks (10R DET, 10R POPT-MSE, and 10R POPT-VAR), indicating relatively poor model simulations with input from the base case rain gauge network. The range of $\overline{NSE}$ values are also much larger for the base case rain gauge network, indicating poor performance relative to variable weather patterns within the case study watershed. The BVR-rectangle plot in Figure 9(b) provides an indication as to the poor performance of the base case network – the distribution (mean and variance) of the total dissolved phosphorus are under predicted, and the timing and shape of flow of the total dissolved phosphorus is poorly simulated. Interpretation of BVR-rectangle plots are discussed in Section 2.4.

The POPT-MSE gauge network generally produces superior model performances in terms of $\overline{NSE}$ compared to the DET and POPT-VAR gauge networks. The boxplot in figure 9(a) shows that the $\overline{NSE}$ values for total dissolved phosphorus simulation throughout the watershed (equation 4) with input from the POPT-MSE gauge network is on average higher across the 25 different weather scenarios as compared to the other methods. The smaller range in $\overline{NSE}$ values over the 25 different weather scenarios also show that the POPT-MSE rain gauge network is more robust to variable weather patterns.

The BVR-rectangle plots in figure 9(b) shows that the POPT-MSE rain gauge network produces input data that results in total dissolved phosphorus modeling results that have better representation of the shape and timing of the total dissolved phosphorus flows in the watershed on average. The DET gauge network on the other hand, generally results in total dissolved phosphorus modeling results that have less bias and a more accurate relative variability. Figure 9(a) shows that the $\overline{NSE}$ values for the POPT-MSE gauge network is higher than that of the DET gauge network, while the BVR-rectangle plot indicates that on average, both networks are comparable. The implications for this result is that the correlation coefficient, $\bar{r}$ is the more heavily weighted component of the $\overline{NSE}$ in equation 6. Additionally, there is a tradeoff between getting the shape and timing of the hydrograph right (controlled by the correlation coefficient) and getting the flow distribution (bias and relative variability) right. This tradeoff is not apparent when looking at the $\overline{NSE}$ results alone in figure 9(a).

Similar trends in flow simulations as well as with the 5, 15 and 20-rain gauge networks were observed, shown in Table 1.

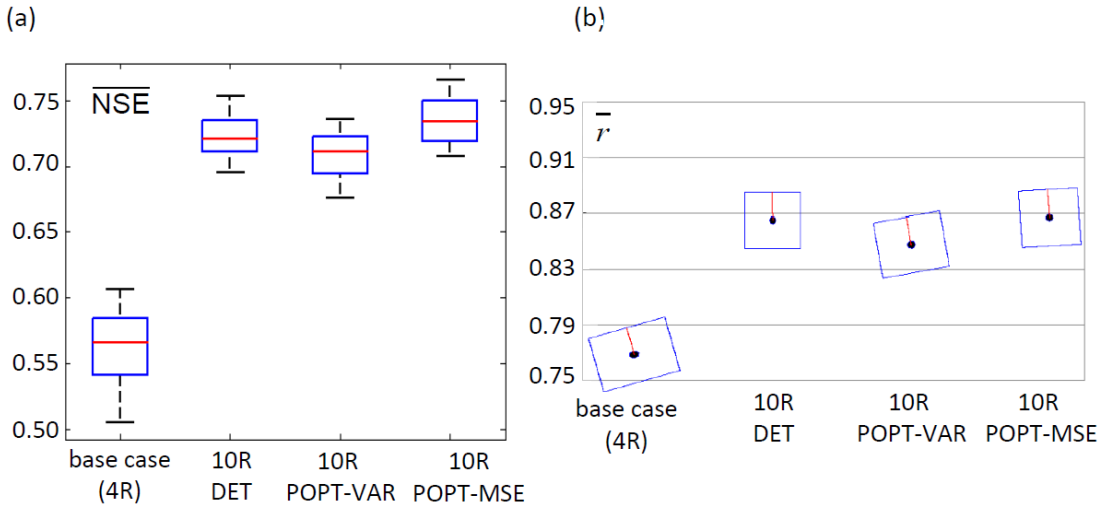


Figure 9: Results for model performance for total dissolved phosphorus simulation for the 10 rain gauge configuration. Part (a) shows a boxplot showing the distribution of $\overline{NSE}$ (equation 4) for total dissolved phosphorus over 25 different weather scenarios and part (b) shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{r}$ averaged over all weather scenarios.

A visual representation of the results for an increasing number of rain gauges in the watershed with land use scenario Cville1 is shown as a boxplot and as a BVR-rectangle plot for total dissolved phosphorus simulation in Figure 10. The rain gauge networks are configured using POPT-MSE method. The boxplot illustrates the trend of improving $\overline{NSE}$ values and the range of the $\overline{NSE}$ over the 25 weather scenarios as the number of rain gauges in a network increases. Improving $\overline{NSE}$ values indicate an improvement in terms of model performance with input from the rain gauge networks. A smaller $\overline{NSE}$ range over the 25 weather scenarios indicate that the rain gauge network is able to better capture variable weather patterns in the watershed and thus is more robust to different weather conditions.

When looking at the individual components of the $\overline{NSE}$, the components $\bar{\alpha}$, $\bar{\beta}_n$, and $\bar{r}$ all improve from the base case up to the 15-rain gauge configuration. The model simulated with input from 15-gauge configuration appears to have a better approximation of the mean and variance of the flow (as indicated by the tilt and the square shape of the BVR-rectangle plot), but has a worse approximation of the shape and timing of the observed output (as indicated by the correlation measure) compared to the 20-rain gauge configuration. Again, the tradeoff between correctly simulating the shape and timing of the total dissolved phosphorus output and correctly estimating the distribution (mean and variance) of the total dissolved phosphorus is seen here.

Generally, $\bar{r}$ seems to be the dominating component of the $\overline{NSE}$. This is demonstrated in both figures 9(a) where the POPT-MSE method has a higher $\overline{NSE}$ score than the deterministic method and 10(a) where the 20-rain gauge configuration has a better $\overline{NSE}$ score than the 15-rain gauge configuration.

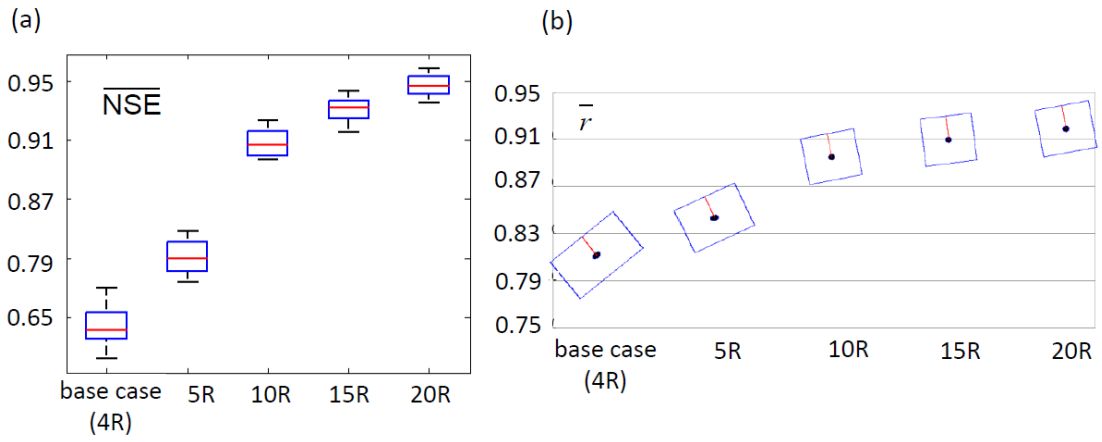


Figure 10: Results for model performance for total dissolved phosphorus simulation for increasing number of rain gauges. Part (a) shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather

scenarios and part (b) shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$ averaged over all weather scenarios.

The best rain gauge network for the exsiting land use case (Cville1) is the 20 rain gauge network chosen using the POPT-MSE method. Figure 11 shows the average rainfall captured by the best rain gauge network for the same 40-year validation data set as in Figure 2 and Figure 5. The best rain gauge network better captures the distribution of rainfall over the watershed compared to the base case rain gauge network shown in Figure 5. Results in this section have demonstrated that when the distribution of rainfall is well captured, the SWAT model is able to capture the timing, shape, mean and variance of both flow and total dissolved phosphorus quite accurately.

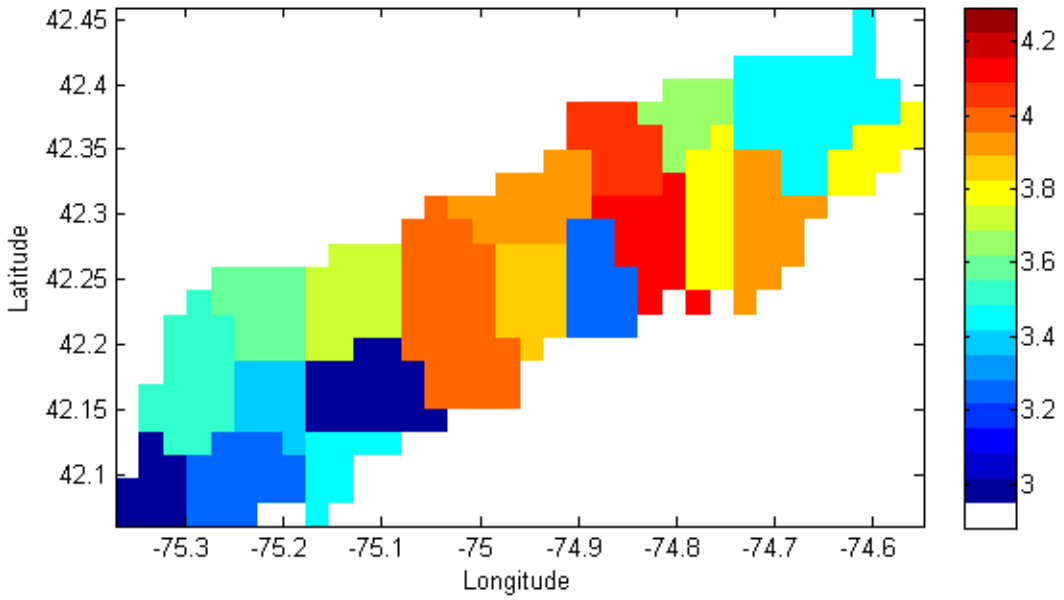


Figure 11: Average daily precipitation (in mm) over a sample of 40-year generated weather data at each gridpoint of the Cannonsville reservoir as captured by the base case rain gauge configuration.

Table 1: Summary of average model performance values for flow and total dissolved phosphorus simulations using observations from different rain gauge network configurations for the Cville1 land use scenario

		NSE	[range]	R	β	α
Flow						
basecase	**N/A**	0.641	[0.059]	0.811	-0.130	0.792
5-gauge	**det**	0.664	[0.048]	0.821	-0.104	0.812
	VAR	0.701	[0.044]	0.843	-0.083	0.857
	MSE	0.701	[0.044]	0.843	-0.083	0.857
10-gauge	**det**	0.777	[0.032]	0.884	-0.024	0.945
	VAR	0.763	[0.041]	0.875	-0.069	0.879
	MSE	0.798	[0.034]	0.895	-0.035	0.937
15-gauge	**det**	0.794	[0.032]	0.893	-0.031	0.926
	VAR	0.787	[0.036]	0.887	-0.064	0.885
	MSE	0.825	[0.034]	0.910	-0.024	0.955
20-gauge	**det**	0.801	[0.032]	0.896	-0.025	0.938
	VAR	0.800	[0.032]	0.894	-0.056	0.897
	MSE	0.846	[0.029]	0.919	-0.035	0.936
Total dissolved phosphorus						
basecase	**N/A**	0.561	[0.101]	0.768	-0.053	0.850
5-gauge	**det**	0.614	[0.096]	0.793	-0.053	0.854
	VAR	0.633	[0.074]	0.805	-0.042	0.886
	MSE	0.633	[0.074]	0.805	-0.042	0.886
10-gauge	**det**	0.724	[0.058]	0.865	-0.002	0.992
	VAR	0.709	[0.060]	0.847	-0.033	0.912
	MSE	0.734	[0.058]	0.867	-0.011	0.973
15-gauge	**det**	0.747	[0.048]	0.872	-0.009	0.964
	VAR	0.741	[0.059]	0.864	-0.030	0.920
	MSE	0.772	[0.046]	0.887	-0.004	0.989
20-gauge	**det**	0.751	0.048	0.875	-0.004	0.983
	VAR	0.752	0.054	0.871	-0.026	0.928
	MSE	0.800	0.043	0.899	-0.011	0.967

4.1.2 Multiple Land Use Cases

The POPT-MSE method is used to configure rain gauges for the three land use cases described in Section 3.3. Table 2 summarizes the results for different rain gauge configurations tested over 25 different weather scenarios and for the three different land use scenarios described in Section 4.1. More detailed results for the Cville2 and Cville3 land use cases can be seen in Appendix B.1 and B.2.

Results for the different land use scenarios show that in general for modeling flow in all land use scenarios, model performance improves as the number of rain gauges placed in the watershed increases. However, for total dissolved phosphorus more rain gauges in the watershed does not necessarily translate into better model performance. The fact that the trend seen in flow simulation performance is not seen in total dissolved phosphorus simulation (which depends on the amount of water in the watershed) performance indicates that an underlying modeling process within SWAT that is pertinent to phosphorus may not necessarily improve with improving flow simulation.

The sensor networks with more rain gauges tend to be more robust to different weather patterns within the watershed as well. This indicates that more rain gauges are able to capture more of the rainfall variability.

Table 2: Summary of average model performance values for flow and total dissolved phosphorus (TDP) simulations with inputs from different rain gauge network configurations chosen using the MSE method and different Cannonsville land use scenarios

	Cville1			
	Flow	[range]	**TDP**	[range]
basecase	0.642	[0.059]	0.561	[0.101]
5R	0.701	[0.044]	0.633	[0.074]
10R	0.798	[0.037]	0.734	[0.058]
15R	0.825	[0.038]	0.772	[0.048]
20R	0.846	[0.041]	0.800	[0.052]
	Cville2			
basecase	0.661	[0.077]	0.57	[0.109]
5R	0.679	[0.053]	0.54	[0.114]
10R	0.741	[0.043]	0.63	[0.068]
15R	0.781	[0.042]	0.73	[0.066]
20R	0.821	[0.038]	0.63	[0.151]
	Cville3			
basecase	0.695	[0.060]	-0.082	[0.435]
5R	0.730	[0.050]	0.426	[0.208]
10R	0.728	[0.052]	0.620	[0.075]
15R	0.722	[0.060]	0.384	[0.197]
20R	0.809	[0.042]	0.436	[0.169]

4.2 Selecting the Best In-stream Gauge Network

Results are presented for the configuration of 3 and 5 in-stream gauge network using the two methods described in Section 2.3.2. The base case sensor network has the four original raingauges and one in-stream gauge, as discussed in Section 3.1. Each of the sensor network configurations are described by the number of rain (denoted R) and in-stream (denoted F) gauges it contains, as well as the method chosen to configure the in-stream gauges. For instance, the (20R, 5FSharp) sensor network configuration denotes a 20-raingauge network and a 5 in-stream gauge network chosen using Sharp's stream ranking method.

First, a discussion of the results for the existing land use case (Cville1) is presented in Section 4.2.1. In particular, this section discusses performance of the in-stream gauge network as a function of how well the SWAT model calibrated to observations from the in-stream gauge network simulates flow and total dissolved phosphorus in the watershed, shown by the $\overline{NSE}$ performance criterion (equation 4) and its components (equations 6-8). The precipitation inputs for calibration and simulation are from the best rain gauge network for the existing land use case. The effect of different techniques and increasing the number of in-stream gauges in the configuration is discussed.

Then, the applicability of the methodology in different land use cases is demonstrated in Section 4.2.2. The effect of phosphorus production as a function of the different land use case is discussed.

4.2.1 Existing Land Use Case (Cville1)

First, the effects of model calibration with additional rain gauge data is investigated. When the model is calibrated at the base case site with observations from the 20-rain gauge network (20R, 1F) instead of the base case raingauge network, an improvement in model performance for flow is seen in Figure 12(a). The BVR-rectangle plot in Figure 12(b) shows that this is because not only is the shape and timing of the hydrograph improved from the base case, but the estimation of the distribution (mean

and variance) improves as well. This indicates that proper capture of rainfall is crucial in proper modeling for flow of the watershed.

In general, the range of the $\overline{NSE}$ values for flow simulation using a model calibrated to the observations from the different sensor networks do not decrease by much relative to the range for $\overline{NSE}$ for the base case network. This is due to the fact that the model parameters that are calibrated are not the same as the "true" parameter values, and some discrepencies in simulation of the watershed processes are bound to occur between all different weather scenarios.

In terms of methods for in-stream gauge selection, Figure 12 shows that adding hydrologically independent in-stream gauges (the "indep" gauge configurations) yields better results in general than placing more gauges upstream from the base case in-stream gauge configuration using Sharp's method. The "indep" in-stream gauge networks result in model performances for flow simulation that are much better at capturing the shape and timing of the observed hydrograph, and also in estimating the distribution of the flows than the "Sharp" in-stream gauge networks.

Increasing the number of in-stream gauges in the watershed to three results in slight improvement from the performance of the model calibrated to the (20R,1F) network in Figure 12. This indicates that for modeling of flow in this watershed, increasing the number of in-stream gauges is not required. Also, the tradeoff between obtaining a good approximation of the distribution of the flows and a good

approximation of the shape and timing of the output graph is again apparent between the (20R, 3Findep) configuration (which has a better approximation of the shape and timing of the hydrograph) and the (20R, 5Findep) configuration (which has a better approximation of the distribution of the flows).

Increasing the number of in-stream gauges in the watershed to five however, does not guarantee an improvement in model performance from the (20R, 1F) sensor network, as evidenced by the (20R, 5Fsharp) network in Figure 12. A possible reason for the trend where model performance suffers when information from more in-stream gauges are included upstream from the base-case in-stream gauge is the fact that the drainage areas for each in-stream gauge site is nested within one another. Information from the catchment area contributing to the most upstream site is included in each of the gauges downstream from it [*Migliaccio and Chaubey*, 2007]. Thus, errors in the most upstream site could be compounded in the subsequent downstream gauges.

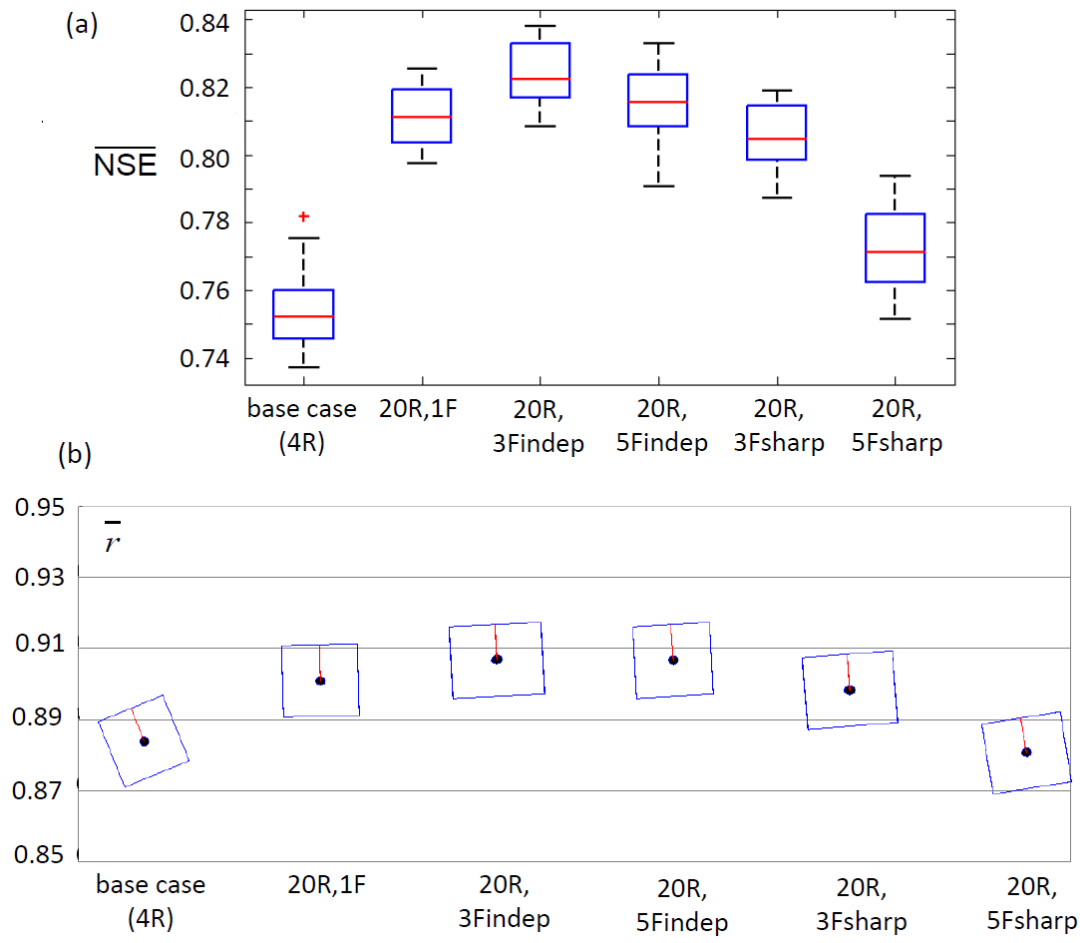


Figure 12: Results for model performance for flow simulation for different sensor network configurations. Part (a) shows a boxplot showing the distribution of $\overline{NSE}$ for flow over 25 different weather scenarios and part (b) shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$ averaged over all weather scenarios.

In terms of the total dissolved phosphorus modeling, the trends are slightly different from flow modeling, shown in Figure 13. The model calibrated to observations at the base case in-stream gauge (both for the base case and also the 20R, 1F configuration) tends to over-predict the variance of the total dissolved phosphorus, as seen in Figure 13(b). When the information from more rain gauges are added to the

basecase rain gauge configuration in the (20R, 1F) configuration, the ability of the model to match the timing and the shape of the total dissolved phosphorus output improves, but the fact that the model overpredicts the variance of the total dissolved phosphorus output results in the relatively poor performance of the sensor network. This indicates that increasing the number of rain gauges is not the most important step needed for total dissolved phosphorus simulation, unlike that in the flow simulation case.

Nutrient modeling in this watershed may benefit from observation from additional in-stream gauging sites. Figure 13(a) shows that adding up to two gauges upstream of the base case in-stream gauge shows improvement (as in the (20R, 3FSharp) case). However, the "indep" in-stream gauge networks like (20R,3Findep) and (20R, 5Findep) have good model performance values as well. The information from additional in-stream gauges generally improve the ability of the model to capture the distribution of the TDP. The timing and shape of the TDP hydrograph is again captured best using information from gauges that are hydrologically independent from one another.

A summary of the results for the different sensor networks in a single land-use case is shown in table 3.

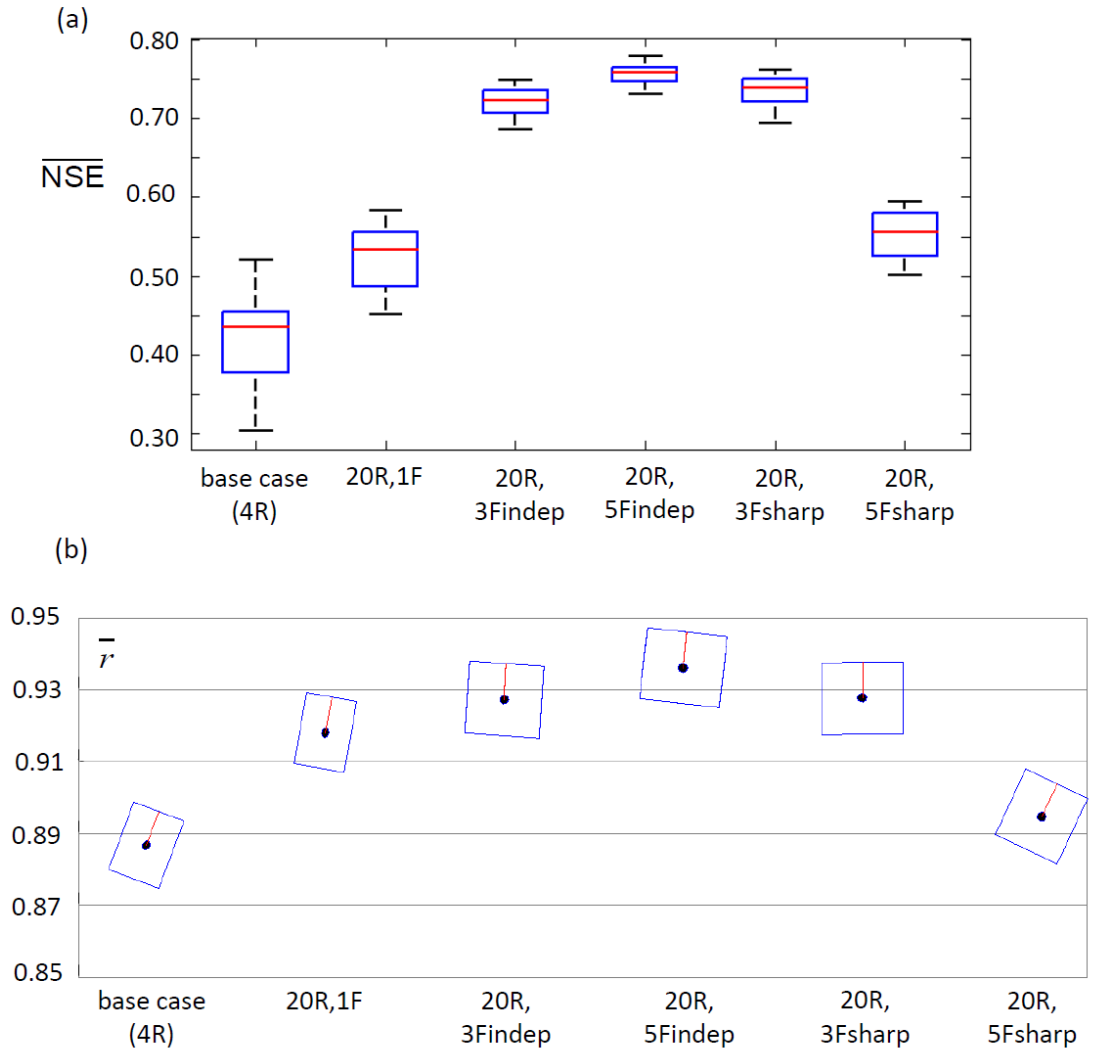


Figure 13: Results for model performance for total dissolved phosphorus simulation for different sensor network configurations. Part (a) shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and part (b) shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta}_n$, and $\overline{R}$.

Table 3: Summary of average model performance values over different weather scenarios for different sensor network configurations

	Flow				
	NSE	**[range]**	$\overline{R}$	$\overline{\beta_n}$	$\overline{\alpha}$
basecase	0.753	[0.045]	0.884	-0.074	1.006
20R,1F	0.811	[0.028]	0.901	-0.004	0.965
20R,3Findep	0.824	[0.030]	0.907	-0.009	0.884
20R, 5Findep	0.816	[0.042]	0.907	-0.011	0.957
20R, 3Fsharp	0.805	[0.032]	0.898	-0.014	0.891
20R, 5Fsharp	0.772	[0.042]	0.881	-0.030	0.945
	Total dissolved phosphorus				
basecase	0.414	[0.216]	0.837	0.070	1.177
20R,1F	0.524	[0.132]	0.868	0.033	1.205
20R,3Findep	0.719	[0.063]	0.877	0.012	0.982
20R, 5Findep	0.756	[0.047]	0.886	0.021	0.948
20R, 3Fsharp	0.734	[0.068]	0.878	-0.001	0.939
20R, 5Fsharp	0.553	[0.093]	0.845	0.083	1.020

4.2.2 Multiple Land Use Cases

The effects of land use on the different ideal in-stream gauges for each of the land use cases are apparent. Table 4 summarizes the results for different sensor network configurations tested over 25 different weather scenarios and for the three different land use scenarios described in Section 3.3. More detailed results for Cville2 and Cville3 land use cases are shown in Appendix 5B.1 and 5B.2

Table 4: Summary of average model performance values for different sensor network configurations. Rbest, the best rain gauge network for each of the land use cases is the 20-rain gauge network for Cville1, the 15-rain gauge network for Cville2, and the 10-rain gauge network for Cville3

	Cville1			
	Flow	[range]	**TDP**	[range]
basecase	0.753	[0.045]	0.414	[0.216]
Rbest ,1F	0.811	[0.028]	0.524	[0.132]
Rbest, 3Findep	0.824	[0.030]	0.719	[0.063]
Rbest, 5Findep	0.816	[0.042]	0.756	[0.047]
Rbest, 3Fsharp	0.805	[0.032]	0.734	[0.068]
Rbest, 5Fsharp	0.772	[0.042]	0.553	[0.093]
	Cville2			
basecase	0.672	[0.077]	0.585	[0.109]
Rbest ,1F	0.738	[0.042]	0.619	[0.068]
Rbest, 3Findep	0.787	[0.028]	0.679	[0.072]
Rbest, 5Findep	0.704	[0.058]	0.657	[0.075]
Rbest, 3Fsharp	0.707	[0.041]	0.663	[0.060]
Rbest, 5Fsharp	0.795	[0.037]	0.732	[0.048]
	Cville3			
basecase	0.695	[0.063]	-0.082	[0.434]
Rbest ,1F	0.701	[0.058]	0.572	[0.106]
Rbest, 3Findep	0.712	[0.045]	0.589	[0.076]
Rbest, 5Findep	0.669	[0.081]	0.560	[0.113]
Rbest, 3Fsharp	0.744	[0.043]	0.675	[0.060]
Rbest, 5Fsharp	0.721	[0.051]	0.569	[0.098]

Some general results remain consistent throughout the three land use cases. As in the Cville1 land use case, flow simulations do not benefit much more from adding in-stream gauges for both the Cville2 and Cville3 land use case. Total dissolved phosphorus simulations generally improve when more in-stream gauges are added.

For the Cville2 land use case, the model calibration benefits from additional observations from two more independent in-stream gauges (as in the (Rbest, 3Findep) network) , or an additional four upstream gauges (as in the (Rbest, 5Fsharp) network).

This is consistently the case for both flow and phosphorus. The (Rbest, 5Fsharp) network works well because of the way the phosphorus is distributed throughout the watershed, as seen in Figure 7.

For the Cville3 land use case, the model calibration benefits from additional in-stream gauge networks selected using Sharp's method rather than the in-stream gauge networks selected using the independent basins method for both flow and total dissolved phosphorus. Again, it appears as though the way the phosphorus is distributed throughout the watershed, Figure 8, is a main factor in the performance.

The results show that the different land use cases affect the ideal in-stream gauge network for each land use case. This is because moving agricultural activity such as corn and hay also affects the amount of water and nutrients generated by the model.

4.3 Proposed Sensor Network for the Cannonsville Watershed (Cville1)

The proposed sensor network for the existing land use case in the Cannonsville watershed is shown in figure 14. The locations of the rain gauge network is well spread out throughout the watershed. For flow calibration, observations at the base case in-stream gauge would suffice, while for phosphorus calibration, observations from all of the five in-stream gauges shown will be beneficial.

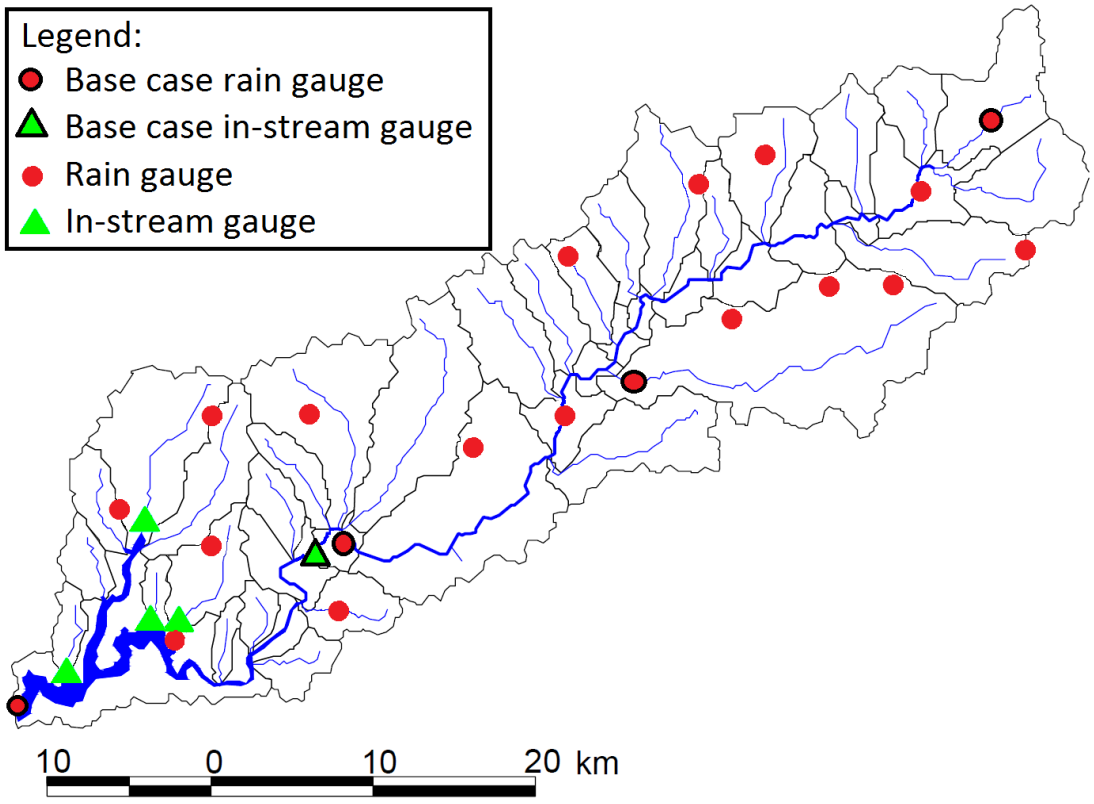


Figure 14: The proposed sensor network for the existing land use case in the Cannonsville watershed

CHAPTER 5
CONCLUSIONS

The methodology discussed in this paper provides a platform for testing several combinations of rain and in-stream gauge configurations in a relatively short amount of time without having to physically place sensors within the environment itself. These sensor network configurations were also tested with many different weather scenarios that each span a period of 30 years to determine the robustness of the sensor network to different weather patterns. The performance of these sensor networks were evaluated and compared against one another through calibration and simulation of the watershed model used by the watershed managers to model conditions in the watershed. It is important to note that the user of this methodology is not limited to the sensor network configuration methods discussed here.

Since a "ground truth" model is used to describe the conditions in the watershed if all inputs and parameters were known, the user is able to gain insight on how additional information from sensor networks affect the modeling of different constituents. This enables the user to make a well-informed decision on the types and quantities of sensors to invest. Specifically for the case study, significant improvements on model performance from the base case configuration for both flow and total phosphorus can be made by adding more rain gauges in the watershed. More rain gauges in the watershed also results in more robust model performance for variable weather patterns as compared to the base case configuration of rain gauges.

As for in-stream gauges, additional investment may not be necessary if the user were concerned with just the modeling of the flows within the watershed, but up to four gauges may be added from hydrologically independent streams to improve the modeling of phosphorus.

The decomposition of a common hydrologic model performance statistic (NSE) into its components lends insight into the various aspects of model performance and how it improves. For instance, it appears that the model performance is largely driven by SWAT's ability to match the timing and shape of the hydrograph (i.e. by the correlation coefficient R). In the case where the model performance is high (i.e high NSE values), figures show that there exists a tradeoff between getting the timing and shape of the hydrograph correctly and the distribution of the flows (bias and variance) correctly. These diagnostics may provide additional insight to watershed managers on what aspects the model is performing well in.

Further studies on this topic are warranted. This paper neglects temporal effects and focuses instead on the spatial effects of sampling; however sampling frequency is another important aspect of sensor network development that should not be ignored. Also, other types of sensors not discussed here such as soil moisture sensors will aid in the development of the understanding of the underlying baseflow in the watershed, a phenomena that is not well understood despite decades of study. Efficient placement of these sensors will be a boon to watershed modelers. In addition, the weather generator used in this study can be parameterized for different types of

climate scenarios, therefore it is a natural step to studying climate change in the watersheds and how better rain and flow gauging can guide the decision-making process.

Further investigation into the multi-site calibration techniques used in this paper can be expanded upon. The single objective optimization assumes that flows and nutrient amounts are independent from one in-stream gauge to another, since no cross correlation effects are considered. Therefore, a more generalized least squares approach that takes into account these cross correlation effects is warranted. A multi-objective approach can also be taken for the modeling of different constituents over at different in-stream gauge sites. This eliminates the required assumption of independence between gauge observations and the subjectivity on the weights on the different constituents. However, this would limit the number of in-stream gauges and constituents to be tested as this would make the problem very large (e.g. for four constituents and five potential in-stream gauge sites the number of objectives can go up to 20) and thus computationally expensive.

With the ever increasing complexity of watershed models, the demand for data which provides useful information to the watershed modeler will surely increase. The wireless sensor network revolution will surely aid in the endeavor to understand better the conditions in the watershed, however efficient placement of these sensors is crucial in order that the information gained from these sensor networks provide the more information with the least cost. The methodology provided in this paper will be a

valuable tool in providing insight into the value of additional data, aiding in investment decisions on hydrometric sensor networks.

APPENDIX A
FURTHER DISCUSSION ON THE METHODOLOGY

A.1 Weather Generation

In the Wilks weather generator [*Wilks*, 2009], the occurrence of precipitation at a currently gauged site is modeled using a two-state first-order Markov chain. Then, conditional on a wet day, the precipitation amounts are drawn from a mixed exponential distribution fitted to historical data. The correlations between occurrences and precipitation amounts at different currently gauged and ungauged sites are also fitted to historical data.

The nonprecipitation variables in the Wilks weather generator (here, daily maximum and minimum temperatures) are generated at a single site using a first order autoregression. In order to perform this analysis, the maximum and minimum temperatures have to be standardized (zero mean, unit variance) where the standardizations are conditional on wet or dry days. This is expanded to simultaneous simulation of minimum and maximum temperatures at different locations by expanding the dimensions of the parameters of the autoregressive formula, and keeping track of the simultaneous correlations between standardized maximum and minimum temperatures at the different locations.

A.2 Deterministic Rain Gauge Configuration

For the deterministic approach, a "base case" raingauge network serves as the starting gauge configuration. The base case rain gauge network can be the existing rain gauge network. The remaining available grid points, *m* are assigned a score based on their proximity and elevation difference from each gauge *n* in the starting gauge configuration, as in equations (A.1-A.3):

$$DET_m = w_1 * D_m + w_2 * H_m \tag{A.1}$$

The DET score is based on the two components: D_m, a distance measure (equation A.2), and H_m, a elevation difference measure (equation A.3).

$$D_m = \frac{\min_n (d_{mn})}{\max_k (\min_n (d_{kn}))} \tag{A.2}$$

In equation A.2, d_{mn} is the distance between ungauged point *m* and gauged point *n*. D_m is the smallest distance d_{mn} over all the n gauged points for an scaled by largest $\min(d_{mn})$ found over all the ungauged points m. The scaling is to give the score D_m a value between 0 and 1.

$$H_m = \frac{\min_n (d_{mn})}{max_k (\min_n (d_{kn}))} \tag{A.3}$$

Similarly, in equation A.3, h_{mn} is the absolute elevation difference between ungauged point *m* and gauged point *n*. H_m is the smallest elevation difference over all the n gauged points scaled by the largest $\min(h_{mn})$ found over all the ungauged points m. The scaling gives the score H_m a value between 0 and 1.

The ungauged point with the highest $score_m$ (equation A.1) will be assigned as the next gauging point, the new rain gauge configuration is updated, and the process is repeated until the desired number of gauges in the configuration is obtained.

The greatest benefit of using the deterministic method is the method's simplicity and speed for solving the gauge placement problem where the number of possible configurations are large. For example placing 10 rain gauges in a site where there are 296 possible locations for a rain gauge could result in 10^{18} different possible combinations. Further, this method ensures geographical diversity in choosing a limited number of rain gauges, which leads to a better idea of the amount and distribution of rain falling within the watershed.

This procedure was tested out for different combinations of w_1 and w_2 on a sample of generated weather data. The results in Figure 15 show the mean squared error of areal rainfall prediction by different rain gauge networks chosen using this method with several combinations of weights w_1 and w_2. The figure shows that a weighting of $w_1 = 0.4$ and $w_2 = 0.6$ in equation A.1 appeared to give the best rain gauge networks with under 20 gauges in terms of predicting the average rainfall over the watershed.

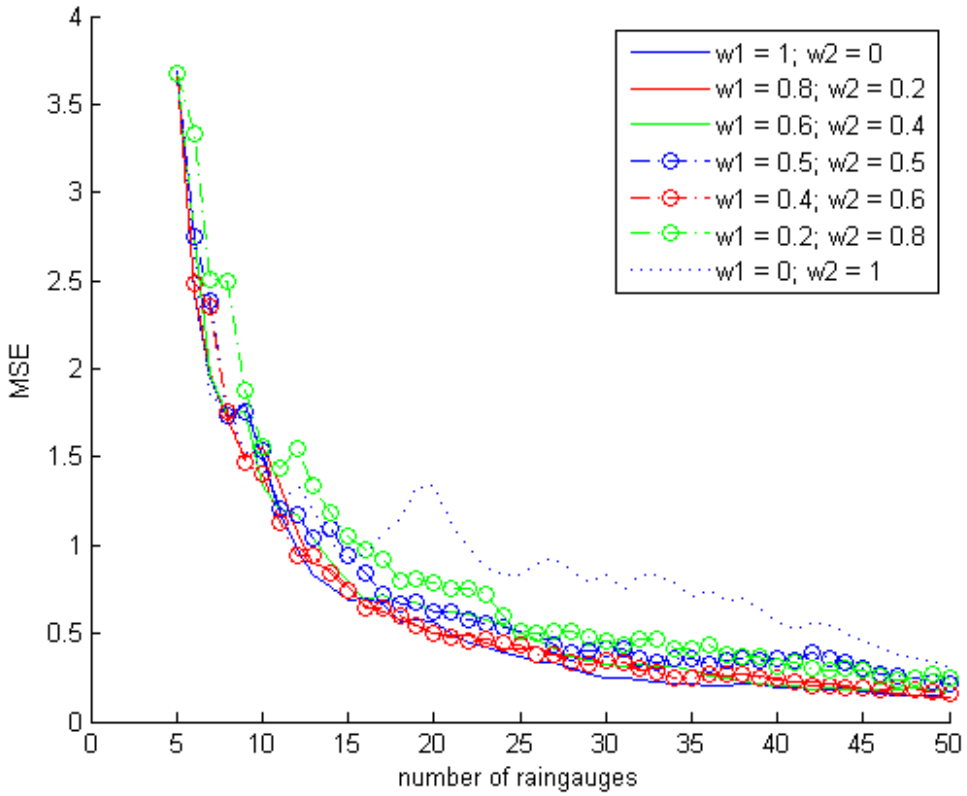


Figure 15: Mean squared error of areal rainfall prediction predicted by rain gauge networks chosen deterministically under different combinations of weights w_1 and w_2 for equation A.1.

A.3 VAR objective function

The VAR criterion describes how well rain gauge network *i* made up of N_i gauges captures the variance of the areal rainfall, rather than areal averages as in the MSE criterion described in Section 2.3.1. This is especially important in large watersheds, where there is a possibility of a large amount of precipitation occurring in one region of the watershed, and almost no precipitation in other parts of the watershed. The variance of the areally averaged rainfall can be formulated as [*Dong et al.*, 2005]:

$$VAR_i = \frac{\bar{s}_i^2}{N_i}[1 + \bar{r}(N_i - 1)] \quad \text{(A.4)}$$

where $\bar{s}_i^2 = \frac{1}{N_i}\sum_{m=1}^{N_i} s_n^2$ is the average temporal variance taken over the gauges in the rain gauge configuration, s_n^2 is the variance of the *n*th gauge, $\bar{r}$ is the mean of the correlation coefficients of all bi-combinations of the rain gauges in the rain gauge configuration *i*.

According to equation A.4, as the number of raingauges approaches infinity, the variance of areally averaged rainfall converges to a linear function of the average point variance and the average correlation coefficient in the watershed [*Dong et al.*, 2005]. The rain gauges are placed so that the estimated VAR for the rain gauge network matches as closely to VAR value calculated if all gridpoints were used.

A.4 Tabu Search Algorithm for Optimization of Precipitation Gauge Location

In our implementation of Tabu Search, the rain gauge configuration consisting of *n* rain gauges at each iteration *i* is defined as $R_i = \{x_1, x_2, ..., x_n : x_k$ is a gridpoint in the weather generator grid$\}$. Each R_i has an associated set of possible moves to candidate solutions $Cand\ (R_i) = \{R_j : R_j$ is in the neighborhood of solution $R_i\}$.

A move is a swap between one of the grid points in R_i with a set of points $g(i) = \{y_1, y_2, ..., y_n\}$ in $G =$ {the set of ordered gridpoints to be swapped in}. The set

G is populated by running the deterministic selection method described in Appendix A.2 until the total number of available gridpoints are exhausted. In each iteration *i*, the set of gridpoints to be swapped in $g(i)$ of length *N* is selected sequentially out of the set *G*. Then the candidate solutions $Cand\ (R_i) = \{R_1, R_2, ..., R_N\}$ are generated by swapping gridpoint x_k in the rain gauge configuration R_i with each of the points in $g(i)$. The elements *k* of R_i to be swapped out are also chosen sequentially.

Among the candidate solutions $Cand\ (R_i)$, the candidate R_j^* with the best cost $Cost\ (R_j^*)$ is selected as the solution for the next iteration R_{i+1}. $Cost\ (R)$ is the MSE (equation 1) or the VAR (equation A.4).

Tabu Search takes advantage of adaptive (flexible) memory by keeping track of a list of previous moves, called the Tabu list, $T(i)$. The length of the list, $Length\ (T)$ is a parameter of the Tabu search algorithm that describes the number of iterations in which a particular move (e.g. swap out grid point *k* for grid point *m* in the Tabu list) is not allowed. Keeping a Tabu list prevents cycling of the algorithm.

A move to a set of candidate solutions from the current solution R_i to obtain the best candidate R_j^* is accepted as long as the move attribute is not in the Tabu list for the current iteration $T(i)$. The cost of the resulting candidate, $Cost\ (R_j^*)$ does not necessarily have to be better than the cost of the current solution, $Cost\ (R_i)$. Therefore,

unlike local search which stops when no new better solution is found in the current neighborhood, Tabu Search allows for a more global search of the solution space.

If the move attribute to obtain candidate solution R_j^* is in the Tabu list, the Tabu can be overridden if $Cost\ (R_j^*)$ is better than the aspiration criterion. The aspiration criterion is defined as the best cost seen so far. If $Cost\ (R_j^*)$ is not better than the aspiration criterion, then no moves are made and a new set $g(i)$ is generated.

Algorithm Tabu Search ();

Ω : Set of feasible rain gauge configurations
N : number of gauges in each rain gauge configuration R_i
R_i : current solution, the set of gridpoints $\{x_1, x_2, ..., x_N\}$
$g(i)$: set of grid points $\{y_1, y_2, ..., y_N\}$ to be swapped in for iteration I
$Cand(R_i)$: set of candidate solutions $\{R_j\}$ that correspond to swapping out x_k in R_i with each of the grid points y in $g(i)$
R^*_j : best solution from candidate solutions $Cand(R_i)$
R_{best} : best rain gauge configuration found so far
$Cost$: Objective function
$T(i)$: Tabu list at iteration i
A : Aspiration level

Begin
Start with an initial rain gauge configuration R_0 and initial set of points $g(0)$
Initialize R_{best}, T and A
For fixed number of iterations ($i = 1$: *maxiterations*) **Do**
Get set of candidate rain gauge configurations $Cand(R_i)$
Find best R^*_j in $Cand(R_i)$
IF move R_i to R^*_j is not in $T(i)$ **THEN**
Accept move and update best solution
Update R_{best}, $T(i)$ and A
ELSE IF $Cost(R^*_j) < A$ **THEN**
Accept move and update best solution
Update R_{best}, $T(i)$ and A
ELSE
$R_{i+1} = R_i$
ENDIF
$i = i + 1$
EndFor
Output R_{best} and A
End

Figure 16: Pseudocode for Tabu Search algorithm adapted from Sait and Youssef [1999]

A.5 In-stream Gauge Configuration Using Sharp's Method

For Sharp's procedure, the farthest upstream portion of the main channel as well as tributaries that do not have other streams flowing into them are given a ranking of 1. At each of the confluence of tributaries and the main channel, the rankings of the tributary and the main channel at that point are summed, until the farthest downstream portion. The first priority stream reach to be gauged is called the centroid of the stream reach, found by dividing the rank at the farthest downstream point by two and finding the stream reach whose rank is closest to that number. Second priority stream reaches are the stream reaches ranked closest to the first quarter and third quarter of the rank at the furthest downstream reach, and lower priority reaches are thus found by continuing to subdivide the rank at the farthest downstream portion by eights, sixteenths, and so forth.

The selection of gauge configurations are done according to the following priorities: 1) the first gauge configuration will contain the highest priority gauge(s), 2) each successive gauge configuration will contain the previous gauge configuration plus additional gauges of equal or lower priority than the lowest priority gauge in the previous gauge configuration, 3) if there is an existing gauge in a subbasin close to the subbasin that contains the current priority stream gauges to be assigned, then the subbasin with the existing flow gauge should be chosen (this is to make sure that we use the information from currently existing gauges).

The gauge locations selected using this method is shown in Figure 17. By the way the stream reaches are ranked, the in-stream gauge locations are necessarily located along the main channel of the Cannonsville watershed.

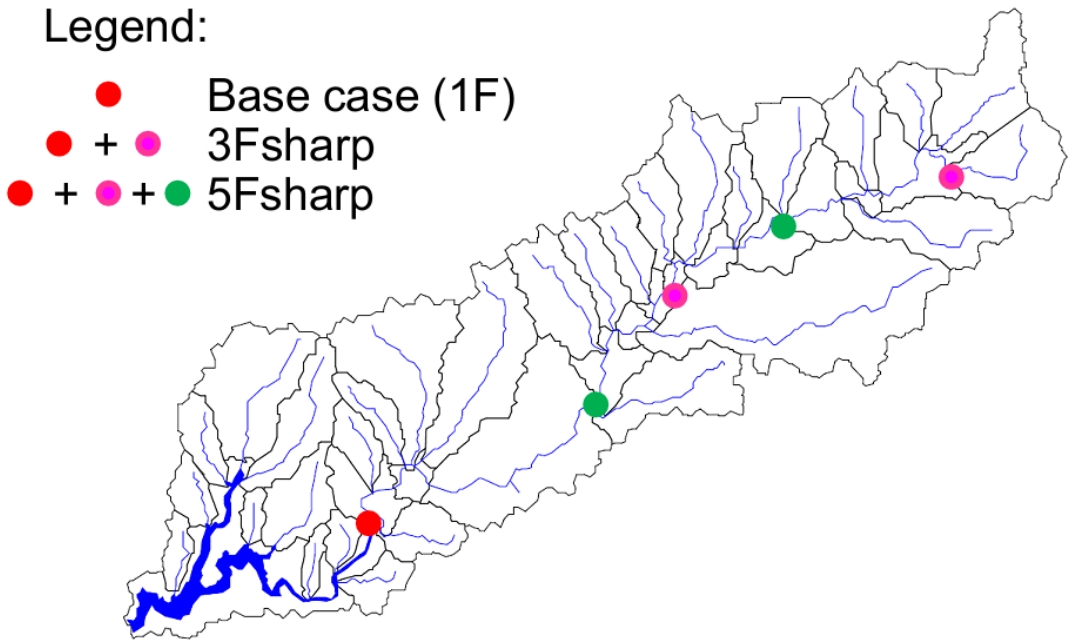


Figure 17: In-stream gauge networks configured according to Sharp's method

A.6 In-stream Gauge Configuration Using the Independent Basins Method

For the second method of locating in-stream gauges, the watershed is divided into several areas that are hydrologically independent, that is, they do not flow into one another. These areas may contain one or several subbasins that were delineated in the semi-distributed watershed model. Then, a potential gauging site is identified as the outlet of each of these hydrologically independent areas. Gauges chosen to be in the in-stream gauge network are prioritized by drainage area behind the gauge site.

The in-stream gauge networks configured for the Cannonsville watershed using the independent basins method is shown in Figure 18.

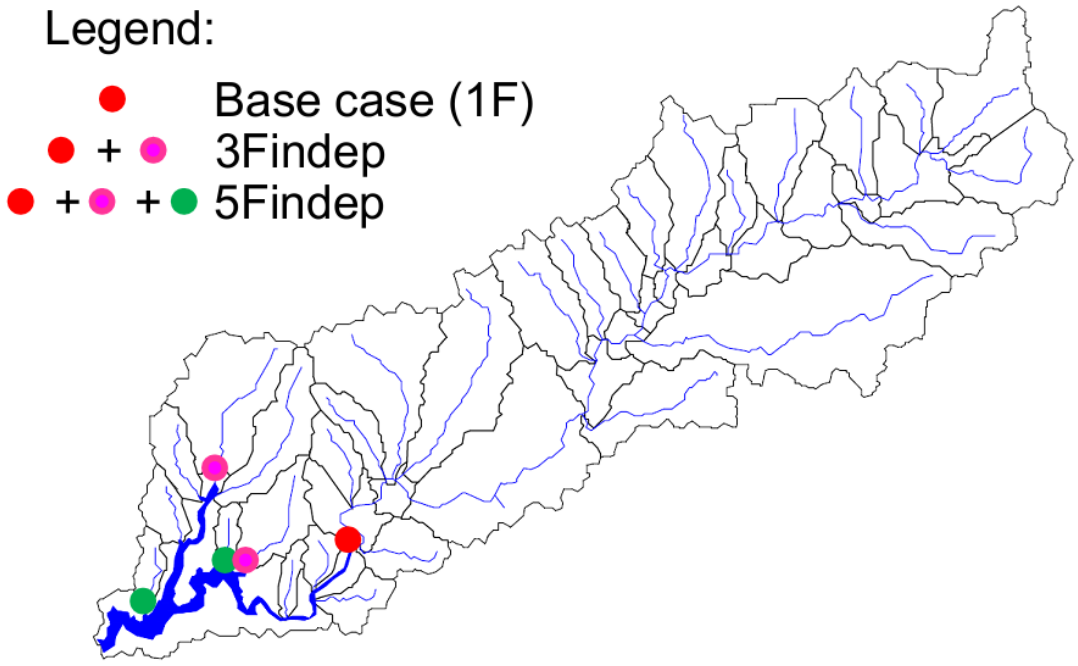


Figure 18: In-stream gauge networks configured using the independent basins method

A.7 Error model equations

For precipitation at gauge m and time t (denoted by $P_{m,t}$), the "observed" data is obtained using equation A.5 [*Bradley et al.*, 2002]:

$$P_{obs,m,t} = P_{true,m,t} * \varepsilon_{P,m,t} \qquad \text{(A.5)}$$

where the error has a log-normal distribution $\varepsilon_{P,m,t} \sim LN(\mu_{\beta,m}, \sigma^2_{\beta,m})$ and the parameters $\mu_{\beta,m}$ and $\sigma^2_{\beta,m}$ are user-defined.

The parameter $\mu_{\beta,m} = \frac{\mu_{observed,m}}{\mu_{true,m}}$ represents the bias of the observed precipitation in terms of the true precipitation, where a value of 1 indicates an

unbiased guage, a value greater than 1 indicates a positively biased gauge and a value less than 1 indicates a negatively biased gauge. The variance of the log-normally distributed error is given by

$$\sigma_{\beta,m}^2 = [(\frac{\sigma_{obs}^2}{\sigma_{true,m}})^2] * (\frac{C_{p,m}^2}{1+C_{p,m}^2}) \quad \text{(A.6)}$$

where $C_{p,m}^2 = \frac{\sigma_{true,m}}{\mu_{true,m}}$ is the coefficient of variation for the true precipitation site m, while $(\frac{\sigma_{obs}^2}{\sigma_{true,m}})^2$ represents the inflation of the standard deviation of the precipitation measured by gauge m as a result of random gauge errors.

For each constituent *k* at gauge *m* and time *t* (denoted by $x_{obs,k,m,t}$), the "observed" data is obtained using Equation A.7

$$x_{obs,k,m,t} = x_{true,k,m,t} * \varepsilon_{x,k,m,t} \quad \text{(A.7)}$$

where the error has a normal distribution $\varepsilon_{P,m,t} \sim N(1, \sigma^2)$ and the parameter σ^2 is user-defined. Note that we have assumed that the error of the in-stream gauge readings are unbiased.

A.8 Calibrated SWAT Parameters

The parameters that are calibrated for the SWAT model are shown in Table 5.

Table 5: Calibrated SWAT parameters [*Neitsch et al.*, 2004; *Woodbury et al.*, 2010]

Parameter	Description	Constituent
SFTMP	Snowfall temperature [°C]	Flow

SMTMP	Snow melt base temperature [°C]	Flow
SMFMX	Melt factor for snow on June 21 [mm H_2O/°C-day]	Flow
TIMP	Snow pack temperature lag factor	Flow
SURLAG	Surface runoff lag coefficient	Flow
GW_DELAY	Groundwater delay time [days]	Flow
ALPHA_BF	Baseflow alpha factor	Flow
GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur [mm H_2O].	Flow
LAT_TTIME	Lateral flow travel time [days]	Flow
ESCO	Soil evaporation compensation factor	Flow
CN2	Initial SCS curve number for moisture condition II. Values can be referenced in a table provided by the SCS.	Flow
DEPTH (SOL_Z) (multiplicative)	Depth from soil surface to bottom of layer [mm]	Flow
BD (SOL_BD)	Moist bulk density [Mg/m^3 or g/cm^3]	Flow
AWC (SOL AWC)	Available water capacity (plant available water) of the soil layer [mm H_2O/mm soil].	Flow
KSAT (SOL_K)	Saturated hydraulic conductivity	Flow
ADJ_PKR	Peak rate adjustment factor for sediment routing in the subbasin (tributary channels)	Sediment
PRF	Peak rate adjustment factor for sediment routing in the main channel	Sediment
SPCON	Linear parameter for calculating the maximum amount of sediment that can	Sediment

	be reentrained during channel sediment routing $\max conc = SPCON * v_{ch,pk}^{SPEXP}$	
SPEXP	Exponent parameter for calculating sediment reentrained in channel sediment routing	Sediment
LAT_SED	Sediment concentration in lateral and groundwater flow (mg/L)	Sediment
SLSUBBSN_f (multiplicative)	Average slope length (m), or the distance that sheet flow is the dominant surface runoff flow process	Sediment
SLSOIL_f (multiplicative)	Slope length for lateral subsurface flow (m)	Sediment
CH_EROD	Channel erodibility factor	Sediment
CLAY_f (multiplicative)	Clay content (% soil weight)	Sediment
ROCK_f (multiplicative)	Rock content (% soil weight)	Sediment
MUSLE_adj	Adjustment factor for the MUSLE equation	Sediment
PPERCO	Phosphorus percolation coefficient (10 m^3/Mg)	Phosphorus
PHOSKD	Phosphorus soil partitioning coefficient (m^3/Mg)	Phosphorus
CMN	Rate factor for humus mineralization of active organic nutrients	Phosphorus
P_UPDIS	Phosphorus uptake distribution parameter	Phosphorus
ERORGP	Phosphorus enrichment ratio for loading with sediment, where the enrichment ratio is the ratio of the concentration of phosphorus transported with the sediment to the concentration of phosphorus in the soil surface layer.	Phosphorus

PSP	Phosphorus availability index, i.e. the fraction of fertilizer phosphorus which is in solution after an incubation period	Phosphorus

A.9 Determining the Weights for the Weighted Objective Function for Calibration

Four different combinations of weights were tested for the calibration objective function (equation 9) to determine which constituent should be more heavily weighted. Case 1 is the set of weights where flow, sediment and phosphorus parameters are equally weighted: $w_{flow} = w_{sed} = \frac{1}{3}$, $w_{TDP} = w_{PP} = \frac{1}{6}$. Case 2 is the set of weights where flow parameters are weighted more: $w_{flow} = \frac{1}{2}, w_{sed} = \frac{1}{4}$, $w_{TDP} = w_{PP} = \frac{1}{8}$. Case 3 is the set of weights where phosphorus parameters are weighted more: $w_{flow} = w_{sed} = w_{TDP} = w_{PP} = \frac{1}{4}$. Finally, case 4 is the set of weights where sediment parameters weighted more: $w_{flow} = w_{TDP} = w_{PP} = \frac{1}{5}, w_{sed} = \frac{2}{5}$. The simultaneous calibration is also compared to the sequential calibration method. The preliminary results are summarized in

Table 6. Based on the results, the weights in Case 4 are used for the simultaneous calibration of all constituents for the SWAT model.

Table 6: Summary of NSE values at a single site for each of the constituents and each of the weight cases compared to the sequential calibration method

	OF value	NSE_{flow}	NSE_{sed}	NSE_{TDP}	NSE_{PP}
Case 1	0.94	0.96	0.94	0.93	0.91
Case 2	0.92	0.94	0.93	0.90	0.88
Case 3	0.92	0.95	0.93	0.92	0.90
Case 4	0.95	0.97	0.95	0.95	0.92
Sequential	N/A	0.98	0.96	0.31	-3.96

APPENDIX B
ADDITIONAL RESULTS

B.1 Extended Results for Cville2 Land Use Case

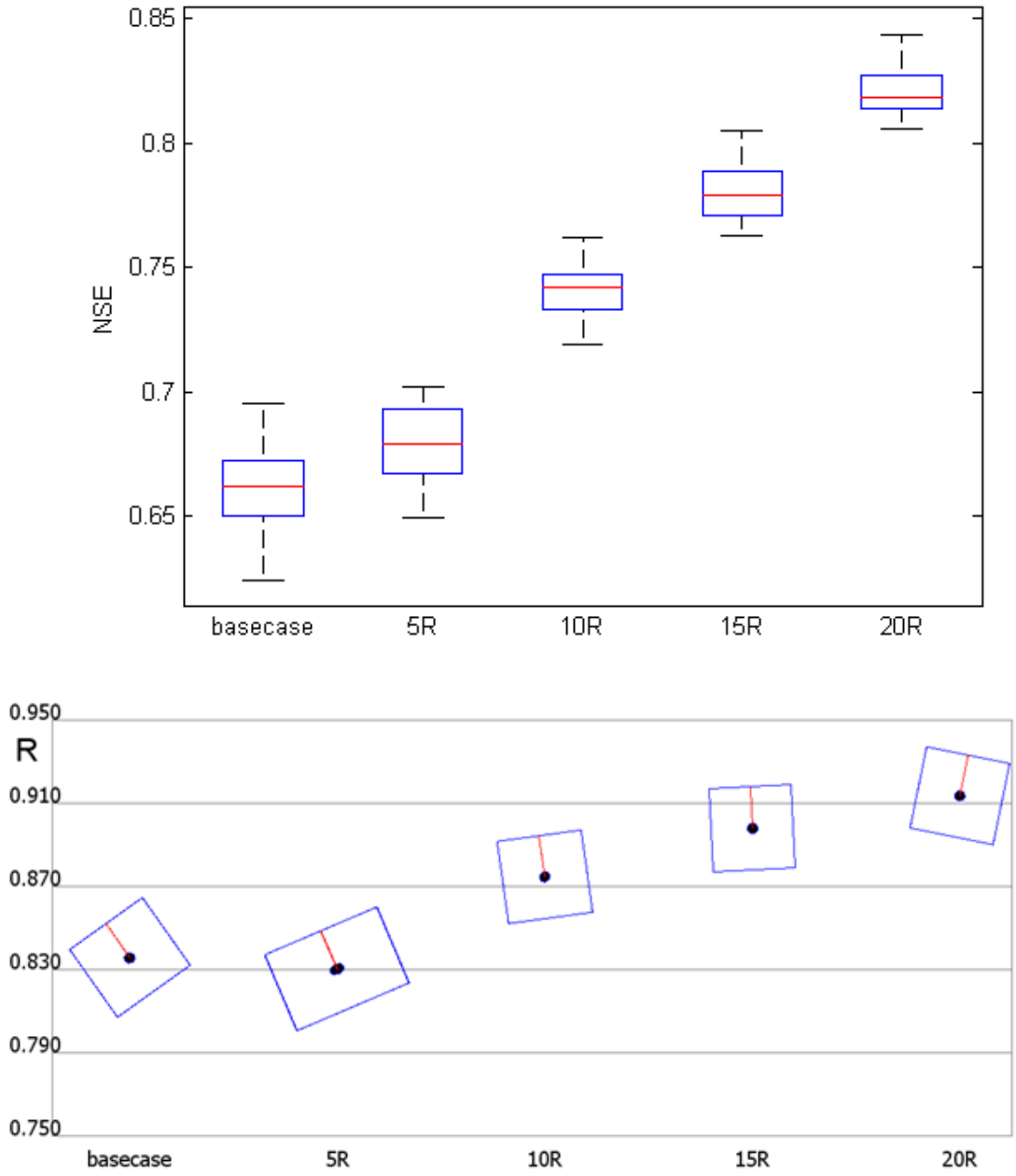


Figure 19: Results for model performance for flow simulation for increasing number of rain gauges in the Cville2 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-

rectangle plot showing the $\overline{\alpha}$, $\overline{\beta}_n$, and $\overline{R}$ averaged over all weather scenarios.

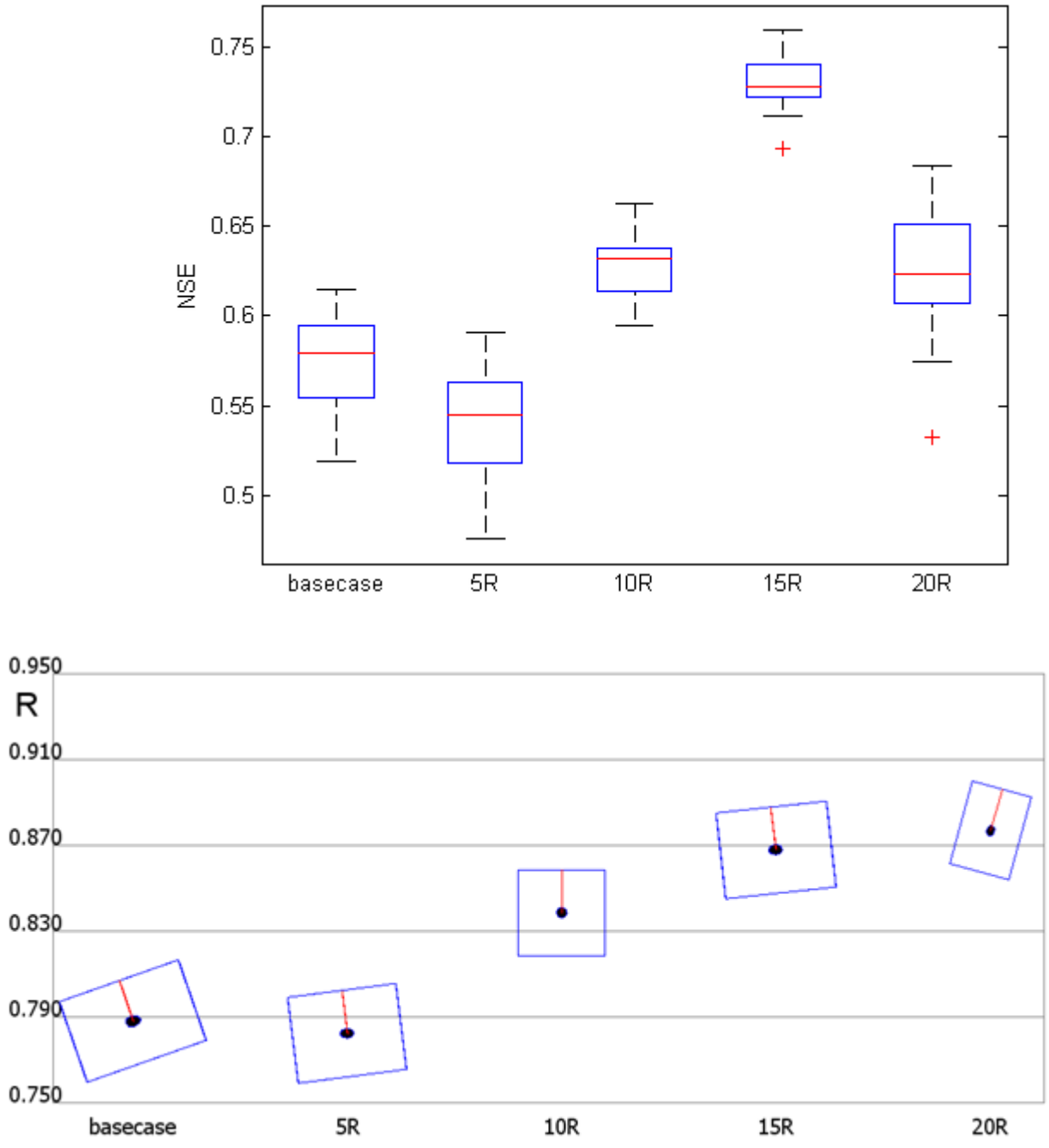


Figure 20: Results for model performance for total dissolved phosphorus simulation for increasing number of rain gauges in the Cville2 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom

figure shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$ averaged over all weather scenarios.

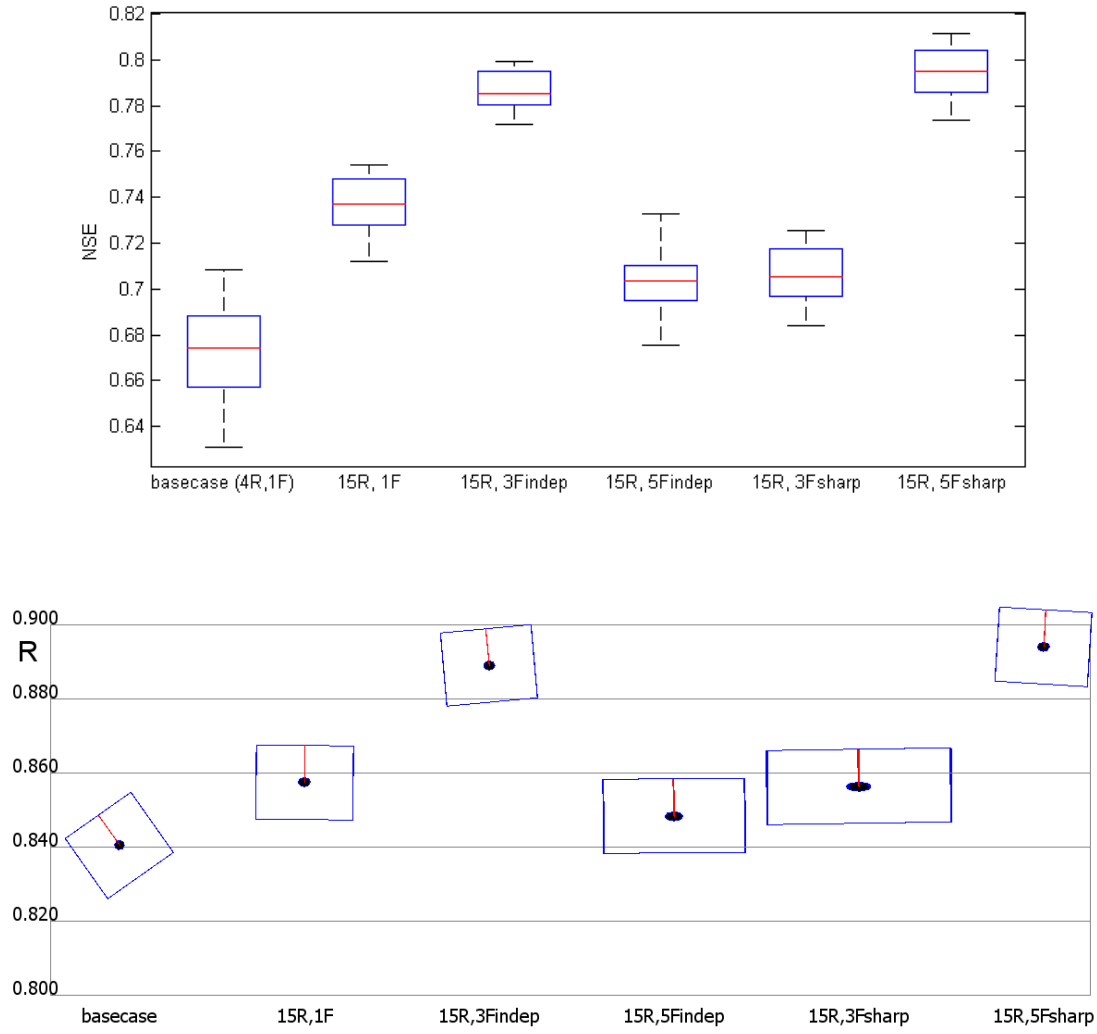


Figure 21: Results for model performance for flow simulation for different sensor network configurations in the Cville2 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$.

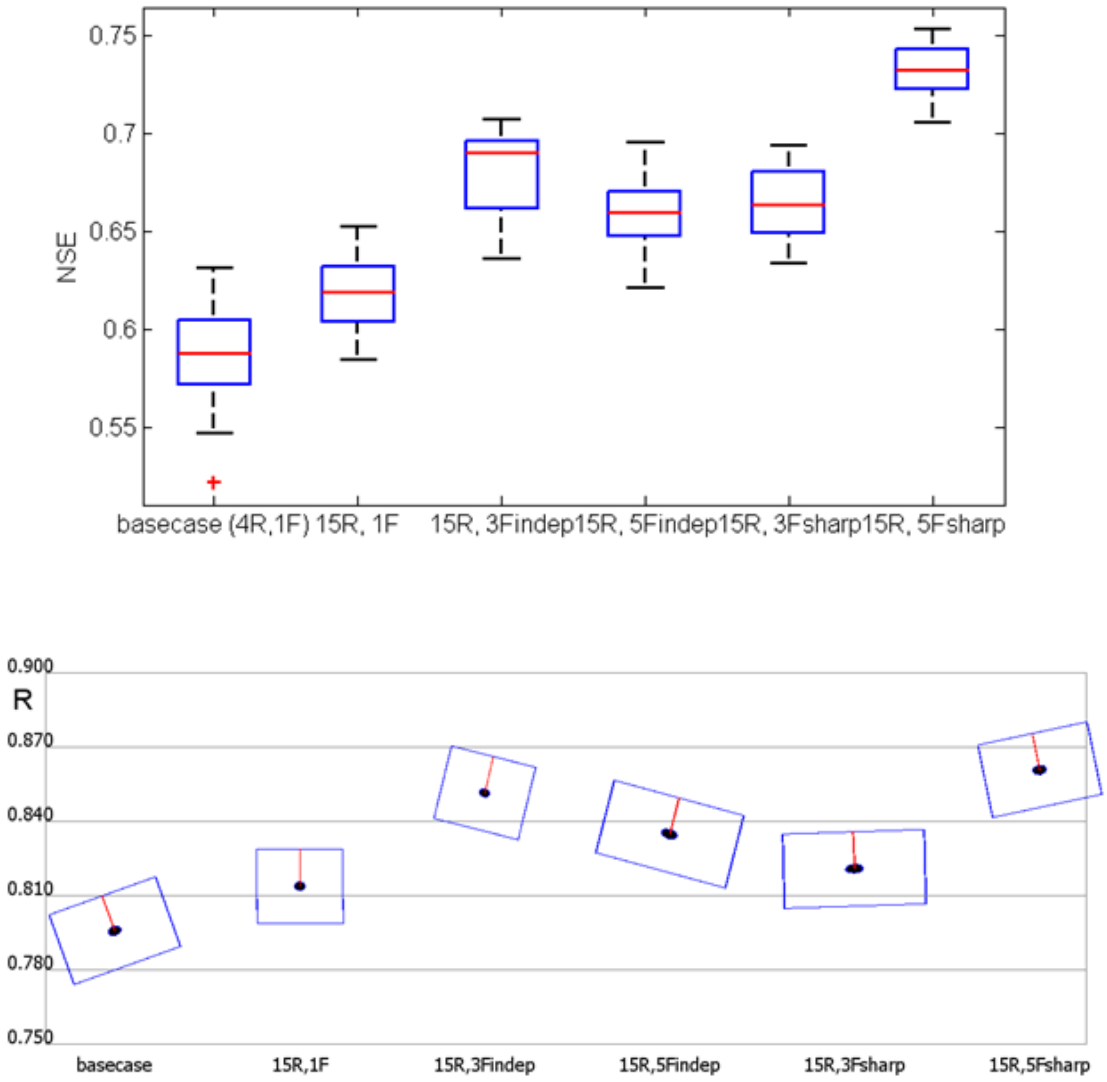


Figure 22: Results for model performance for total dissolved phosphorus simulation for different sensor network configurations in the Cville2 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta}_n$, and $\overline{R}$.

B.2 Extended results for the Cville3 Land Use Case

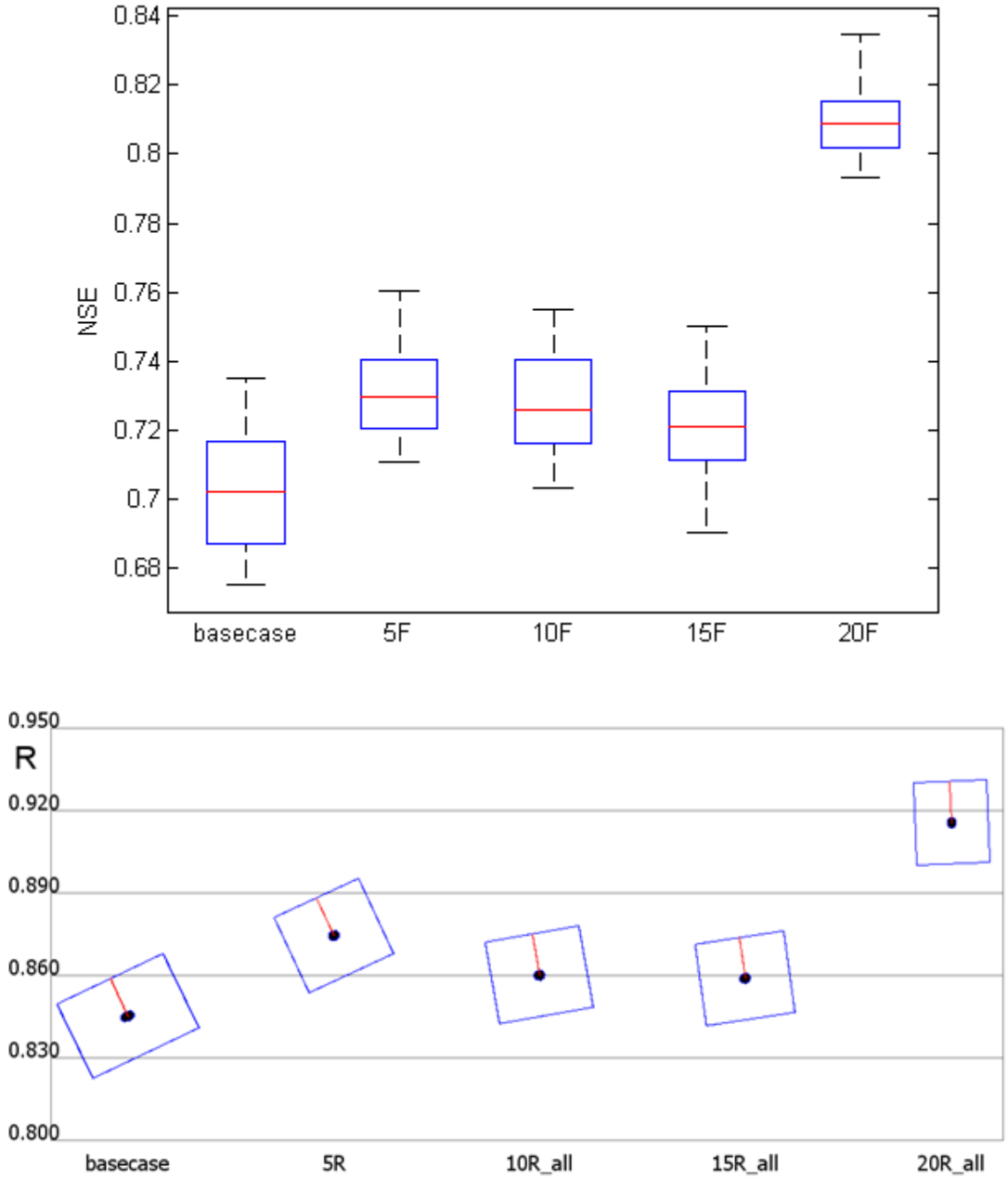


Figure 23: Results for model performance for flow simulation for increasing number of rain gauges in the Cville2 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-

rectangle plot showing the $\overline{\alpha}$, $\overline{\beta}_n$, and $\overline{R}$ averaged over all weather scenarios.

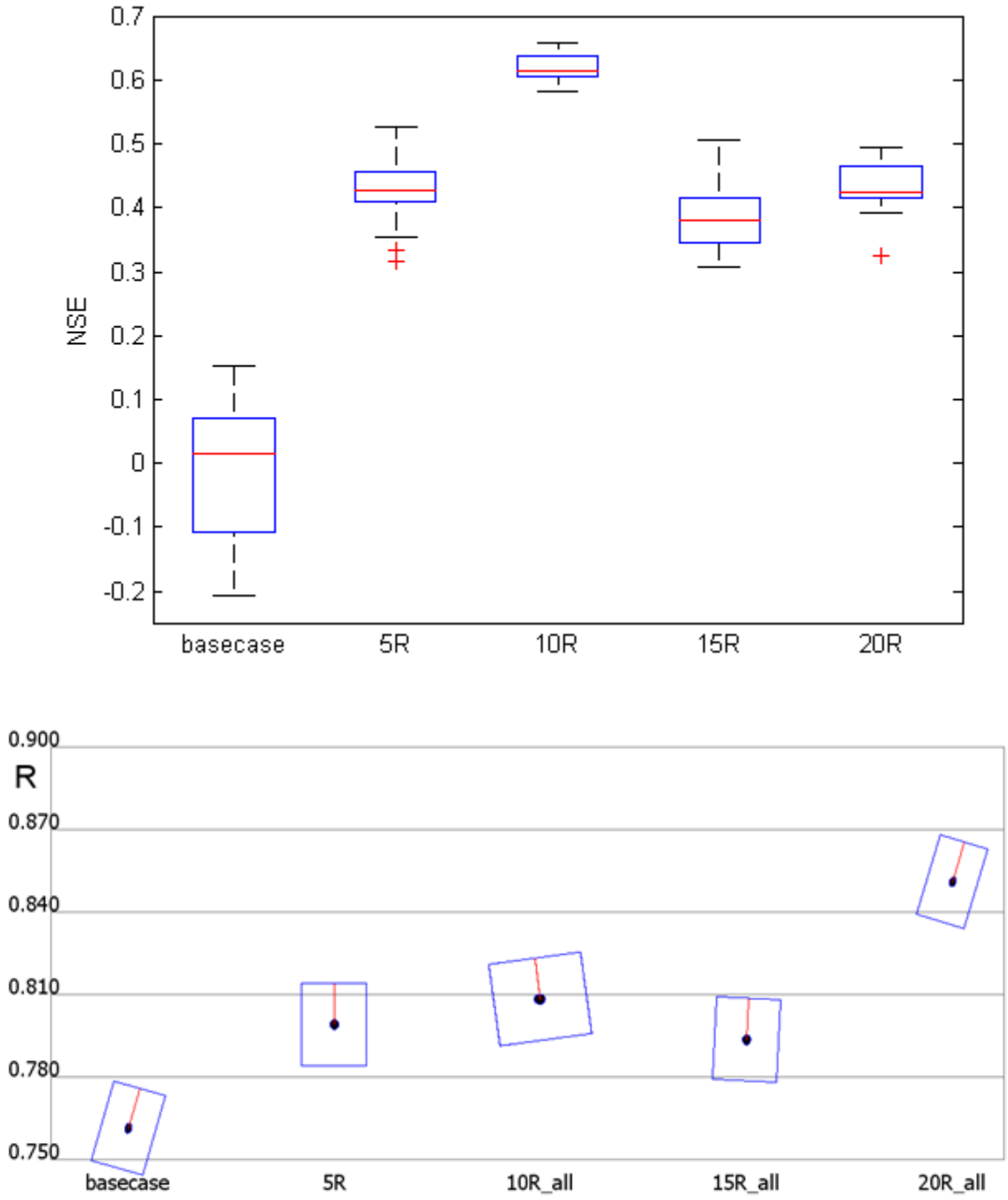
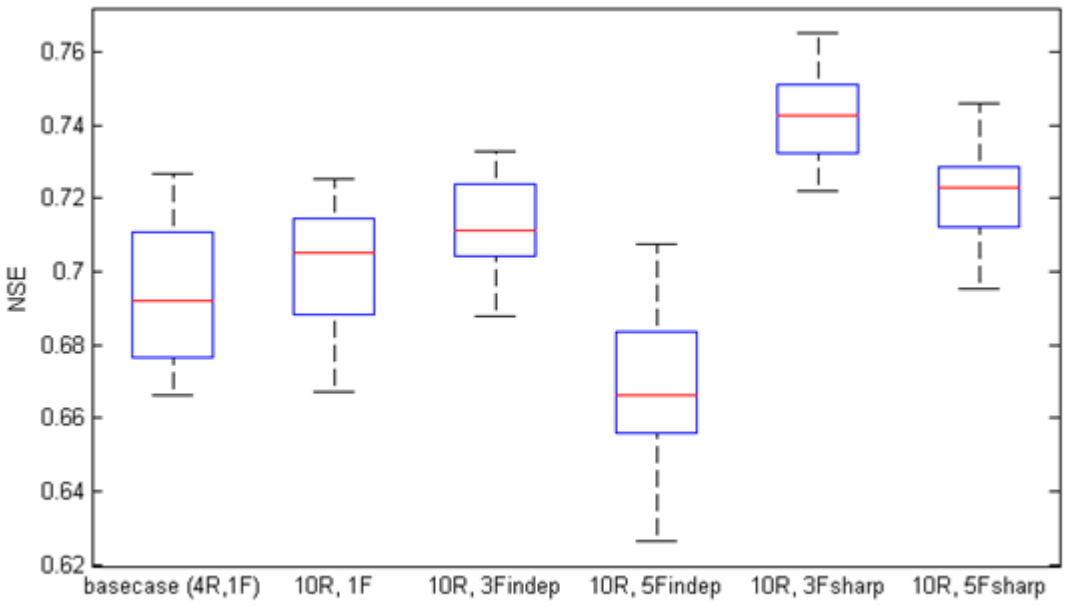


Figure 24: Results for model performance for total dissolved phosphorus simulation for increasing number of rain gauges in the Cville3 land use case. The top figure shows a boxplot showing the distribution of average NSE for total

dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$ averaged over all weather scenarios.



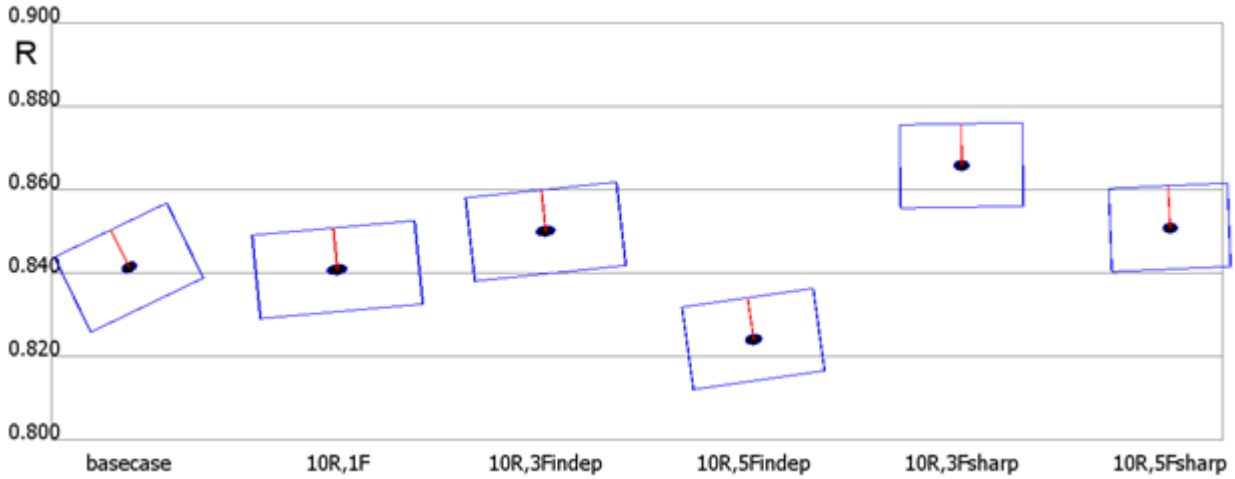


Figure 25: Results for model performance for flow simulation for different sensor network configurations in the Cville3 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$.

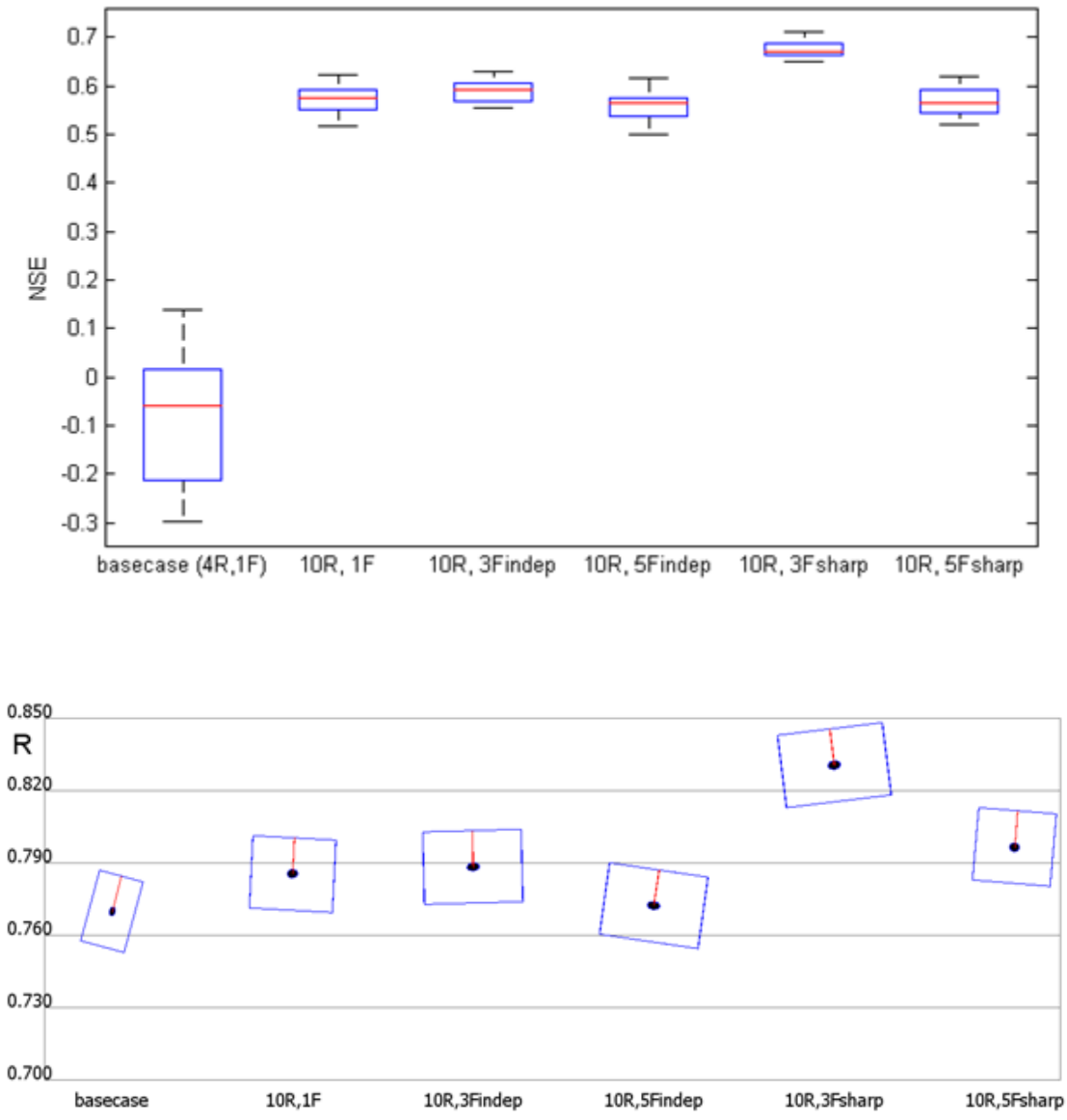


Figure 26: Results for model performance for total dissolved phosphorus simulation for different sensor network configurations in the Cville3 land use case. The top figure shows a boxplot showing the distribution of average NSE for total dissolved phosphorus over 25 different weather scenarios and the bottom figure shows a BVR-rectangle plot showing the $\overline{\alpha}$, $\overline{\beta_n}$, and $\overline{R}$.

REFERENCES

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, doi:10.1061/(ASCE)1084-0699(2005)10:5(363). [online] Available from: http://link.aip.org/link/JHYEFF/v10/i5/p363/s1&Agg=doi

Bradley, A. A., C. Peters-Lidard, B. R. Nelson, J. A. Smith, and C. B. Young (2002), Raingage Network Design Using Nexrad Precipitation Estimates, *Journal of the American Water Resources Association*, *38*(5), 1393–1407, doi:10.1111/j.1752-1688.2002.tb04354.x. [online] Available from: http://doi.wiley.com/10.1111/j.1752-1688.2002.tb04354.x

Bras, R. L., D. G. Tarboton, and C. Puente (1988), Hydrologic Sampling - A Characterization in Terms of Rainfall and Basin Properties, *Journal of Hydrology*, *102*, 113–135.

Bárdossy, A., and T. Das (2008), Influence of rainfall observation network on model calibration and application, *Hydrology and Earth System Sciences*, *12*(1), 77–89, doi:10.5194/hess-12-77-2008. [online] Available from: http://www.hydrol-earth-syst-sci.net/12/77/2008/

Cao, W., W. B. Bowden, T. Davie, and A. Fenemor (2006), Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability, *Hydrological Processes*, *20*(5), 1057–1073, doi:10.1002/hyp.5933. [online] Available from: http://doi.wiley.com/10.1002/hyp.5933 (Accessed 9 March 2012)

Dong, X., C. M. Dohmen-Janssen, and M. Booij (2005), Appropriate Spatial Sampling of Rainfall or Flow Simulation, *Hydrological Sciences Journal*, *50*(2), 279–297, doi:10.1623/hysj.50.2.279.61801. [online] Available from: http://www.informaworld.com/openurl?genre=article&doi=10.1623/hysj.50.2.279.61801&magic=crossref||D404A21C5BB053405B1A640AFFD44AE3 (Accessed 4 January 2011)

Easton, Z., D. Fuka, M. Walter, D. Cowan, E. Schneiderman, and T. Steenhuis (2008), Re-conceptualizing the soil and water assessment tool (SWAT) model to predict runoff from variable source areas, *Journal of Hydrology*, *348*(3-4), 279–291, doi:10.1016/j.jhydrol.2007.10.008. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169407005756

Van Griensven, A., and W. Bauwens (2003), Multiobjective autocalibration for semidistributed water quality models, *Water Resources Research*, *39*(12), 1–9,

doi:10.1029/2003WR002284. [online] Available from: http://www.agu.org/pubs/crossref/2003/2003WR002284.shtml (Accessed 7 September 2010)

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of Hydrology*, *377*(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169409004843

Hirsch, B. R. M., and J. E. Costa (2004), U . S . Stream Flow Measurement, *Current*, *85*(20), 18–20.

Migliaccio, K. W., and I. Chaubey (2007), Comment on Cao W , Bowden BW , Davie T , Fenemor A . 2006 . " Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability ." Hydrological Processes 20 (5): 1057 – 1073, *Hydrological Processes*, *3228*(April), 3226–3228, doi:10.1002/hyp.

Mishra, A. K., and P. Coulibaly (2009), DEVELOPMENTS IN HYDROMETRIC NETWORK DESIGN : A REVIEW, *Reviews of Geophysics*, *47*, 1–24, doi:10.1029/2007RG000243.1.INTRODUCTION.

Nash, J. E., and J. V. Sutcliffe (1970), River Flow Forecasting through Conceptual Models Part I - A discussion of Principles, *Journal of Hydrology*, *10*, 282–290.

Neitsch, S. ., J. G. Arnold, J. R. Kiniry, R. Srinivasan, and J. R. Williams (2004), *Soil and Water Assessment Tool Input/Output File Documentation*.

Neitsch, S. ., J. G. Arnold, J. R. Kiniry, and J. R. Williams (2005), Soil and Water Assessment Tool Theoretical Documentaion, , 1–476.

Reed, P. M. et al. (2006), Bridging river basin scales and processes to assess human-climate impacts and the terrestrial hydrologic system, *Water Resources Research*, *42*(7), W07418, doi:10.1029/2005WR004153. [online] Available from: http://www.agu.org/pubs/crossref/2006/2005WR004153.shtml (Accessed 28 November 2012)

Sait, S. M., and H. Youssef (1999), Tabu Search (TS), in *Iterative Computer Algorithms with Applications in Engineering*, pp. 183–235, IEEE.

Sanders, T. G., R. C. Ward, J. C. Loftis, T. D. Steele, D. D. Adrian, and V. Yevjevich (1987), Location of Water Quality Monitoring Stations, in *Design of Networds for Monitoring Water Quality*, pp. 98–150, Water Resources Publications.

Schaefli, B., and H. V Gupta (2007), Do Nash values have value ?, *Hydrological Processes*, *21*(15), 2075–2080, doi:10.1002/hyp.

Strobl, R. O., P. D. Robillard, R. L. Day, R. D. Shannon, and a J. McDonnell (2006), A water quality monitoring network design methodology for the selection of critical sampling points: part II., *Environmental monitoring and assessment*, *122*(1-3), 319–34, doi:10.1007/s10661-006-0358-4. [online] Available from: http://www.ncbi.nlm.nih.gov/pubmed/16502278 (Accessed 23 February 2012)

Strobl, R. O., P. D. Robillard, and P. Debels (2007), Critical sampling points methodology: case studies of geographically diverse watersheds., *Environmental monitoring and assessment*, *129*(1-3), 115–31, doi:10.1007/s10661-006-9346-y. [online] Available from: http://www.springerlink.com/content/yvpp0429320g13u7

Telci, I. T., K. Nam, J. Guan, and M. M. Aral (2009), Optimal water quality monitoring network design for river systems., *Journal of environmental management*, *90*(10), 2987–98, doi:10.1016/j.jenvman.2009.04.011. [online] Available from: http://www.ncbi.nlm.nih.gov/pubmed/19501953 (Accessed 3 August 2010)

Tolson, B., and C. Shoemaker (2007a), Cannonsville Reservoir Watershed SWAT2000 model development, calibration and validation, *Journal of Hydrology*, *337*(1-2), 68–86, doi:10.1016/j.jhydrol.2007.01.017. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022169407000273

Tolson, B. A., and C. A. Shoemaker (2007b), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res., 43, W01413*, *doi*, 101029/.

Vandenberghe, V., W. Bauwens, and P. a. Vanrolleghem (2007), Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns, *Environmental Modelling & Software*, *22*(5), 725–732, doi:10.1016/j.envsoft.2005.12.019. [online] Available from: http://linkinghub.elsevier.com/retrieve/pii/S1364815206000466 (Accessed 22 May 2012)

White, K. L., and I. Chaubey (2005), Sensitivity Analysis, Calibration and Validations for a Multisite and Multivariable SWAT Model, *Journal Of The American Water Resources Association*, *41*(5), 1077–1089.

Wilks, D. S. (2009), A gridded multisite weather generator and synchronization to observed weather data, *Water Resources Research*, *45*(10), 1–11, doi:10.1029/2009WR007902. [online] Available from: http://www.agu.org/pubs/crossref/2009/2009WR007902.shtml

Woodbury, J., C. A. Shoemaker, D. Cowan, and Z. Easton (2010), A comparison of a SWAT model for the Cannonsville Watershed with and without Variable Source Area Hydrology, , 1–28.

Zhang, X., R. Srinivasan, and M. Van Liew (2010), On the use of multi-algorithm, genetically adaptive multi-objective method for multi-site calibration of the SWAT model, *Hydrological Processes*, *24*(8), 955–969, doi:10.1002/hyp.7528. [online] Available from: http://doi.wiley.com/10.1002/hyp.7528