ABSTRACT

DAVENPORT, JUSTIN MARC. Hierarchical Model Prediction of Flow and Nutrient Load in Ungauged Watersheds: an Application in the Falls Lake Basin, North Carolina. (Under the direction of Daniel Obenour).

Water quantity and quality is an international concern and the resolution of this concern is hindered by the lack of adequate measurements of flow and nutrient concentration data in many of the world's watersheds. An accurate historical dataset is often necessary to model larger-scale water systems such as lakes, riverine systems, and reservoirs. While efforts can be made to increase the number of monitoring stations in watersheds, different modelling techniques can be used to supplement historical datasets through estimation of flows and nutrient loads. Our research is focused on the Neuse River basin upstream of Falls Lake, a reservoir in North Carolina. The Weighted Regressions on Time, Discharge, and Season (WRTDS) approach was used to create a more complete nutrient loading dataset on a weekly time scale for the monitored watersheds. Hierarchical (multi-level) models were then developed to estimate historical watershed flows and nutrient loads in ungauged and partially-gauged tributary watersheds within the Neuse River basin upstream of Falls Lake. The hierarchical models developed also provided realistic estimates of uncertainty for the predictions of flows and loads where 95% of the observations fall within 95% predictive intervals. Land use classifications, flows and loads in a completely gauged watershed (the Eno River), year, and a seasonality term were all considered as candidate predictor variables for inclusion in the hierarchical models. The flow model, containing predictor variables of the completely gauged watershed flow and the area ratio, explained 82% of the variability in weekly

tributary watershed flows during calibration and 81% of the variability during validation. Three nutrient models were developed for organic nitrogen, total nitrogen, and total phosphorous, and had the watershed flow and the completely gauged watershed nutrient load as predictor variables. During calibration, the nutrient models explained 84% (organic nitrogen), 92% (total nitrogen), and 71% (total phosphorous) of the variability in weekly watershed nutrient loads; while during validation they explained 82% (organic nitrogen), 83% (total nitrogen), and 45% (total phosphorous) of the variability. Both the flow and nutrient models are capable of providing a complete history of flows and nutrient loads, with quantification of the uncertainties, for the Falls Lake basin since the completion of the Falls Lake dam in 1982.

Hierarchical Model Predictions of Flows and Nutrient Loads in the Falls Lake Basin from 1982-2014

by
Justin Marc Davenport

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Civil Engineering

Raleigh, North Carolina

2016

APPROVED BY:

_____ _____
Dr. Sankar Arumugam Dr. Tarek Aziz

_____
Dr. Daniel Obenour
Chair of Advisory Committee

BIOGRAPHY

I was born in Albany, Georgia in 1992 to Don and Jayne Davenport. My father attended the University of Georgia and my mother attended North Carolina State University. I have one brother Clinton Davenport, and two sisters Lindsay Davenport and Leah Szatkowski. We moved to Colfax, NC during my elementary school years. During high school I was involved in the Boy Scouts of America and on a FIRST robotics team. I attended North Carolina State University starting in 2010 to pursue a dual degree of a B.S. in Environmental Engineering and B.S in Science, Technology, and Society. I completed my undergraduate career in May 2014 as a valedictorian with a 4.0 GPA. I immediately entered graduate school to pursue my M.S. of Civil Engineering degree, where I find myself now completing this a little over two years later.

ACKNOWLEDGMENTS

I would like to thank North Carolina State University for the opportunity to pursue both undergraduate and graduate degrees. I would like to thank Dr. Daniel Obenour for the advisement, encouragement, and challenges that he provided during the performance of my master's research. I would like to thank doctoral student Jonathan Miller for his comments, questions, and help throughout the research process, and Kevin Li for his comments with regards to both the research and writing. Lastly, I would like to thank my parents, friends, family, and church family for all of the encouragement and support that they have provided me over my entire life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Water quantity and quality issues are prevalent throughout the world and will intensify as the world's population increases. The National Academy of Engineering identified providing access to clean drinking water as a grand challenge of engineering for the first half of this century (Perry et al., 2008). Issues related to water quantity include an increased demand for water from growing populations and industry and changing climate conditions that affect availability of water (Huang & Xia, 2001). However, it is not enough to have an adequate quantity of water, but it is also necessary to ensure that the water is of sufficient quality for its beneficial uses and to be safe for human contact (Fewtrell & Bartram, 2013). One complicating factor in resolving water quantity and quality questions is the lack of flow and pollutant concentration data available in many watersheds. Models can be developed to provide estimates of flows and nutrient concentrations and to understand how water systems are influenced by a variety of factors, and these insights can help engineers create solutions for how adequate water quantity and quality can be provided.

Hydrological flow models are created for the purpose of understanding what watershed factors affect flow in streams and to answer management questions about watershed flow quantity. Watershed hydrological flow modelling is performed using a variety of methods that range in complexity, data input required, and what the model reveals about quantity and quality of watershed outflows (Archfield & Vogel, 2010; Beven, 1989; Karunanithi, Grenney, Whitley, & Bovee, 1994; Reckhow, 1999). Some

models, such as flood prediction models, are designed to only determine extreme flows (Lima & Lall, 2010). Rainfall-runoff models are also used to estimate flows in ungauged watersheds (Makungo, Odiyo, Ndiritu, & Mwaka, 2010; Micovic & Quick, 1999). Other models such as the Soil and Water Assessment Tool (SWAT) and the Storm Water Management Model (SWMM) are more complex and data intensive, utilizing information about the watershed from GIS, precipitation, evaporation, land use, soil, discharge, and withdrawal data to model watershed flows on a daily or sub-daily time scales (Arnold et al., 2012; Jayakrishnan, Srinivasan, Santhi, & Arnold, 2005; Spruill, Workman, & Taraba, 2000; Srinivasan, Ramanarayanan, Arnold, & Bednarz, 1998). These hydrologic models and others are capable of estimating flows from watersheds at varying degrees of scale in order to inform decision makers in regards to partially gauged watersheds.

Watershed nutrient (e.g. nitrogen and phosphorous) models vary greatly in their ability to predict nutrient concentrations and often require a considerable amount of data for calibration, which in many cases are not available, or cover a limited time period (Migliaccio, Chaubey, & Haggard, 2007; Moatar & Meybeck, 2005; Shimoda & Arhonditsis, 2015). Four statistical models for gauged watersheds with limited nutrient data include Load Estimator (LOADEST), Weighted Regressions on Time, Discharge, and Season (WRTDS), Spatially Referenced Regressions on Watershed Attributes (SPARROW), and SWAT. LOADEST uses a multiple linear regression technique with varying degrees of complexity and can output daily pollutant loads and concentrations for periods of record containing a complete flow record and at least a partial nutrient

record (Runkel, Crawford, & Cohn, 2004). WRTDS uses a similar multiple linear regression approach for estimating nutrient loads and concentrations, but it provides a unique regression for each prediction point by weighting observations with covariate values similar to those of the prediction point (Hirsch, Moyer, & Archfield, 2010). SPARROW combines statistical and mechanistic approaches to calculate loads and flows from regional-sized watersheds on monthly or more commonly yearly time scales (Preston, Alexander, Woodside, & Hamilton, 2009). SWAT incorporates land use and other physical data to make predictions on daily time scales in watersheds using a mechanistic formulation with numerous rate parameters that are usually calibrated or estimated from literature (Arnold et al., 2012; Jayakrishnan et al., 2005). While WRTDS and LOADEST work well when expanding the information known about flow-gauged watersheds, they have limited utility for making predictions in ungauged watersheds. SWAT and SPARROW can make predictions in ungauged watersheds, but there is the potential to over parameterize the models, leading to poor predictions (Beven, 2006; Ficklin & Barnhart, 2014; Wellen, Arhonditsis, Labencki, & Boyd, 2014). Uncertainty estimation presents a further challenge for existing flow and nutrient models; frequently, these values are unreported, however the characterization of uncertainty is becoming more widespread (Harmel & Smith, 2007; Reckhow, 1994; Wagener & Montanari, 2011).

In this study, an approach to predict weekly flows and nutrient loads in partially gauged and ungauged watersheds was developed using statistical hierarchical (multi-level) models. Hierarchical models have previously been used in hydrology in

applications such as estimation of stream flow in gauged watersheds from tree rings (Devineni, Lall, Pederson, & Cook, 2013), for flood frequency analysis (Micevski, Kuczera, & Franks, 2004; Yan & Moradkhani, 2015), and through the combination of multiple modelling techniques (Marshall, Nott, & Sharma, 2007). The watersheds were modelled hierarchically allowing for the flow and nutrient models to characterize variability at an “event” level (weekly observation) and at the watershed level. This approach is distinctive when compared to the aforementioned nutrient models in that it allows for the prediction of flows and nutrient loads in ungauged watersheds while accounting for variations between watersheds by allowing model parameters to vary between watersheds. To illustrate this statistical modelling approach, several watersheds that flow into Falls Lake reservoir in central North Carolina (NC) will be used. The Falls Lake basin was chosen as the example due to extensive eutrophication in the receiving reservoir (Stow, Borsuk, & Reckhow, 2011; Stow, Borsuk, & Stanley, 2001), and because it includes many partially gauged and ungauged watersheds, meaning there are significant spatial and temporal gaps in the flow and nutrient load record. The goal of the modeling effort is to create a complete flow and nutrient dataset for all tributaries to the lake, which could then be used to help model water quality issues in the reservoir.

METHODS

Study Area

Falls Lake reservoir serves as a primary drinking water source for the city of Raleigh, NC. The study period begins in October 1982 with the completion of the Falls Lake dam and continues to the end of 2014. At normal pool elevation, the reservoir has a surface area of 50 km^2. The Falls Lake basin (Figure 1) covers an area of 1996 km^2 in Durham, Granville, and Wake counties, NC. The Eno River (ER) watershed, highlighted in tan in Figure 1 is the only continuously gauged watershed in the basin, having a flow and nutrient record absent of significant gaps, and it covers 18.4% of the entire Falls Lake basin. Partially gauged watersheds, in light grey in Figure 1, with incomplete flow and nutrient data, cover 43.5% of the basin; and 38.1% of the basin has negligible gauge information. The Little River Tributary (LT) and Mountain Creek (MC) flow into the Little River Reservoir. The outflow from that reservoir is then measured by the Little River (LR) USGS gauge station.

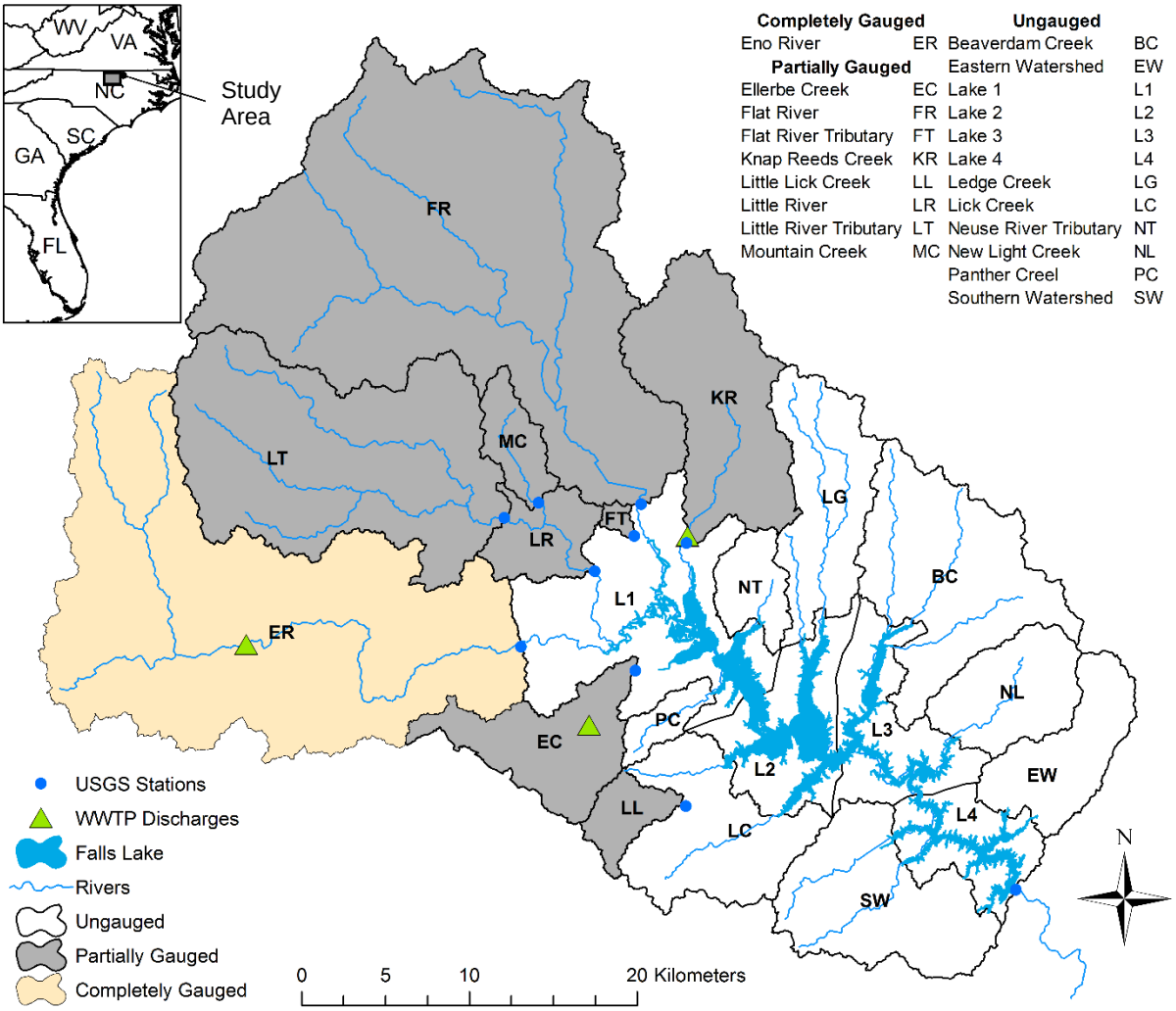


Figure 1. Map of Falls Lake Basin study area.

Flow and Nutrient Data

Flow data come from USGS gauge stations located within the Upper Neuse basin from nine different tributaries (Table 1) (US Geological Survey, 2015). The data were recorded as daily discharges, but were converted to a weekly time scale to mitigate temporal auto-correlation and to reduce some of the noise inherent in daily

flow data. Weeks with missing data were omitted from the calibration dataset. Table 1 shows some summary statistics for the weekly flows in the gauged watersheds.

Table 1. Flow summary statistics for gauged and partially gauged watersheds.

USGS Station	Site	Period of Record (%)	Weekly Flows (cms) Min.	Mean	Max.
02086849	Ellerbe Creek (EC)	61.6	0.1	1.2	9.9
02085070	Eno River (ER)	100.0	0.007	3.4	54.5
02086500	Flat River (FR)	73.4	0.5	3.9	66.2
0208650112	Flat River Tributary (FT)	58.5	0.0003	0.03	1.0
02086624	Knap of Reeds Creek (KR)	69.5	0.04	1.1	14.8
0208700780	Little Lick Creek (LL)	40.3	0.001	0.3	3.3
0208524975	Little River (LR)	60.9	0.01	1.8	59.4
0208521324	Little River Tributary (LT)	85.6	0.0003	1.9	33.5
0208524090	Mountain Creek (MC)	63.3	0.0003	0.2	6.6

Nutrient data from the EPA STORET database and the Upper Neuse River Basin Association (UNRBA) (Upper Neuse River Basin Association, 2015; US Environmental Protection Agency, 2015) were comprised of organic nitrogen (ON), total nitrogen (TN), and total phosphorus (TP); ON is Total Kjeldahl Nitrogen less ammonia nitrogen. Yearly statistics regarding the observed nutrient concentrations are shown in Appendix A. As with flows, weekly loads were desired for the nutrient modelling to reduce some of the noise in daily nutrient data along with reducing the potential for temporal auto-correlation. To obtain weekly loads from the concentration data WRTDS was used to calculate daily concentration values in the gauged watersheds (Hirsch & De Cicco, 2015; Hirsch et al., 2010). The accuracy of the WRTDS model on a daily scale (Figure 2) was evaluated using the Nash-Sutcliffe

Efficiency (NSE) (Equation 4) (Nash & Sutcliffe, 1970). The ON and TN are more accurate than the TP for most rivers. Then the daily WRTDS concentration values were multiplied by the daily USGS flow data to determine daily loads, that were then aggregated into weekly loads. Based on the data input requirements of WRTDS, only a fraction of the period of record was obtainable in the partially gauged watersheds (Table 2).

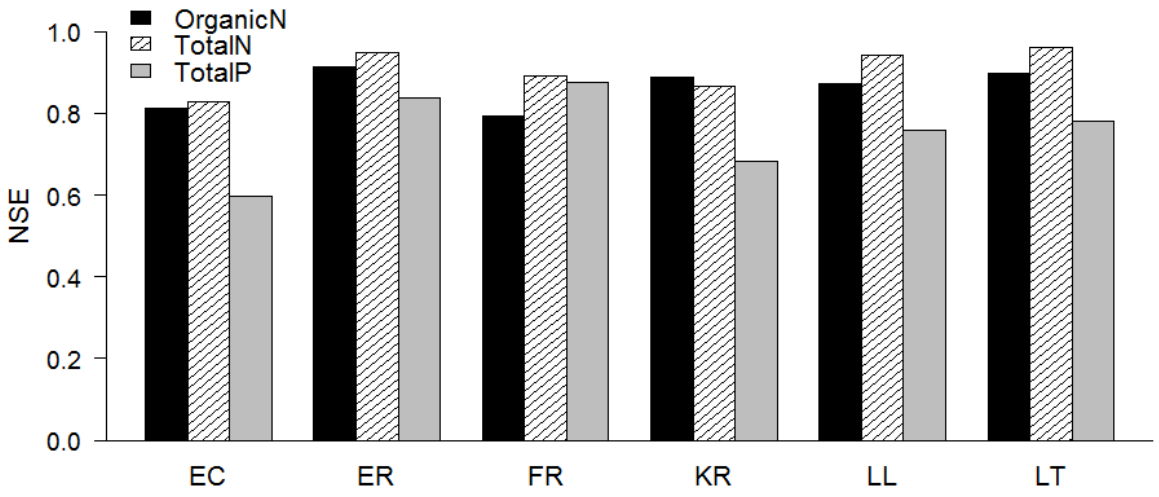


Figure 2. Nash-Sutcliffe Efficiency values for the WRTDS models of daily nutrient loads.

Table 2. Nutrient period of record and weekly load data for gauged and partially gauged watersheds.

USGS Station	Site	Period of Record (%)	Avg. Weekly Loads (kg/d) Organic Nitrogen	Total Nitrogen	Total Phosphorous
02086849	Ellerbe Creek (EC)	48.2	117	156	74
02085070	Eno River (ER)	100.0	129	267	57
02086500	Flat River (FR)	67.3	180	301	25
02086624	Knap of Reeds Creek (KR)	24.7	41	81	14
0208700780	Little Lick Creek (LL)	40.3	25	72	9
0208521324	Little River Tributary (LT)	81.5	99	162	22

Accounting for Withdrawals and Discharges

Five watersheds contained significant anthropogenic discharges from wastewater treatment plants and/or withdrawals for drinking water. Three watersheds contained discharges (Ellerbe Creek (EC), Knap of Reeds Creek (KR), and ER). The average discharge to EC, KR, and ER as a percentage of their average flow was 31%, 9%, and 1%. The average withdrawal from the Flat River (FR), LR, KR, and ER as a percentage of their average flow was 12%, 31%, 9%, and 1% (North Carolina Division of Water Resources, 2015).

For periods of missing discharge and withdrawal (flow or nutrient load) data, missing values where imputed using a watershed specific linear regression. The selected parameters for the flow linear regressions were the same for all watersheds and included the flow in the ER and the year. The coefficients of the parameters for the flow linear regressions are shown in Appendix B. The selected parameters of the linear regressions for nutrients included year, month, month squared, the flow in the

ER, the flow from the wastewater treatment plant, and a plant upgrade variable (Appendix B).

Flow and nutrient exchanges were taken into account before hierarchical modelling by adjusting the observed flow and loading values to reflect non-discharge and non-withdrawal watershed output. Withdrawals were added to the observed watershed flows, while discharges were subtracted from the observed watershed flows. For nutrient loading, discharge loads were removed from the observed watershed loads. To account for the load removed by withdrawals it was assumed that the concentration at the point of withdrawal was the same as the concentration at the gauge station and as a result the loads were scaled to account for the removal of the load. For the ungauged portion of the period of record the watershed nutrient concentration was calculated from the adjusted watershed loads. The watershed nutrient concentration was then multiplied by the withdrawal flow to serve as an estimate of the nutrient load removed from the watershed.

Candidate Predictor Variables

Two types of candidate predictor variables were considered in the hierarchical models: watershed-level variables and event-level variables. Watershed-level variables are variables whose values remain constant for all observations within a given watershed but the values still vary between watersheds. Event-level variables are variables whose values vary both within and between watersheds.

Land use predictor variables were considered in the hierarchical flow and nutrient models to account for how different types of land impact the flow and nutrient loads (Beaulac & Reckhow, 1982; Brabec, Schulte, & Richards, 2002; Line, White, Osmond, Jennings, & Mojonnier, 2002; Tong & Chen, 2002). The land use data came from the National Land Cover Dataset (NLCD) for 2001 which is close to the midpoint of the study period (Multi-Resolution Land Characteristics Consortium, 2001). Land use data were grouped from the original Andersen Level I Classification Scale (Andersen, 1976) into larger subgroups of water, forest, grass, wetlands, urban, bare and agriculture (pasture plus cultivated crops) (Table 3).

Table 3. Gauged watershed land use characterization by percentage of total area.

Site	Total Area (km^2)	Land Use (%)						
		Agriculture	Bare	Forest	Grass	Urban	Water	Wetlands
Ellerbe Creek (EC)	54.9	3.0	0.2	19.6	2.4	74.8	0.2	0.0
Eno River (ER)	366.9	18.0	0.1	60.1	4.0	16.7	0.7	0.4
Flat River (FR)	442.7	27.3	0.1	56.9	6.2	6.1	0.7	2.7
Flat River Tributary (FT)	4.0	0.8	0.0	70.3	2.9	1.0	0.0	25.0
Knap of Reeds (KR)	102.6	14.7	0.4	68.7	6.6	7.2	1.6	0.8
Little Lick Creek (LL)	18.0	4.0	0.1	31.1	2.8	59.1	0.3	2.6
Little River (LR)	252.7	27.5	0.1	58.8	4.6	7.0	0.9	1.1
Little River Tributary (LT)	200.7	28.1	0.1	60.8	4.8	5.9	0.4	0.0
Mountain Creek (MC)	20.8	32.7	0.1	53.5	4.2	9.1	0.4	0.0

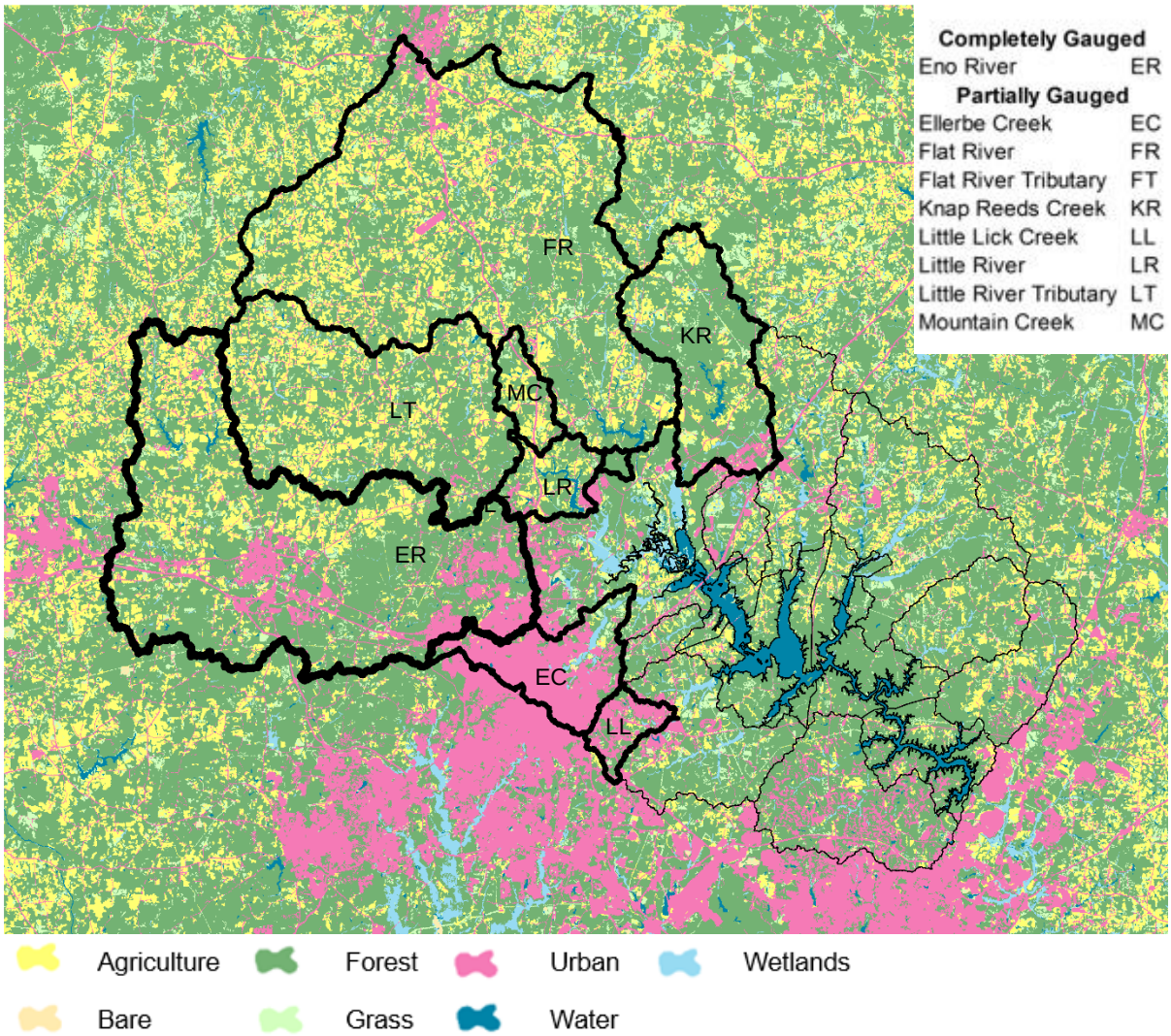


Figure 3. Map of 2001 NLCD subgroups for Falls Lake basin.

Land use area ratios (Equation 1) were calculated for each land use subgroup by dividing the land use subgroup of the watershed by the land use subgroup of the ER watershed. The “R” subscript is an indicator of which watershed the land use area ratio is being calculated. For example, the agriculture area ratio for EC equals the agricultural area of the EC watershed divided by the agricultural area of the ER

watershed (Equation 1). The ratio of the total area of a particular watershed to the total area of the ER watershed was also included as a candidate predictor variable. These land use related variables were candidate predictor variables in the model as watershed-level variables.

$$\text{Land Use Area Ratio} = \frac{\text{Land Use Area}_{\text{R}}}{\text{Land Use Area}_{\text{Eno}}} \qquad \text{Equation 1}$$

The potential event level variables included year, cos(time), and sin(time); where “time” is the month and day as a decimal of the total year multiplied by 2π. The cos(time) and sin(time) are variables that account for seasonality, (Runkel et al., 2004). The year term accounts for long-term trends in flow and nutrient loads. In addition, the flow model included the ER watershed flow (Q_E), and the nutrient models included the ER load (L_E) and the watershed specific flow (Q_R). In addition, a lag-1 version of the ER watershed flow and ER load were included to address potential autocorrelation. The watershed specific flow predictor variable was the observed USGS flows.

Model Formulation

A hierarchical model is essentially a multiple linear regression with intercepts and slopes that vary by group (i.e. watershed). In addition, the varying slopes and intercepts are modelled as members of a common hyperdistribution (Gelman & Hill,

2006) (Figure 4). The advantage of modelling watersheds hierarchically is that the unexplained variability among watersheds can be accounted for along with the variability among individual observations.

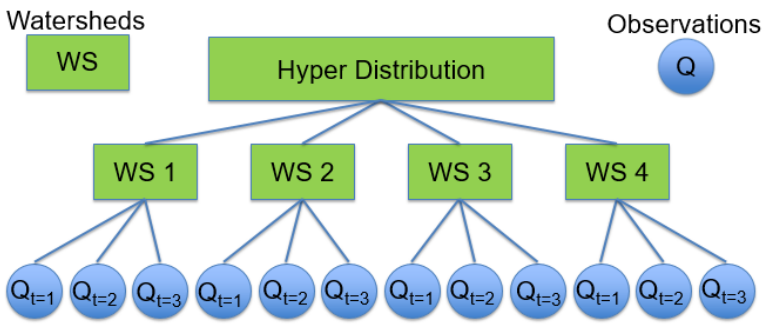


Figure 4. Illustration of a simple hierarchical structure of groups (watersheds) and observations.

The group-level variations in model parameters (slopes and intercepts) are often referred to as random effects, and these model formulations are also frequently known as multi-level models (Steenbergen & Jones, 2002). The LMER function from the lme4 package in R was used to develop the hierarchical models (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2015). The general model equation for a hierarchical model is:

$$y = \alpha_{0,R} + x_{\alpha} * \alpha_{R} + x * \beta + \epsilon \qquad \text{Equation 2}$$

where y is an observation (or prediction), $\alpha_{0,R}$ is a watershed-specific hierarchically modeled intercept value (random effect) that is drawn from a hyperdistribution centered on γ_0 (a fixed effect), x_α is a vector of predictor variables that are subject to watershed-specific hierarchically modeled slope values (α_R, random effects) that are drawn from a hyperdistribution centered on γ_x, x is a vector of additional predictor variables that are subject to only fixed-effect slope parameters (β). The residual value, ϵ, is drawn from a zero-centered normal distribution and represents error in the model at the event level.

Heteroskedacity, reflecting an increase in the variance of the residuals as observed flows and nutrient concentrations increased, was apparent and Box-Cox transformations were used to mitigate this heteroskedacity on the transformed scale (Box & Cox, 1964). Predictor variables were also often transformed to maintain approximate linearity between predictors and responses. The equation for the Box-Cox transformation is as follows:

$$T(y_i) = \frac{(y_i + \lambda_2)^{\lambda_1} - 1}{\lambda_1} \text{ when } \lambda_1 \neq 0$$

$$T(y_i) = \ln(y_i + \lambda_2) \text{ when } \lambda_1 = 0 \qquad \text{Equation 3}$$

The value of λ_1 serves as the power of the transformation; if λ_1 is equal to zero then the variable will undergo a natural log transformation, and if the value of λ_1 is equal to one then there is essentially no transformation (Box & Cox, 1964). The value

of λ_2 serves as a shift parameter. With an appropriate λ_1 value, the variance of the residuals on the transformed scale was approximately constant. The λ_1 and λ_2 values were selected based on a preliminary analysis to reduce heteroscedasticity, while maintaining predictive performance.

To back-transform predictions from the transformed scale to the original scale the following procedure was used. For each prediction, five thousand random draws were taken from the normal distribution centered at the value of the prediction, and using the standard deviation of the residuals on the transformed scale. These draws were then transformed back to the original scale (Equation 4). The best estimate (mean) and predictive interval could then be determined directly from this back-transformed predictive distribution.

$$\mathrm{y_i} = (\mathrm{T(y_i)} * \lambda_1 + 1)^{\frac{1}{\lambda_1}} - \lambda_2 \text{ when } \lambda_1 \neq 0$$
$$\mathrm{y_i} = \mathrm{e}^{\mathrm{T(y_i)}} - \lambda_2 \text{ when } \lambda_1 = 0 \qquad \text{Equation 3}$$

Variable Selection

Before model development, the watershed and event level variables were checked for correlation. Any specific land use variables that was not highly correlated with total area ($|r| < 0.8$) was included as a predictor variable (Table 4). In the flow and nutrient load models, the watershed level variables that were non-correlated were total area ratio and urban ratio. Agricultural area ratio was not included in the nutrient

models since it was highly correlated (r^2=0.98) with the total area ratio even though existing literature on land use and nutrient loadings (Beaulac & Reckhow, 1982; Line et al., 2002) would suggest that it could be a significant factor. All event level variables were non-correlated and included in the models as candidate predictor variables.

Table 4. Pearson's correlation coefficient (r^2) matrix for candidate predictor variables

	Total Area	Agriculture	Bare	Forest	Grass	Urban	Water	Wetlands
Total Area	1	0.98	0.67	0.98	0.98	0.18	0.81	0.83
Agriculture	0.98	1	0.59	0.98	0.96	0.12	0.76	0.72
Bare	0.67	0.59	1	0.69	0.76	0.13	0.83	0.66
Forest	0.98	0.98	0.69	1	0.98	0.12	0.85	0.77
Grass	0.98	0.96	0.76	0.98	1	0.14	0.85	0.76
Urban	0.18	0.12	0.13	0.12	0.14	1	0.10	0.52
Water	0.81	0.76	0.83	0.85	0.85	0.10	1	0.66
Wetlands	0.83	0.72	0.66	0.77	0.76	0.52	0.66	1

Models were formulated with two random effects, the intercept and an additional random effect selected for its importance in predicting watershed flows and nutrient loads: the ER flow for the flow model and the watershed specific flow for the nutrient models. Additional candidate predictor variables were evaluated using a backwards variable selection function (bfFixefLMER_F.fnc) in R at a 99% confidence level (R Core Team, 2015; Tremblay & Ransijn, 2015). If more than two predictor variables remained after selection, the model was re-evaluated for inclusion of interaction terms between the remaining predictor variables based on the same backwards variable selection process at a 99% confidence level. The predictor variables were then pruned by evaluating whether their inclusion increased the portion

of variance explained by the model by more than 2% (change in NSE > 0.02). This stringent method of model selection reflects a preference for parsimonious models that are less subject to over-parameterization.

Model Performance Metrics and Cross-Validation

Summary statistics that were calculated during the calibration and validation phase include the Nash-Sutcliffe efficiency (NSE), average bias, relative bias, and mean absolute error (MAE). The NSE, which is often used to evaluate hydrologic models is determined as follows in Equation 4 where y is the observed discharge, y' is the modelled discharge, and $\bar{\mathrm{y}}$ is the average discharge:

$$\mathrm{NSE} = 1 - \frac{\sum(\mathrm{y} - \mathrm{y}')^2}{\sum(\mathrm{y} - \bar{\mathrm{y}})^2} \qquad \text{Equation 4}$$

The NSE is thus equivalent to how the coefficient of determination (R^2) is commonly defined (Faraway, 2014), but it is unique from the square of Pearson's correlation coefficient (r^2) in that the NSE can take on negative values indicating models that perform worse than just making predictions based on the average value (Jain & Sudheer, 2008; Nash & Sutcliffe, 1970). Bias is the expectation of prediction minus observation, so that if a model has a negative bias then it is under predicting, and if it has a positive bias then it is over predicting. The relative bias scales the bias

to the size of the watersheds by dividing the bias by the average watershed flow or nutrient load. The MAE is the average of the absolute values of the residuals.

Validation was performed by removing one watershed at a time and using the remaining watersheds as the calibration dataset. The excluded watershed was used for validation of the calibrated model and it's NSE, relative bias, and MAE was compared to when the watershed was included in the data. Validating on the excluded watershed evaluates the model's ability to predict flows and nutrient concentrations in a "completely unknown" watershed (Chatfield, 2006; Elsner & Schmertmann, 1994). The random effect for the "completely unknown" watershed was simulated from the distribution of the calibrated random effects, improving estimations of the observations and uncertainties. This method of validation mimics the process that will occur when using the full calibrated model to make flow and concentration predictions for ungauged watersheds.

Watershed Flow and Load Estimation in Spatial and Temporal Gaps

For ungauged watersheds and periods in partially gauged watersheds with missing data, flows and nutrient loads were estimated using the calibrated flow and nutrient models. A Monte Carlo simulation approach is used to fully characterize the predictive distributions of both flows and loads in these spatial and temporal gaps. First, the unknown flows were estimated using the calibrated flow models and the simulate function in the lme4 package in R (R Core Team, 2015; Tremblay & Ransijn, 2015). For simulations in partially gauged watersheds, this function uses the

calculated random effects, but in the ungauged watersheds these random effects are unknown and so they are simulated from the random effect hyperdistribution. In addition, the function includes the residual error in the simulations. The result of simulation is distribution of flow predictions for all watersheds across the entire period of record. For the purpose of examining the output, results were aggregated across multiple watersheds, and then aggregated temporally to examine the results on a yearly basis. Thus, the flow simulation process accounts for the uncertainty from the fixed effects, the random effects, and the event level (residual) variance.

The nutrient load estimation process is the same as the flow estimation process with one change due to the inclusion of watershed flow as a predictor variable. For time periods in partially gauged and ungauged watersheds with no observed flow data, watershed flow values were randomly drawn from the simulated flow distributions described above. These simulated flows were then used in the nutrient load model for the predictor variable values. Thus, the load models reflect the additional uncertainty associated with the stochasticity in the flow estimates.

RESULTS

Flow Model Calibration

The optimized flow model provides insight into how the predictor variables are related to flows in the watersheds. The calibrated model has three fixed effects: ER flow $T(Q_E)$, area ratio T(AR), and the interaction between the ER flow and area ratio $T(Q_E)*T(AR)$ (Equation 5). The "T()" notation indicates a value that is on the Box-Cox transformed scale. The model has two random effects, the intercept and the ER flow, that are drawn from a joint normal distribution (Equation 6). Table 5 displays which parameters were kept, removed due to non-significance, or pruned for not improving the model's explanation of variance by more than 2%.

Table 5. Variable selection results for flow model. NS indicates removal due to not being significant. Prune indicates removal because it did not improve the model's explanation of variance by more than 2%. Keep indicates a parameter was included in the final model.

Variable	
Area Ratio, T(AR)	Keep
Urban Area Ratio, T(UR)	NS
Eno River Flow, $T(Q_E)$	Keep
Lag-1 Eno Flow, $T(Q_{E,-1})$	NS
$T(Q_E)*T(AR)$	Keep
Year	Prune
cos(time)	NS
sin(time)	Prune

To address heteroscedasticity in residuals, the response variable (watershed flow) is subject to a Box-Cox power transformation with λ_1=0.25 and λ_2=0.05, such

that the response unit is $m^{0.75}\,s^{-0.25}$. The predictor variables are subject to the same transformation to maintain approximate linearity between predictor and response. The fixed effect parameters for $T(Q_E)$, $T(AR)$, and $T(Q_E)*T(AR)$ are positive and significant at more than a 99% level (Table 6). No correlations above r^2=0.5 are observed among fixed effects estimates or random effects estimates. The standard deviation of the residuals of the calibrated flow model is 0.53 $m^{0.75}\,s^{-2.25}$. The residuals have an average lag-1 correlation coefficient of 0.15 across all watersheds, and range from 0.10 to 0.22.

$$T(Q_R) = \alpha_{0,R} + T(Q_E) * \alpha_R + T(AR) * \beta_{AR} + T(Q_E) * T(AR) * \beta_{QAR} + \epsilon \qquad \text{Equation 5}$$

$$\begin{pmatrix} \alpha_{0,R} \\ \alpha_R \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_0 \\ \gamma_{Q_E} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_{Q_E} \\ \rho\sigma_0\sigma_{Q_E} & \sigma_{Q_E}^2 \end{pmatrix} \right) \qquad \text{Equation 6}$$

Table 6. Flow model specification during calibration. Units for the estimates and standard errors of the intercept and area ratio parameter estimates are $m^{0.75}s^{-0.25}$. The watershed flow and interaction parameter estimates and standard errors are unit-less.

Fixed Effects	Estimates	Std. Error
Intercept, γ_0	0.04	0.10
Eno River Flow, γ_{Q_E}	0.92	0.06
Area Ratio, β_{AR}	0.79	0.06
Interaction, β_{QAR}	0.26	0.04
Random Effects	**Std. Dev**	**Range**
Intercept, $\alpha_{0,R}$	0.16	-0.24 - 0.20
Eno River Flow, α_R	0.10	-0.17 - 0.13

The flow model calibration is evaluated based on NSE (Figure 5), bias (Figure 6), relative bias (Figure 7) and MAE (Figure 8). The calibrated flow model has a NSE of 0.82, an under predicting bias of 0.10 cms, and a MAE of 0.54 cms, when evaluating the fit of the model to observations across all partially-gauged watersheds. The average flow in the partially gauged watersheds is 1.41 cms, and therefore the relative bias of the calibrated flow model is an under prediction of 7.0%. Plots of predicted versus observed flows for selected watersheds are shown in Figure 9. The calibrated flow model performs well for most watersheds, with all but one of the NSE values greater than 0.6. The greatest NSE is 0.93 in the LT and the lowest NSE is 0.38 in the Flat River Tributary (FT). The relative bias is highest in the FT at 92% and the lowest is Little Lick Creek (LL) with -3%. The MAE is largest in the FR with a value of 1.46 cms, and the smallest is in the FT with 0.04 cms. The FT has relatively low mean flow so even though MAE and bias are lowest in the FT it has the largest relative bias. In addition, the FT has the lowest NSE value, and the MAE of 0.04 cms is greater than the average flow 0.03 cms. The NSE indicates a relatively poor fit of the model which is attributed to the relatively small size of the FT watershed with a total area of 4 km^2. Smaller watershed sizes result in smaller flows representing less flow variability than is present in larger watersheds such as the FR watershed. The larger watersheds have a greater influence on model calibration due to their increased flow variability.

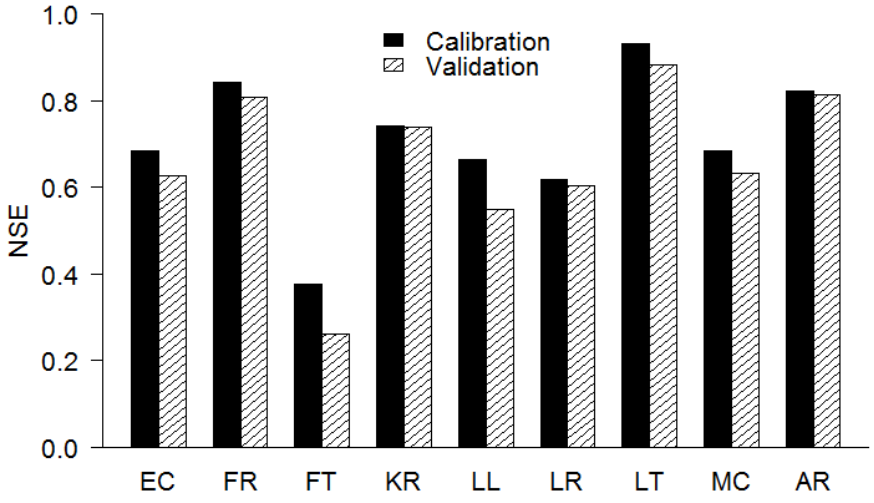


Figure 5. Flow validation and calibration Nash-Sutcliffe Efficiencies (NSE) for the partially gauged watersheds with flow data. AR is the combination of all the watersheds together.

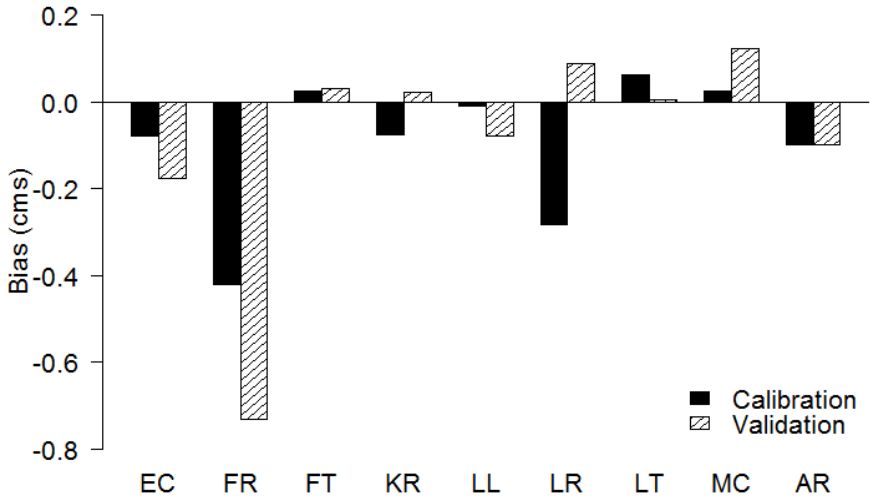


Figure 6. Flow validation and calibration biases (predicted – observed) for the partially gauged watersheds with flow data. AR is the combination of all the watersheds together. Negative bias indicates an under prediction.

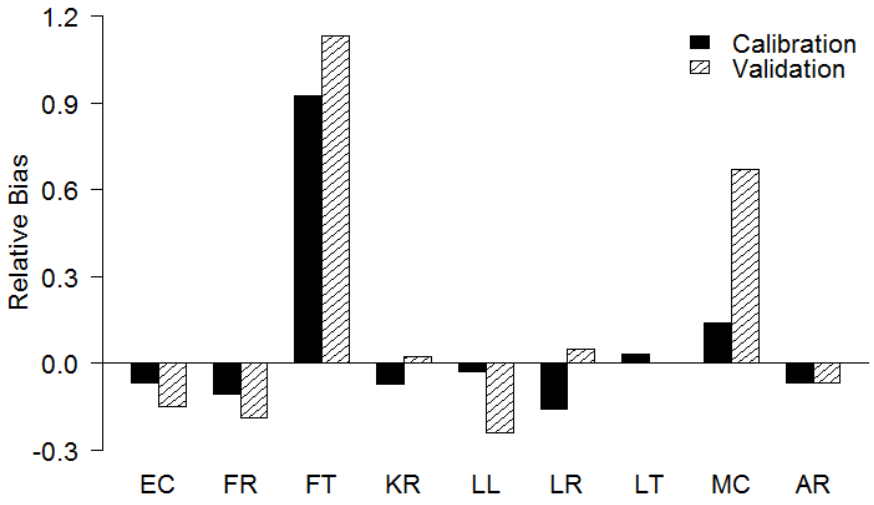


Figure 7. Flow validation and calibration relative biases for the partially gauged watersheds with flow data. AR is the combination of all the watersheds together. Negative relative bias indicates an under prediction.

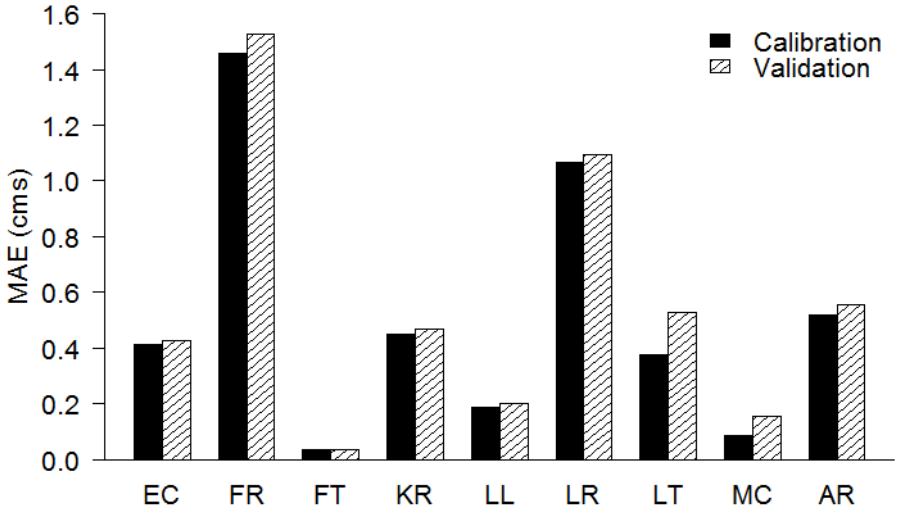


Figure 8. Flow validation and calibration mean absolute errors (MAE) for the partially gauged watersheds with flow data. AR is the combination of all the watersheds together.

Flow Model Validation

The flow model was cross-validated to assess its ability to predict flows in streams outside of the calibration dataset, simulating predictions in ungauged watersheds. The NSE for all watersheds combined is 0.80 during validation, compared to 0.82 during calibration. The overall bias remains an under prediction of 0.1 cms, and accordingly the relative bias also remains the same -7%. The MAE increases from 0.54 cms during calibration to 0.58 cms during validation across all watersheds. The largest decline in NSE occurrs in LL where the NSE decreases from 0.66 to 0.55. Five of the eight watersheds are over predicted while the other three are under predicted. The FT again has the largest relative bias but lowest MAE due to the aforementioned issues with small flow watersheds. Plots of observed and predicted flows during the validation process can be seen in Figure 9 for selected watersheds with plots for the other watersheds shown in Appendix C. The value of parameter estimates during validation (Table 7) fall within the expected range of parameter estimates and standard errors seen in calibration (Table 6).

Table 7. Parameter estimates during model flow validation.

Excluded Watershed	β_0	β_{Q_E}	β_{AR}	β_{QAR}
EC	0.03	0.91	0.79	0.26
FR	-0.01	0.91	0.76	0.26
FT	0.06	0.90	0.82	0.24
KR	0.06	0.92	0.79	0.26
LL	0.04	0.92	0.80	0.27
LR	-0.04	0.98	0.75	0.29
LT	0.12	0.87	0.82	0.24
MC	0.04	0.92	0.76	0.26

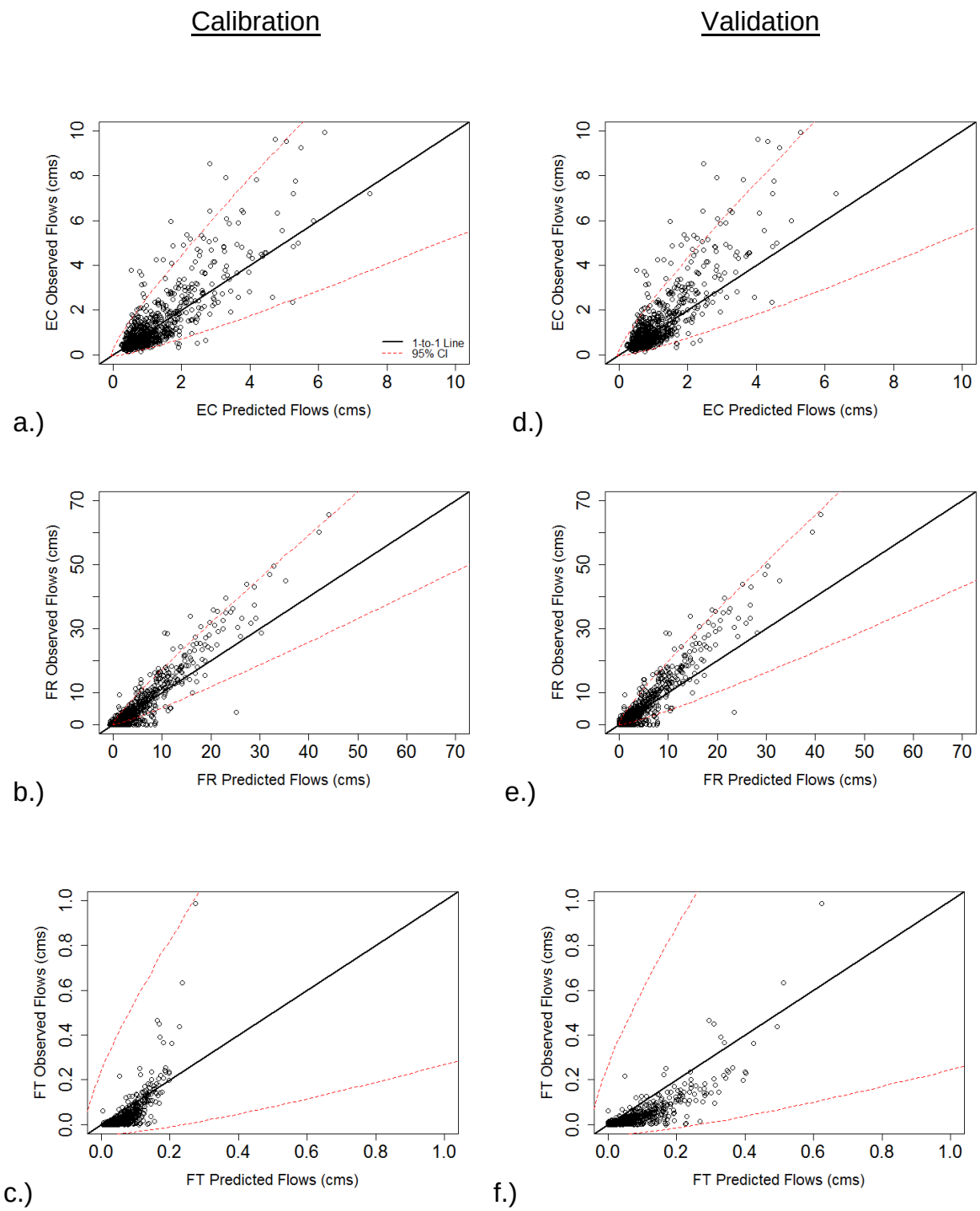


Figure 9. Predicted versus observed plots of flows during calibration (a,b,c) and validation (d,e,f). The selected watersheds are Ellerbe Creek (EC, most urbanized), the Flat River (FR, largest watershed), and Flat River Tributary (FT, smallest watershed).

Nutrient Model Calibration

The nutrient models reveal factors that are related to ON, TN, and TP loads in watersheds. The ON model has three fixed effects, the intercept, ER load $T(L_E)$ and the flow in the watershed $T(Q_R)$. The TN model has two fixed effects, the intercept and $T(Q_R)$. Table 8 displays which parameters were kept, removed due to non-significance, or pruned for not improving the model's explanation of variance by more than 2%.

Table 8. Variable selection results for nutrient load models. NS indicates removal due to not being significant. Prune indicates removal because it did not improve the model's explanation of variance by more than 2%. Keep indicates a parameter was included in the final model.

Variable	ON	TN	TP
Area Ratio, T(AR)	NS	Prune	NS
Urban Area Ratio, T(UR)	NS	Prune	NS
Watershed Flow, $T(Q_R)$	Keep	Keep	Keep
Eno River Load, $T(L_E)$	Keep	NS	Keep
Lag-1 Eno River Load, $T(L_{E-1})$	NS	NS	NS
$T(Q_R)*T(L_E)$	NS	NS	NS
Year	Prune	NS	Prune
cos(time)	Prune	NS	Prune
sin(time)	Prune	Prune	Prune

All parameter estimates for the nutrient models are positive and significant at a 99% confidence level (Table 9). No interaction terms are selected. All of the nutrient models have two random effects: the intercept and the flow in the watershed. The nutrient models are transformed with a Box-Cox transformation of $\lambda_1 = 0.25$ and $\lambda_2 = 0$ to minimize heteroskedacity. The standard deviation of the residuals is 1.09, 1.78,

and 1.98 $kg^{0.25}\,d^{-2.25}$ for ON, TN, and TP, respectively. The average lag-1 correlation of the residuals is 0.16, 0.20, and 0.21 for ON, TN, and TP, respectively.

$$T(L_R) = \alpha_{0,R} + T(Q_R) * \alpha_R + T(L_E) * \beta_{L_E} + \epsilon \qquad \text{Equation 7}$$

$$\begin{pmatrix} \alpha_{0,R} \\ \alpha_R \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_0 \\ \gamma_{Q_R} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_{Q_R} \\ \rho\sigma_0\sigma_{Q_R} & \sigma_{Q_R}^2 \end{pmatrix} \right) \qquad \text{Equation 8}$$

Table 9. Nutrient model specification during calibration. The modelled nutrients were organic nitrogen (ON), total nitrogen (TN), and total phosphorous (TP). Q_R is the watershed flow, L_E is the eno load term. Units of the intercept term are $kg^{0.25}d^{-0.25}$. Units for the watershed flow term are $kg^{0.25}s^{0.25}m^{-0.75}d^{-0.25}$. The Eno load term has no units.

Fixed	ON		TN		TP	
Effects	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
Intercept, γ_0	5.75	0.60	9.02	0.63	1.50	.72
River Flow, γ_{Q_R}	2.29	0.16	2.91	0.31	0.73	.17
Eno Load, β_{L_E}	0.17	0.01			0.42	0.01
Random						
Effects	Std. Dev.	Range	Std. Dev.	Range	Std. Dev.	Range
Intercept, $\alpha_{0,R}$	1.32	-1.17 - 2.26	1.40	-1.55 - 1.31	1.60	-1.17 - 2.72
River Flow, α_R	0.34	-0.56 - 0.37	0.68	-0.75 - 0.86	0.37	-0.32 - 0.41

The nutrient model calibration is evaluated based on NSE (Figure 10), bias (Figure 11), relative bias (Figure 12), and MAE (Figure 13). When evaluated for all watersheds combined the nutrient models has NSE values of 0.84, 0.92, and 0.71; bias values of -4.8, 0.9, and -2.1 kg/day; relative bias of -4.6, 0.5, and -7.2%; and MAE values of 28, 40, and 17 kg/day for ON, TN, and TP, respectively. Negative bias values indicate that the model is under predicting the nutrient loads. The average load in the

partially gauged tributaries is 106, 175, and 30 kg/day for ON, TN, and TP, respectively, such that overall relative biases are less than 8%. Plots of observed versus predicted values for each nutrient for the largest watershed, the FR, are shown as an example of the results (Figure 14) while plots of other watersheds are in Appendix D. The lowest NSE is 0.34 in EC when modelling TP, while the largest NSE is 0.97 in KR when modelling ON. The least biased watershed is ON in the FR with an over prediction of 1% of the average load, and the most biased watershed is TP in EC with an under prediction of -33% of the average load. On average the TN models perform the best, and the TP models were the worst-performing models.

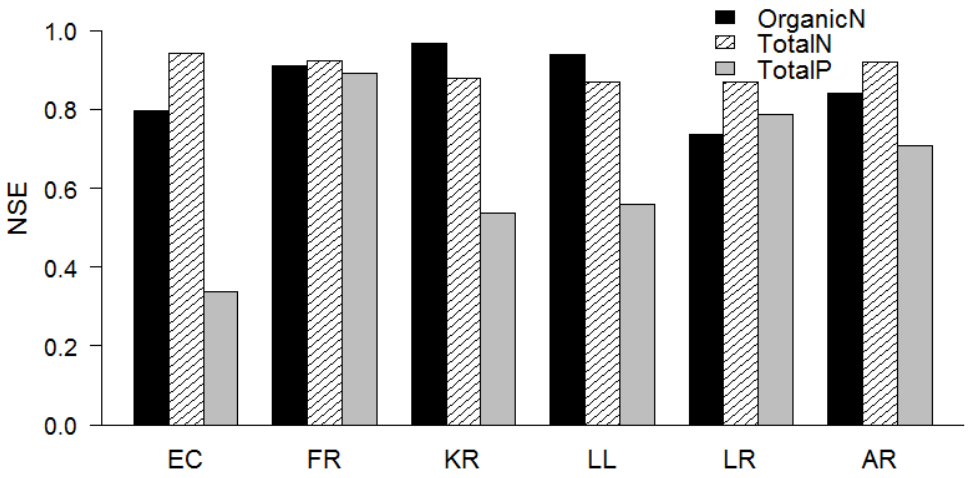


Figure 10. The Nash-Sutcliffe Efficiencies (NSE) for the calibration of the nutrient loading model. AR indicates all of the watersheds combined.

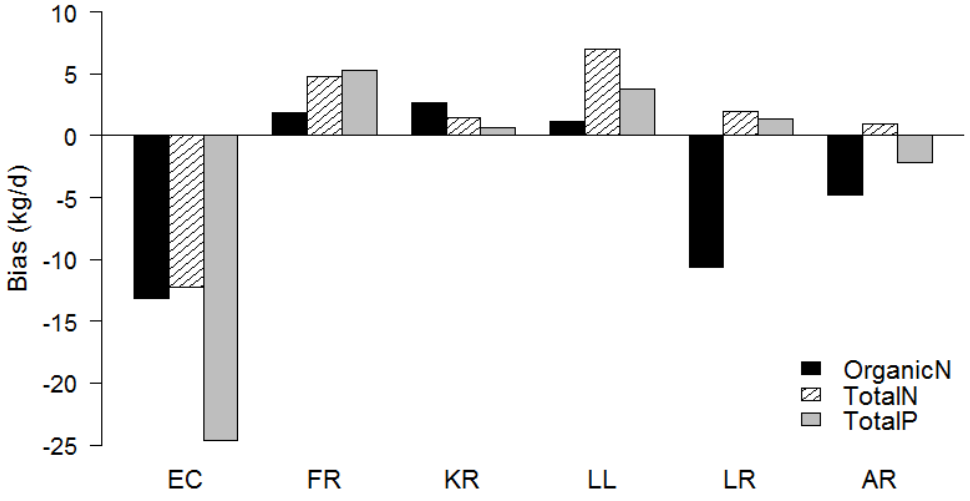


Figure 11. Bias in kg/day during the nutrient loading model calibration. AR indicates all of the watersheds combined. A negative bias indicates an under prediction.

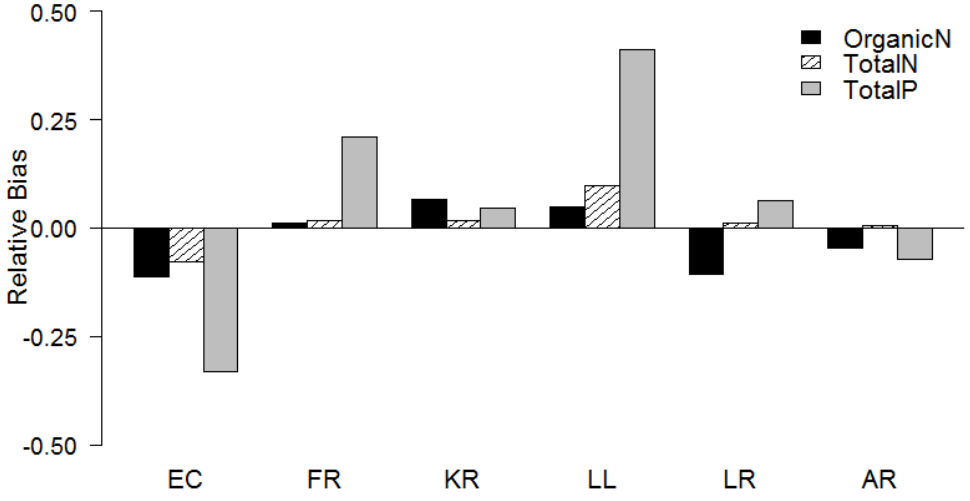


Figure 12. Relative bias during the nutrient loading model calibration. AR indicates all of the watersheds combined. A negative bias indicates an under prediction.

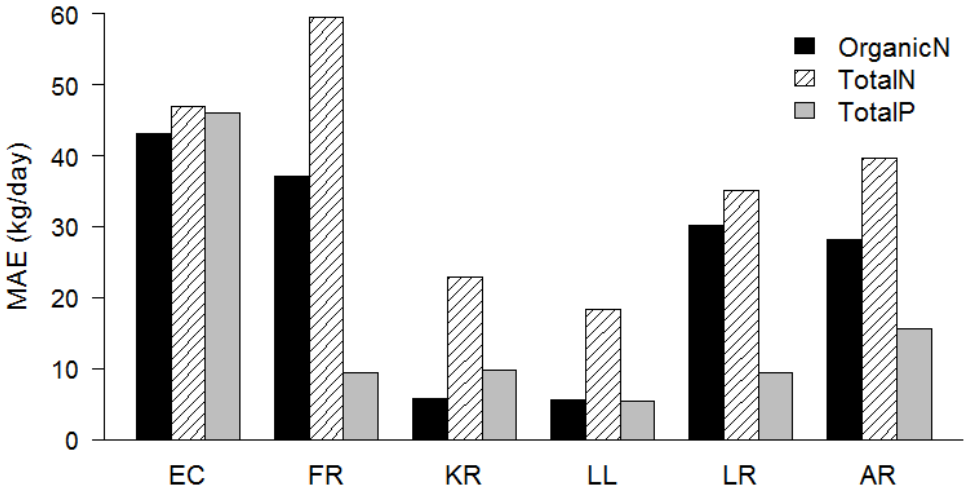


Figure 13. Mean absolute errors (MAE) during the nutrient loading model calibration. AR indicates all of the watersheds combined.

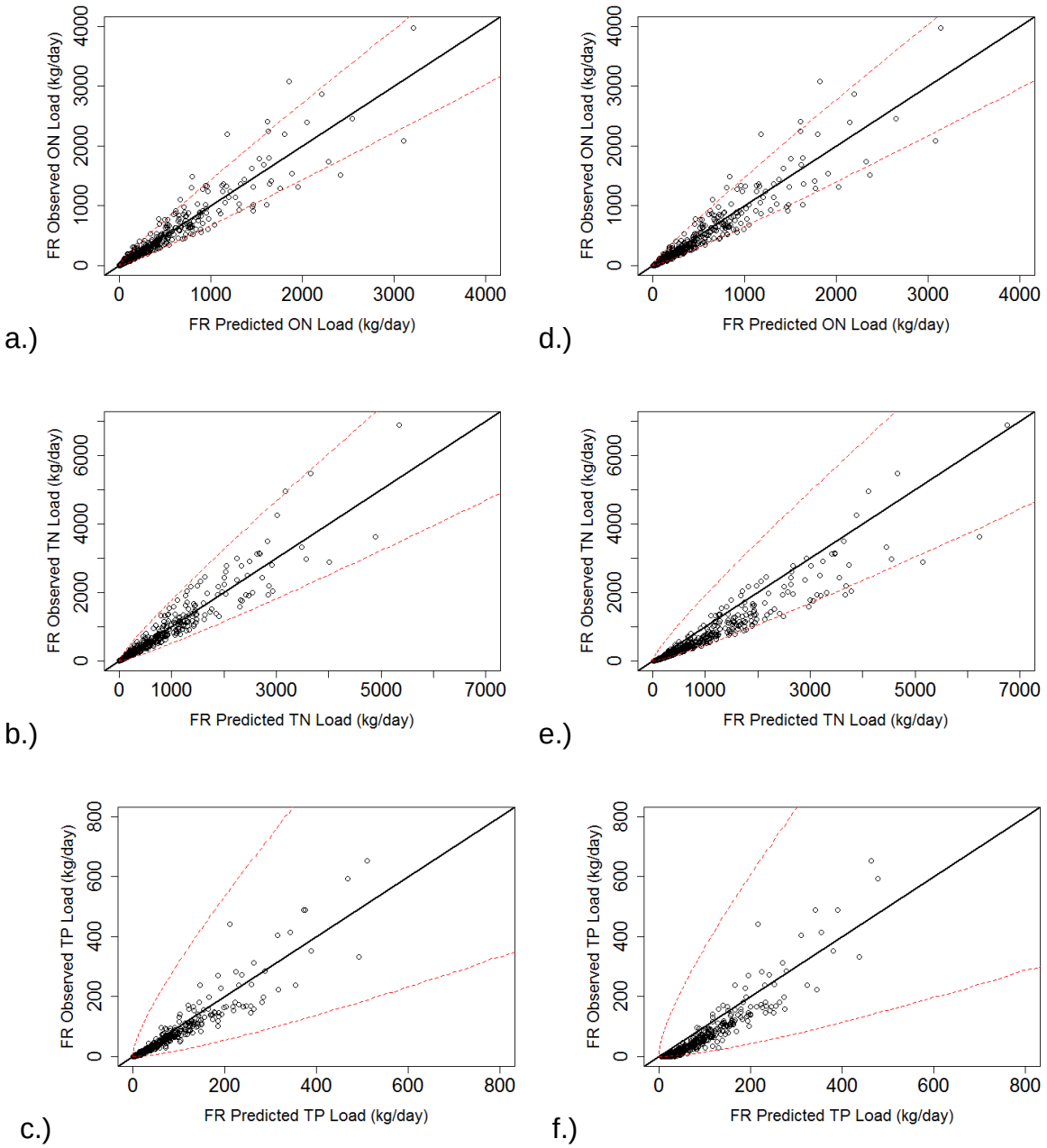


Figure 14. Predicted versus observed plots of nutrient loads during calibration (a,b,c) and validation (d,e,f). The largest watershed, the Flat River (FR), was chosen as an example watershed. The three nutrients modelled are organic nitrogen (ON), total nitrogen (TN), and total phosphorous (TP).

Nutrient Model Validation

The model was cross-validated to assess its ability to predict nutrient loads in watersheds outside of the calibration dataset in the same way as flow models were. Validation plots for FR can be seen in panels D, E, and F of Figure 14 while validation plots for the other watersheds can be seen in Appendix D. When evaluating the fit of the nutrient models to observations across all partially gauged watersheds the NSE (Figure 15) decreases by 0.02 (ON), 0.08 (TN), and 0.26 (TP). Two watersheds, EC and LL, have negative NSE values during TP validation of -0.24 and -0.26 respectively. Both of these watersheds have larger urban areas, with the EC watershed being 75% urban and LL is 59% urban. The effect of greater urban areas increases the variance in the weekly loads when compared to the other watersheds that are less urbanized. The watershed with the greatest NSE is ON in KR with a value of 0.96.

The bias for all watersheds during validation (Figure 17) was 2.7, 39.1, and 0.4 kg/day for ON, TN, and TP. The nutrient load in the EC watershed is under predicted for all three nutrient models, while the nutrient loads in FR, KR, and LT are over predicted in all three models. The ON and TP nutrient loads are over predicted in the LL watershed, while the TN load is under predicted. When the bias values are scaled to the average load, the relative bias (Figure 18) reveals that TP is the most biased at 26%, and ON is the least biased nutrient model at 2%. The MAE (Figure 16) is 45, 85, and 28 kg/day for ON, TN, and TP. When the MAE is compared with the average nutrient load in the watersheds of 106, 175, and 30 kg/day it is clear that combined

with the other performance metrics that the TP model does not perform as well as the ON and TN models. The parameter estimates for the TP model are more varied during validation than the estimates for the ON and TN models (Table 10). This could be a contributing factor to the TP model not performing as well in validation.

Table 10. Parameter estimates of nutrient models used in validation: organic nitrogen (ON), total nitrogen (TN), and total phosphorous (TP).

	ON			TN		TP		
Excluded Watershed	β_0	β_{Q_R}	β_{L_E}	β_0	β_{Q_R}	β_0	β_{Q_R}	β_{L_E}
EC	5.78	2.59	0.11	8.70	2.69	2.02	1.17	0.27
FR	5.37	2.13	0.23	9.36	2.88	1.49	0.84	0.56
KR	5.78	2.19	0.17	8.95	3.09	1.37	0.74	0.44
LL	5.51	2.22	0.20	8.69	3.06	1.56	0.76	0.45
LT	6.37	2.34	0.14	9.41	2.81	1.54	0.57	0.45

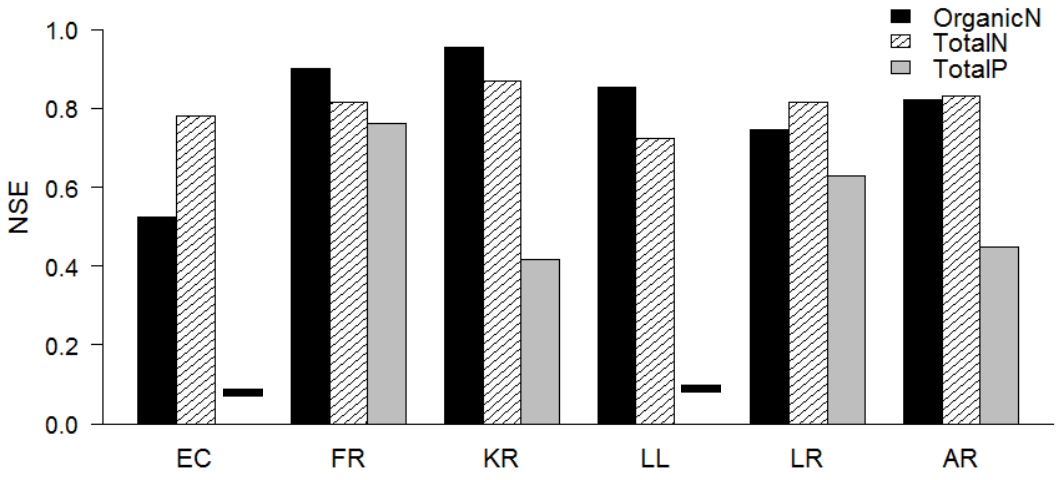


Figure 15. The Nash-Sutcliffe Efficiencies (NSE) during the nutrient loading model validation. AR indicates all of the watersheds combined. Ellerbe Creek and Little Lick Creek had negative NSE values of -0.24 and -0.26.

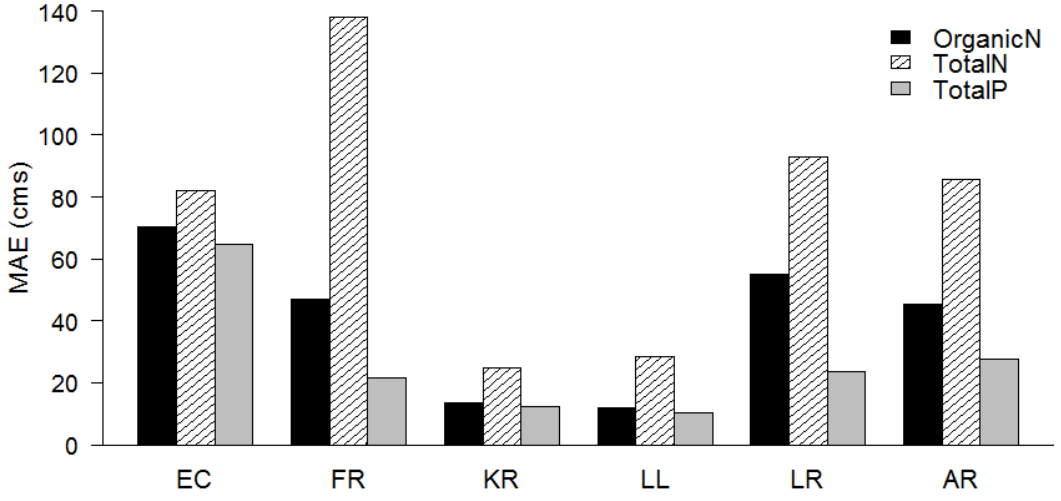


Figure 16. Mean absolute errors during the nutrient loading model validation. AR indicates all of the watersheds combined. A negative relative bias indicates an under prediction.

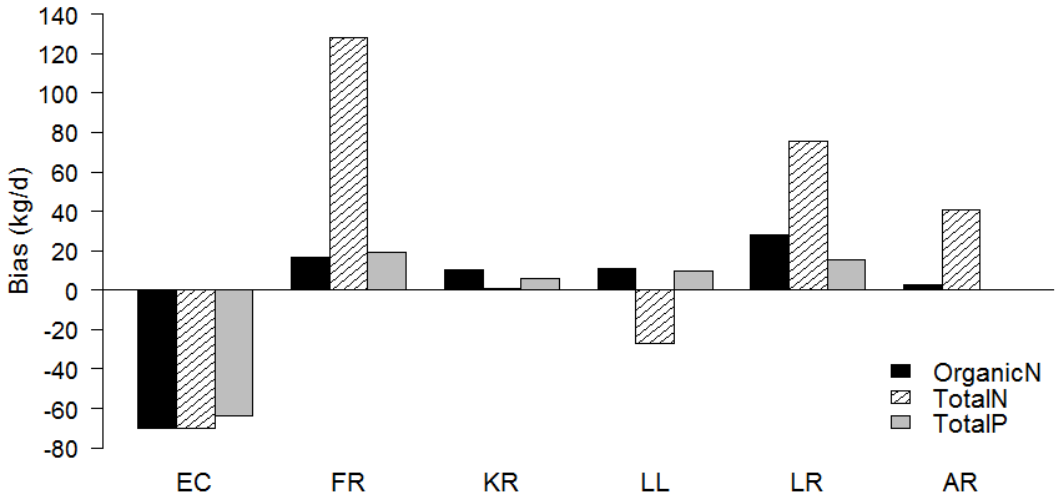


Figure 17. Bias in kg/day during the nutrient loading model validation. AR indicates all of the watersheds combined. A negative bias indicates an under prediction.

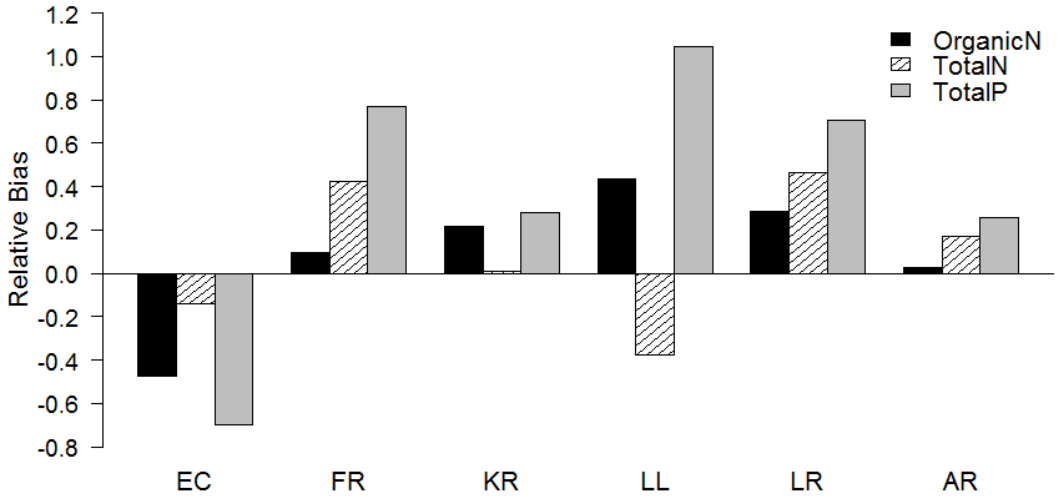


Figure 18. Relative bias during the nutrient loading model validation. AR indicates all of the watersheds combined. A negative relative bias indicates an under prediction.

Watershed Predictions

The ungauged watershed flow estimations are shown in Figure 19 and the nutrient load estimations are shown in Figure 20 as yearly averages. The combination of all watersheds (ungauged, partial, completely) flow estimations are shown in Figure 21 and the nutrient load estimations for all watersheds are shown in Figure 22 as yearly averages. The flow and nutrient load uncertainty estimates account for the variability in the fixed effects, the random effects, and the event level (residual) variance. The nutrient loads uncertainty also includes the uncertainty from the simulated watershed flow as a stochastic predictor variable. Therefore, the uncertainty in the flow estimates is less than the uncertainty in the nutrient load estimates. The high flow estimate in 2003 (Figure 19, 21) seems to relate well with the high flows (7.23 m^3s^{-1}) seen in the ER flow (Appendix E). Similarly, the nutrient loads in 2003 were higher than in other years (Figure 20, 22) as nutrient load is positively related to watershed flow. This same relationship with flows and nutrient loads can also be seen in dry years such as those in 2011 and 2012. The flow estimates seem to indicate the possibility of a slight downwards trend in the flow over the years. If there is a downward trend in flow this would also be expected to affect the nutrient loading unless nutrient concentrations also changed. There does appear to be a negative trend in the nutrient load estimations however, the relationship between flows and loads confounds the ability to make judgments about the condition of the watersheds flowing into the lake and whether they are improving or deteriorating.

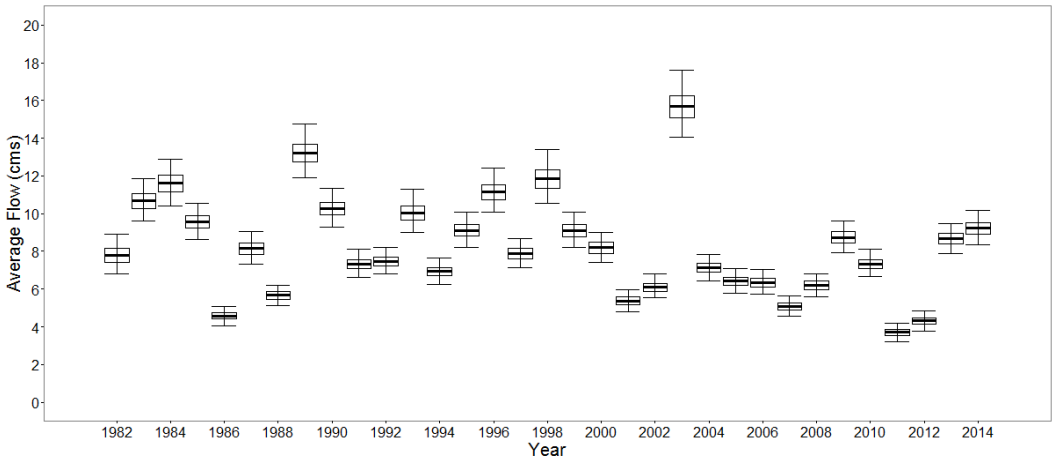


Figure 19. Flows from completely ungauged watersheds. The middle bar represent the average flow, the box represents the 25% and 75% quartiles, and the error bars are the 95% confidence intervals.

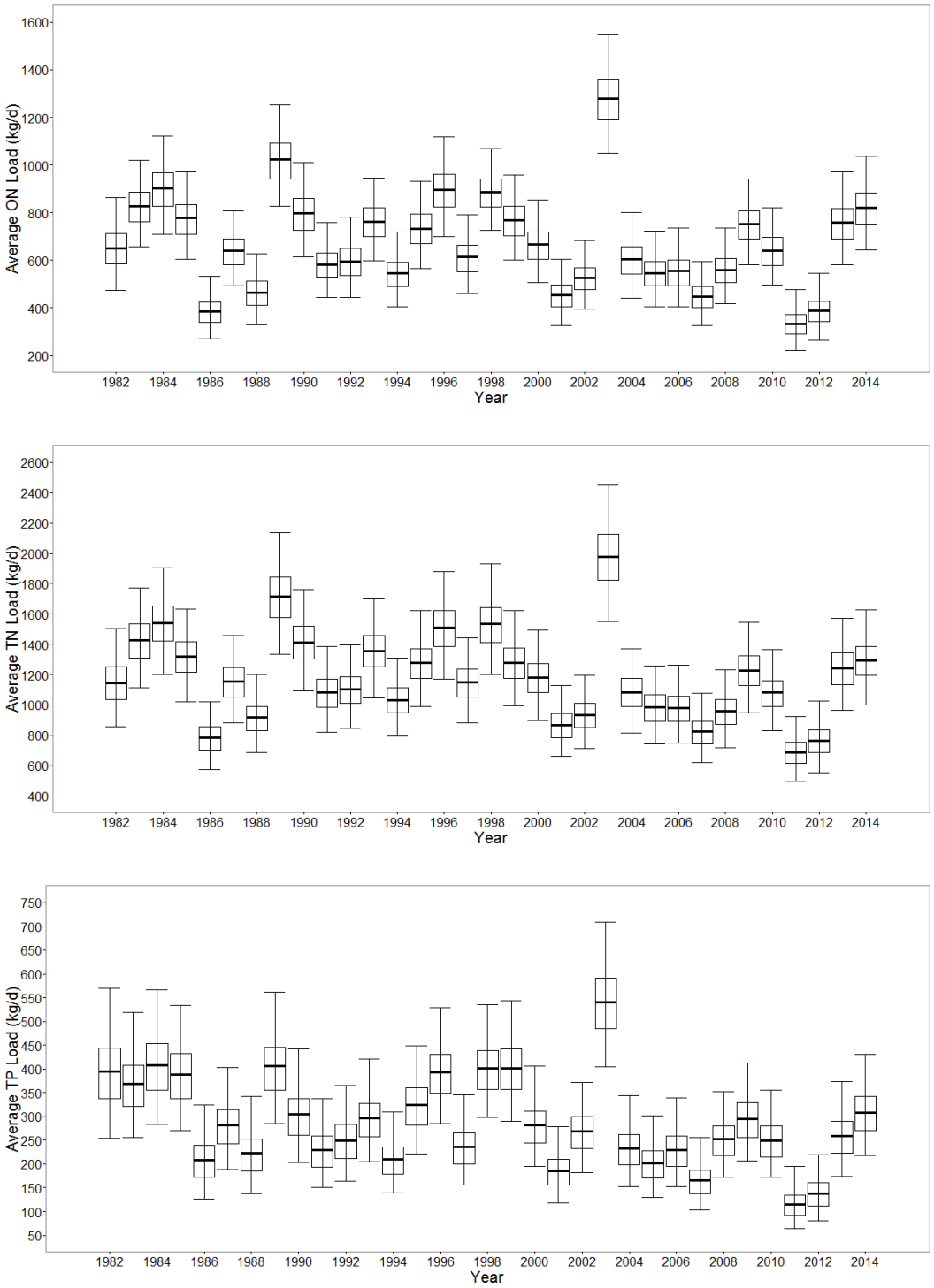


Figure 20. Nutrient loads from completely ungauged watersheds. The middle bar represents the average load, the box represents the 25% and 75% quartiles, and the error bars are the 95% confidence intervals.

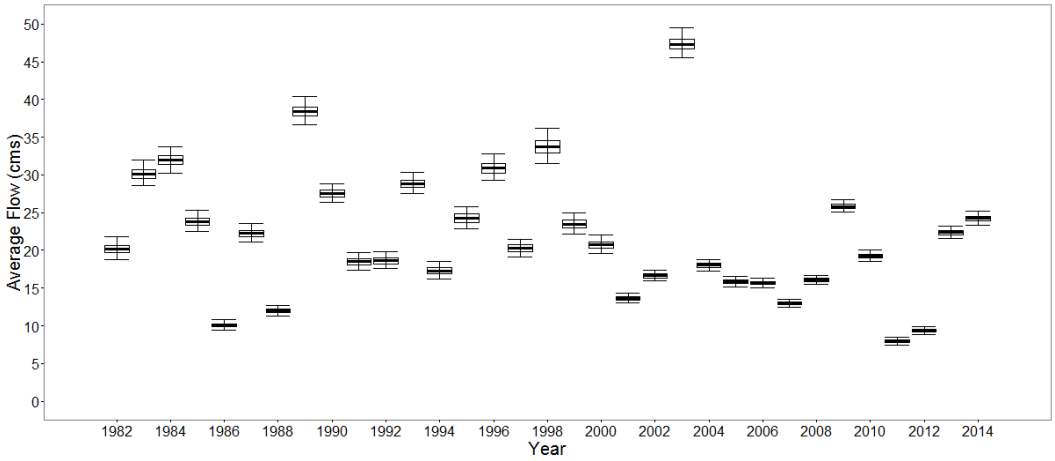


Figure 21. Flows from all watersheds combined (ungauged, partial, and complete). The middle bar represents the average flow, the box represents the 25% and 75% quartiles, and the error bars are the 95% confidence intervals.

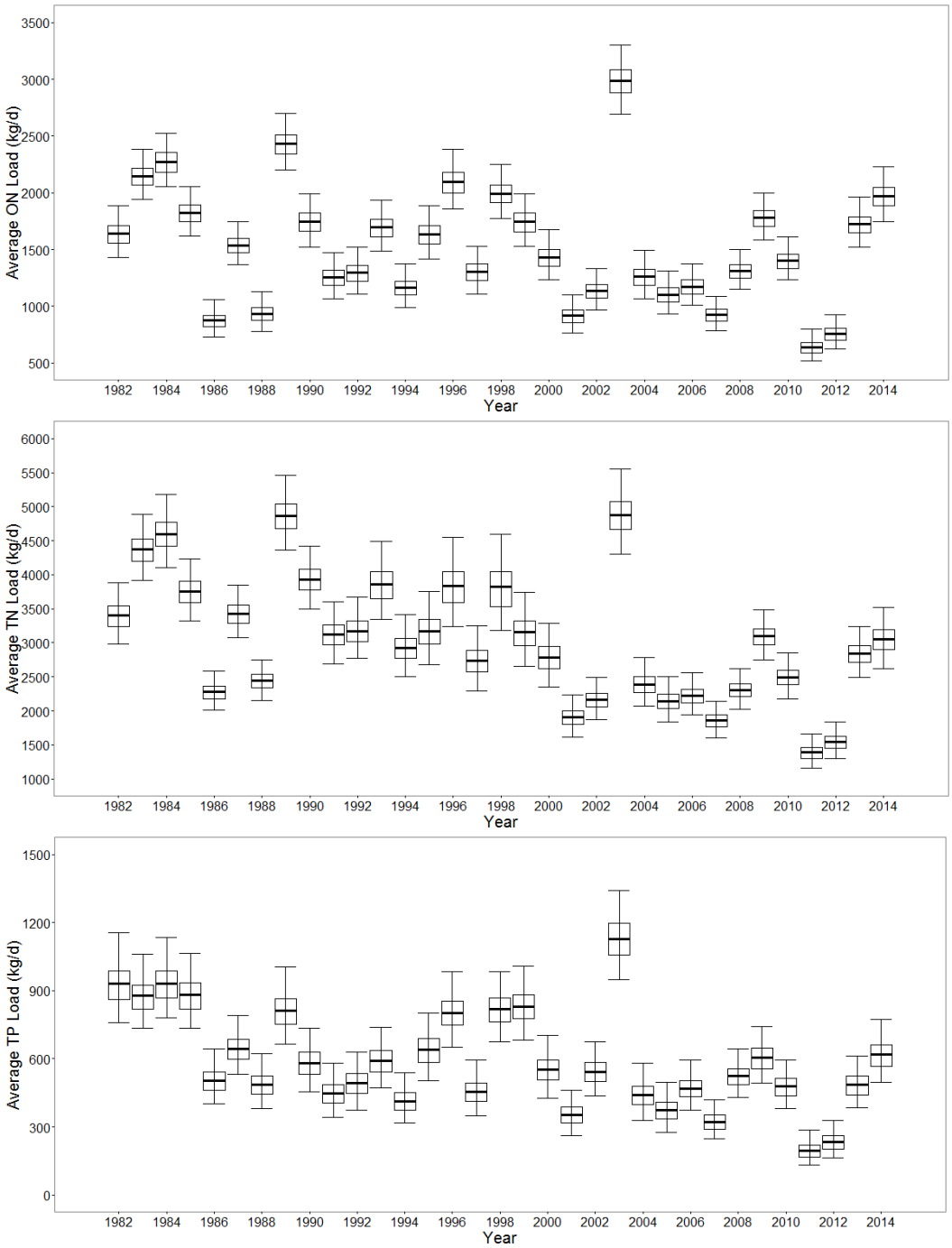


Figure 22. Nutrient loads from all watersheds combined (ungauged, partial, and complete). The middle bar represents the average nutrient load, the box represents the 25% and 75% quartiles, and the error bars are the 95% confidence intervals.

DISCUSSION

Flow Model

The flow model parameterization is harmonious with the existing understanding about influences of watershed flows as it has positive coefficients for both the ER flow and the area ratio. It is expected that the larger a watershed is in comparison to the reference watershed, the ER watershed, the greater amount of flow would be expected and the same holds true for increasingly smaller watersheds (Booth, Hartley, & Jackson, 2002; Brabec et al., 2002). In addition, the inclusion of area ratio as both an individual and interaction term is significant in that it the inclusion of both terms could characterize the differences in base flow and runoff flow between watersheds. However, other predictor variables that might be expected to be included such as urban area ratio were not selected.

The flow model's predictive power comes from including sufficient predictors to characterize flow variability without over parameterizing the model, and the use of a hierarchical structure to account for variations among watersheds that cannot be explained from selected predictor variables. The flow model also includes an accounting of withdrawals and discharges prior to predictor variable selection to characterize the output from the watershed if no withdrawals or discharges occurred. This allows the model to make predictions and include variables that characterize the natural output of the watersheds.

The flow model may appear somewhat similar to the simple area ratio method for estimating flows, which is the idea that the flow in a watershed is directly related to

the flow in another watershed (ER) multiplied by a ratio of the watershed areas. However, the model outperformed the simple area ratio method despite the similarities. In the partially gauged watersheds, the simple area ratio method (after appropriately accounting for discharges and withdrawals, as in the hierarchical model) had a NSE of 0.71, while the hierarchical flow model had a NSE of 0.82 during the calibration phase and a NSE of 0.80 in the validation phase. The flow model should perform as well, if not better, in nearby ungauged watersheds, when compared with the simple area ratio method.

Nutrient Model

The nutrient models predict nutrient loads well (e.g., NSE>0.70 for the majority of watersheds during calibration and validation) with the ON and TN models performing better than the TP model. The nutrient models (ON, TN, and TP) are able to predict nutrient loads in the partially gauged watersheds using the watershed-specific flow (all nutrients) and the ER load (ON and TP) as predictors. The watershed flow parameter is positive and significant for all models as would be expected when predicting watershed loads, since the nutrient load in a watershed is directly related to flow and concentration. As with the flow model, the contributions from wastewater treatment plants and the losses from withdrawals are accounted for prior to hierarchical model development, improving the predictive performance of the model.

Furthermore, the hierarchical nutrient load model performance can be compared to the performance of a simple area-ratio method for estimating nutrient

loads, which directly relates the load in a watershed to the load in another watershed (ER) multiplied by a ratio of the watershed areas. The discharges and withdrawals were accounted for in the simple area ratio method in the same manner as they were in the hierarchical model. For the simple area ratio method, the NSE values for ON, TN, and TP are 0.65, 0.74, and 0.35. Inclusion of the hierarchical structure and the ER load as a predictor variable improves the predictions and the NSE values to 0.84, 0.92, and 0.71 for ON, TN, and TP during calibration. During validation, the NSE values are 0.83, 0.82, and 0.45 for ON, TN, and TP, respectively. Greater improvement is seen in the calibration results than the validation results, but the validation results are still an improvement over the area-ratio method. The hierarchical nutrient models' predictive power comes from using a linear regression approach between loads and watershed flow that is increased by a hierarchical structure and the use of ER load as an additional predictor variable for the ON and TP models.

Uncertainty Estimation

Improvements in the estimation of flow and load from partially gauged watersheds is likely enough of a reason to justify a hierarchical modelling process over a simpler process, and the probable improvement in ungauged watershed estimations is a further benefit of a hierarchical modelling approach. Furthermore, the models are capable of accounting for uncertainty in the estimates of flows and nutrient loads in both partially gauged and ungauged watersheds (Table 11). For all models there is

uncertainty from the variance in the fixed effects and the variance at the event level. The uncertainty estimates in ungauged watersheds also include uncertainty from the random effects. The nutrient load uncertainty for periods of no observed flow in watersheds, includes the uncertainty derived from using watershed flow based on a predictive distribution from the flow model. In addition, for all uncertainty estimations, the hierarchical models can account for the increasing variance in the residuals through the use of the Box-Cox transformation. The calculation of both estimates and uncertainty aids decision makers in understanding the spatial and temporal trends of flows and nutrient loads in these watersheds.

Table 11. Sources of uncertainty for different watershed gauge levels and models. For partially gauged watersheds the watershed flow uncertainty only contributes uncertainty in periods with no observed flow.

Watershed Gauge Level	Flow Model	Nutrient Model
Partially Gauged	Fixed Effects Residual Variance	Fixed Effects Residual Variance Watershed Flow Uncertainty WRTDS Uncertainty
Completely Ungauged	Fixed Effects Random Effects Residual Variance	Fixed Effects Random Effects Residual Variance Watershed Flow Uncertainty

WRTDS

As with most watersheds, the gauged watersheds used in this study had limited nutrient sampling data. As such, it was critical to use a complementary model, such as WRTDS, to estimate daily and weekly loads for these watersheds, and thus provide a calibration dataset for the hierarchical nutrient models. The WRTDS models were

accurate on a daily time scale (NSE > 0.75, Figure 2) for all nutrients in all watersheds with the exception of TP in EC (NSE = 0.60). It would be expected that the weekly loads would be more accurate than the daily loads as some of the variability associated with daily measurements would be reduced. However, there is still uncertainty associated with these estimates of load. This uncertainty from the WRTDS models has not been propagated through the hierarchical nutrient models. Therefore, it would be expected that the uncertainty estimates are somewhat underestimated. The propagation of the WRTDS uncertainty through the hierarchical nutrient models is one potential area of improvement for the hierarchical modelling process.

Hierarchical Modelling Challenges

The hierarchical models developed used a linear regression approach, and therefore the results are expected to be unbiased. However, this is confounded to some degree by the fact that they are unbiased on the transformed scale rather than the original, non-transformed scale. While the possibility of a bias from the transformation process could exist, examining the bias results (Figures 6, 11, and 17), the biases do not seem to be consistently positive or negative in the flow and nutrient load model predictions. Another potential issue faced was the heteroskedacity of the residuals leading to inaccurate characterization of uncertainty. The solution of using the Box-Cox transformation provides a more realistic characterization of the uncertainty in the model predictions by reducing the heteroskedacity in the residual

errors on the transformed scale. When back transformed to the original scale, the uncertainty and the variance of the residuals increases when the predictions increase (Table 12). If no transformation was used, the model could not account for this change in variance of the residuals, and the uncertainty in our estimates could be severely understated at higher values of flow and load. Also, it was assumed that the residual variance was equal in all watersheds. The uncertainty characterization could be made more accurate if the residual variance was not assumed to be equal in all watersheds.

Table 12. Uncertainty estimation accuracy over a range of flows and nutrient loads. Normal refers to the range of flows and loads less than the 90% quantile value. High refers to flows and loads greater than the 90% quantile value.

Calibration Results						
	Normal Flows/Loads			High Flows/Loads		
Model	Within 95% CI	Outside 95% CI	% Within	Within 95% CI	Outside 95% CI	% Within
Flow	7386	385	95	817	48	94
ON	3833	136	97	420	21	95
TN	3728	194	95	428	10	97
TP	3796	173	96	386	25	93
Validation Results						
	Normal Flows/Loads			High Flows/Loads		
Model	Within 95% CI	Outside 95% CI	% Within	Within 95% CI	Outside 95% CI	% Within
Flow	7334	437	94	796	69	94
ON	3759	210	95	406	35	92
TN	3756	166	96	413	23	95
TP	3797	172	96	403	38	91

Some limitations that affect the performance of the hierarchical models are a small data set with regard to the number of available watersheds and a lack of diversity in these watersheds. The watersheds available in this study had heavily correlated

land use values for all land uses except for urban area indicating a lack of diversity in the watershed datasets. When examining nutrient loading it might be expected that agricultural land use would be a significant factor(Beaulac & Reckhow, 1982; Johnes, 1996; Line et al., 2002; Tong & Chen, 2002). However, agriculture and total area are highly correlated (r = 0.99) making it infeasible to discern whether agriculture or total land area is more important in determining nutrient load. The limited data set limits the ability of the model to select significant predictor variables whose inclusion explains greater than 2% of the model's explanation of variability. For example, currently in the data set, there are two watersheds that have major reservoirs, and if more such watersheds had available data it is possible that the model could discern the impact of these impoundments on flows and loads. Also, it was expected that the urban area ratio would be selected as a factor influencing flows (Booth et al., 2002; Brabec et al., 2002) and nutrient loads (Beaulac & Reckhow, 1982; Line et al., 2002), but it was not selected. An increase in the number of watersheds would allow the model to establish stronger relationships between predictor variables and watershed flows. Currently, this variability is being explained through the random effects, but additional predictor variables could improve the predictive capacity of the model by reducing the variability explained by the random effects. This improvement in performance would be expected in ungauged watersheds where random effects are unknown, have to be simulated, and add additional uncertainty to the model estimations.

The inclusion of ER flow and load as predictor variables relates the completely gauged watershed to the partially and ungauged watersheds. No explicit temporal

trend is selected for the flow or nutrient load models (year, sin(dtime), cos(dtime)), but this is expected as temporal variation both on a seasonal and a decadal scale should be captured, to some extent, by the variations in the flows and nutrient loads in the ER watershed (Appendix E). The most prominent pattern is seen by examining the average monthly flows and nutrient loads; where in the early part of the year there are high flows and loads and then during the summer months the flow and loads are lower.

If there were no completely gauged watersheds available, the efficacy of this methodology could be called into question. Yet, the use of additional candidate predictor variables could be included to overcome the lack of a completely gauged watershed, such as precipitation or the output from a mechanistic model such as SWAT. If more than one watershed was completely gauged, then all such watersheds could be included as candidate predictor variables individually or by formulating some combination of the watersheds.

FUTURE WORK

One opportunity to improve the hierarchical model is to expand the data set of available watersheds thus increasing both the total area and diversity of watersheds covered. A data set that covers a larger area could justify inclusion of a third “regional” level to the hierarchical structure. Regions could be separated based on aspects such as geology, types of agriculture, and climate. One example of this could be modelling the Neuse River basin in North Carolina that begins in the Piedmont region (hilly with

clay soils) and ends in at the coast (flat and sandy soils). In addition, if more watersheds with greater diversity were included, additional candidate predictor variables such as precipitation, soil type, agriculture type, and multiple completely gauged watersheds could be included. With an increase in watershed diversity it would be expected that a greater number of predictor variables, such as urban area or agricultural area, would ultimately be selected.

Another opportunity to enhance the model could be changing to a daily time step. The benefit of doing so would be greater temporal resolution in the flow and nutrient outputs which might be desired for downstream modelling efforts. However, two challenges of decreasing the temporal scale is the development of methods to account for the increase in the variability in the data, and the increase in auto-correlation between observations and model residuals. If a reduction to a daily time step was desired the model would likely need to include an auto-regressive component to account for autocorrelation. On a weekly time step, for the hierarchical models developed the lag-1 correlations in the residuals are 0.15, 0.16, 0.20, and 0.21 for flow, ON, TN, and TP, respectively. These values would likely increase if the temporal step was reduced and the inclusion of lagged ER flow and load terms which were not selected as significant during model development could become significant predictor variables. If the autocorrelation is not accounted for, the effective sample size of the data set is decreased and there is an increased chance to over parameterize the model.

CONCLUSION

The performance of flow and nutrient models during calibration (NSE > 0.70 when evaluated for all watersheds combined) demonstrates the ability of this hierarchical model approach to estimate flows and nutrient loads in partially gauged watersheds. The results of the hierarchical model do not show a consistent bias towards under or over prediction. The model accounts for seasonal and decadal variability not through predictor variables such as year, cos(time), and sin(time), but through the temporal variability in the ER flow and loads which serve as predictors in the models. The hierarchical model is beneficial in that it can be used to eliminate temporal gaps in the flow and nutrient record while also providing estimates of uncertainty.

The hierarchical model, in addition to estimating flows and loads in partially gauged watersheds, is able to estimate flows and loads in completely ungauged watersheds. Validation results, which provide an indication of model performance in ungauged watersheds, are more varied with the TP model having the lowest NSE of 0.44; while the flow, ON, and TN models all performed well with NSE greater than 0.80. Lower validation results in the case of the TP model highlight a potential limitation to hierarchical modeling, in that there should be sufficient number of watersheds to develop a model capable of making accurate flow and nutrient load estimates. Models based on few watersheds, such as the five watersheds with nutrient observations available in this study, provides limited ability to identify predictor

variables explaining variability across watersheds. However even with the limited data available, the hierarchical flow and nutrient models are capable of making reasonably accurate predictions with estimates of uncertainty on a weekly time scale.

.

REFERENCES

Anderson, J. R. (1976). A land use and land cover classification system for use with remote sensor data (Vol. 964). US Government Printing Office.

Archfield, S. & Vogel, R. (2010). Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water Resources Research*, *46*(10).

Arnold, J., Moriasi, D., Gassman, P., Abbaspour, K., White, M., Srinivasan, R., … others. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE*, *55*(4), 1491–1508.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Beaulac, M. N. & Reckhow, K. H. (1982). AN EXAMINATION OF LAND USE-NUTRIENT EXPORT RELATIONSHIPS1. *JAWRA Journal of the American Water Resources Association*, *18*(6), 1013–1024.

Beven, K. (1989). Changing ideas in hydrology—the case of physically-based models. *Journal of Hydrology*, *105*(1), 157–172.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1), 18–36.

Booth, D. B., Hartley, D. & Jackson, R. (2002). Forest cover, impervious-surface area, and the mitigation of stormwater impacts. *Journal of the American Water Resources Association*, *38*(3), 835–845.

Box, G. E. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252.

Brabec, E., Schulte, S. & Richards, P. L. (2002). Impervious surfaces and water quality: a review of current literature and its implications for watershed planning. *Journal of Planning Literature*, *16*(4), 499–514.

Chatfield, C. (2006). Model Uncertainty. *Encyclopedia of Environmetrics*. doi:10.1002/9780470057339.vam030

Devineni, N., Lall, U., Pederson, N. & Cook, E. (2013). A tree-ring-based reconstruction of Delaware River basin streamflow using hierarchical Bayesian regression. *Journal of Climate*, *26*(12), 4357–4374.

Elsner, J. B. & Schmertmann, C. (1994). Assessing forecast skill through cross validation. *Weather and Forecasting*, *9*(4), 619–624.

Faraway, J. J. (2014). *Linear Models with R*. CRC Press.

Fewtrell, L. & Bartram, J. (2013). Water quality: guidelines, standards and health: assessment of risk and risk management for water-related infectious disease. *Water Intelligence Online*, *12*, 9781780405889.

Ficklin, D. L. & Barnhart, B. L. (2014). SWAT hydrologic model parameter uncertainty and its implications for hydroclimatic projections in snowmelt-dependent watersheds. *Journal of Hydrology*, *519*, 2081–2090.

Gelman, A. & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Harmel, R. D. & Smith, P. K. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology*, *337*(3), 326–336.

Hirsch, R. M. & De Cicco, L. A. (2015). *User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data*.

Hirsch, R. M., Moyer, D. L. & Archfield, S. A. (2010). Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs1. *JAWRA Journal of the American Water Resources Association*, *46*(5), 857–880.

Huang, G. H. & Xia, J. (2001). Barriers to sustainable water-quality management. *Journal of Environmental Management*, *61*(1), 1–23.

Jain, S. K. & Sudheer, K. (2008). Fitting of hydrologic models: a close look at the Nash-Sutcliffe index. *Journal of Hydrologic Engineering*, *13*(10), 981–986.

Jayakrishnan, R., Srinivasan, R., Santhi, C. & Arnold, J. (2005). Advances in the application of the SWAT model for water resources management. *Hydrological Processes*, *19*(3), 749–762.

Johnes, P. J. (1996). Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: the export coefficient modelling approach. *Journal of Hydrology*, *183*(3), 323–349.

Karunanithi, N., Grenney, W. J., Whitley, D. & Bovee, K. (1994). Neural networks for river flow prediction. *Journal of Computing in Civil Engineering*, *8*(2), 201–220.

Lima, C. H. & Lall, U. (2010). Spatial scaling in a changing climate: A hierarchical bayesian model for non-stationary multi-site annual maximum and monthly streamflow. *Journal of Hydrology*, *383*(3), 307–318.

Line, D. E., White, N. M., Osmond, D. L., Jennings, G. D. & Mojonnier, C. B. (2002). Pollutant export from various land uses in the Upper Neuse River Basin. *Water Environment Research*, 100–108.

Makungo, R., Odiyo, J., Ndiritu, J. & Mwaka, B. (2010). Rainfall-runoff modelling approach for ungauged catchments: A case study of Nzhelele River sub-quaternary catchment. *Physics and Chemistry of the Earth, Parts A/B/C*, *35*(13), 596–607.

Marshall, L., Nott, D. & Sharma, A. (2007). Towards dynamic catchment modelling: a Bayesian hierarchical mixtures of experts framework. *Hydrological Processes*, *21*(7), 847–861.

Micevski, T., Kuczera, G. & Franks, S. W. (2004). A Bayesian Hierarchical Nonhomogeneous Regional Flood Model for New South Wales, Australia. In *Critical Transitions in Water and Environmental Resources Management* (pp. 1–10).

Micovic, Z. & Quick, M. (1999). A rainfall and snowmelt runoff modelling approach to flow estimation at ungauged sites in British Columbia. *Journal of Hydrology*, *226*(1), 101–120.

Migliaccio, K., Chaubey, I. & Haggard, B. (2007). Evaluation of landscape and instream modeling to predict watershed nutrient yields. *Environmental Modelling \& Software*, *22*(7), 987–999.

Moatar, F. & Meybeck, M. (2005). Compared performances of different algorithms for estimating annual nutrient loads discharged by the eutrophic River Loire. *Hydrological Processes*, *19*(2), 429–444.

Multi-Resolution Land Characteristics Consortium. (2001). National Land Cover Database 2001. Retrieved July 1, 2015, from http://www.mrlc.gov/nlcd01_data.php

Nash, J. & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290.

North Carolina Division of Water Resources. (2015). Local Water Supply Plans. Retrieved December 1, 2015, from http://www.ncwater.org/Water_Supply_Planning/Local_Water_Supply_Plan/

Perry, W., Broers, A., El-Baz, F., Harris, W., Healy, B., Hillis, W. D. & others. (2008). Grand challenges for engineering. *National Academy of Engineering, Washington, DC*.

Preston, S. D., Alexander, R. B., Woodside, M. D. & Hamilton, P. A. (2009). *SPARROW Modeling: Enhancing Understanding of the Nation's Water Quality*. US Department of the Interior, US Geological Survey.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austia. Retrieved from https://www.R-project.org/

Reckhow, K. H. (1994). Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. *Ecological Modelling*, *72*(1), 1–20.

Reckhow, K. H. (1999). Water quality prediction and probability network models. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(7), 1150–1158.

Runkel, R. L., Crawford, C. G. & Cohn, T. A. (2004). *Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers*.

Shimoda, Y. & Arhonditsis, G. B. (2015). Integrating hierarchical Bayes with phosphorus loading modelling. *Ecological Informatics*, *29*, 77–91.

Spruill, C., Workman, S. & Taraba, J. (2000). Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. *Transactions of the ASAE*, *43*(6), 1431.

Srinivasan, R., Ramanarayanan, T. S., Arnold, J. G. & Bednarz, S. T. (1998). Large area hydrologic modeling and assessment part II: Model application1.

Steenbergen, M. R. & Jones, B. S. (2002). Modeling multilevel data structures. *American Journal of Political Science*, 218–237.

Stow, C. A., Borsuk, M. E. & Reckhow, K. H. (2011). Nitrogen TMDL development in the Neuse River Watershed: an imperative for adaptive management. *Journal of Contemporary Water Research and Education*, *122*(1), 4.

Stow, C. A., Borsuk, M. E. & Stanley, D. W. (2001). Long-term changes in watershed nutrient inputs and riverine exports in the Neuse River, North Carolina. *Water Research*, *35*(6), 1489–1499.

Tong, S. T. & Chen, W. (2002). Modeling the relationship between land use and surface water quality. *Journal of Environmental Management*, *66*(4), 377–393.

Tremblay, A. & Ransijn, J. (2015). LMERConvenienceFunctions: Model Selection and Post-hoc Analysis for (G)LMER Models. Retrieved from https://CRAN.R-project.org/package=LMERConvenienceFunctions

Upper Neuse River Basin Association. (2015). UNRBA Flow and Water Quality Database. Retrieved February 9, 2015, from https://www.unrba.org/

US Environmental Protection Agency. (2015). STOrage and RETrieval and Water Quality eXchange. Retrieved July 1, 2015, from https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange

US Geological Survey. (2015). USGS Surface-Water Data for the Nation. Retrieved August 1, 2015, from http://waterdata.usgs.gov/nwis/sw

Wagener, T. & Montanari, A. (2011). Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resources Research*, *47*(6).

Wellen, C., Arhonditsis, G. B., Labencki, T. & Boyd, D. (2014). Application of the SPARROW model in watersheds with limited information: a Bayesian assessment of the model uncertainty and the value of additional monitoring. *Hydrological Processes*, *28*(3), 1260–1283.

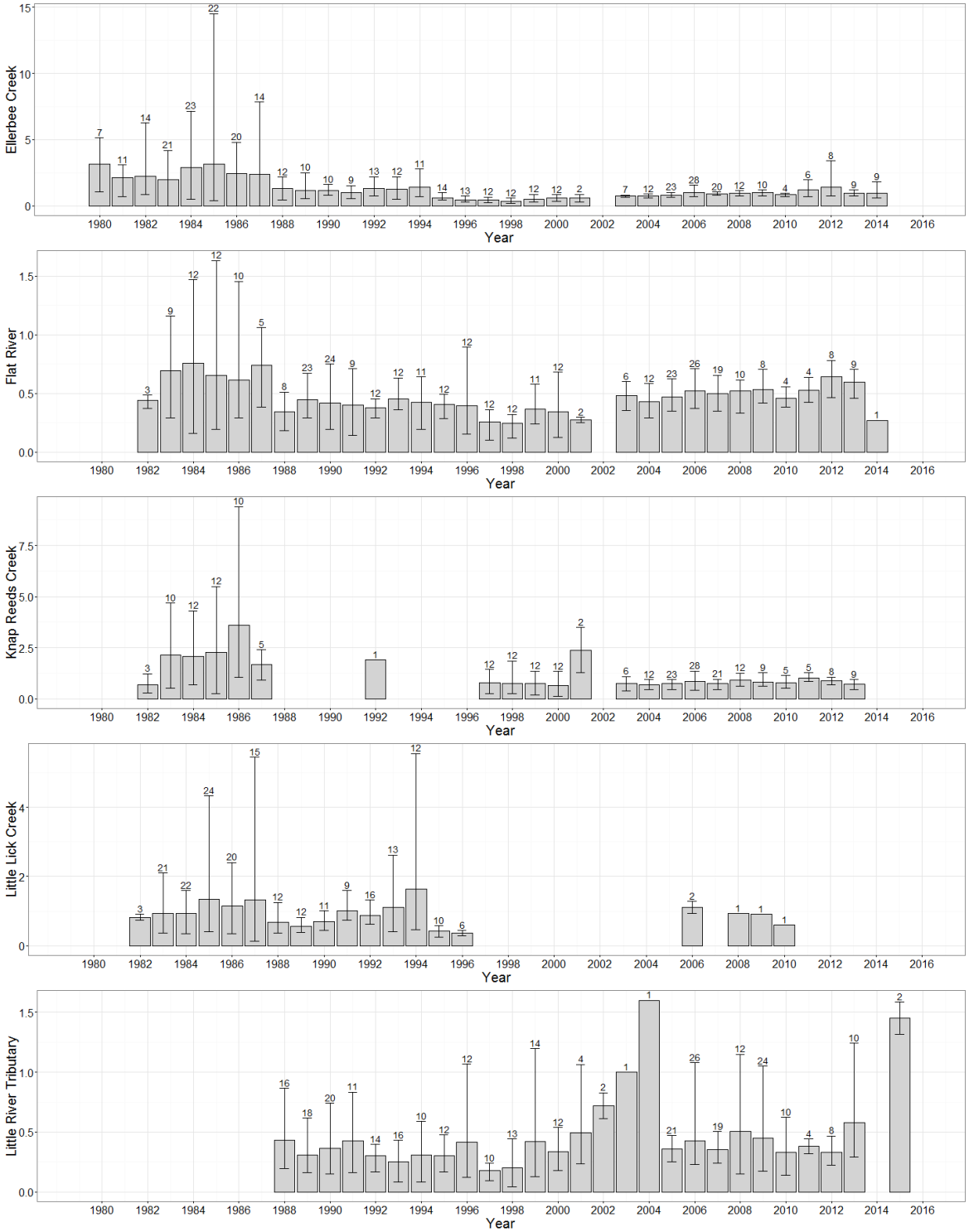
Yan, H. & Moradkhani, H. (2015). A regional Bayesian hierarchical model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, *29*(3), 1019–1036.

APPENDICES

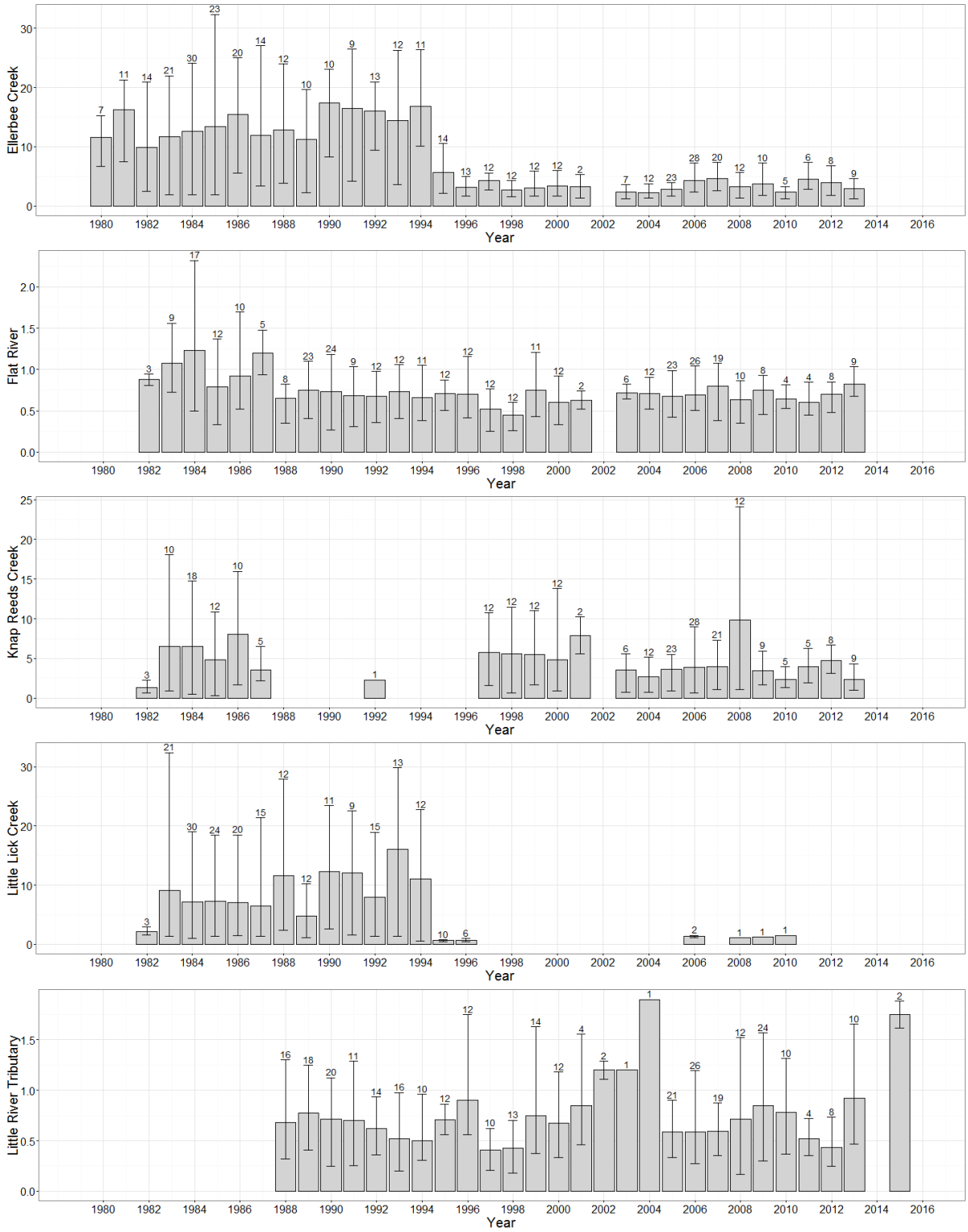
Appendix A. *Observed Watershed Nutrient Concentrations*

The bar height is the average of the nutrient concentrations in that year. The error bars indicate the 5% and 95% quantiles of the observed nutrient concentrations. The number on top of the 95% error bar is how many nutrient concentration observations were recorded in the given year.

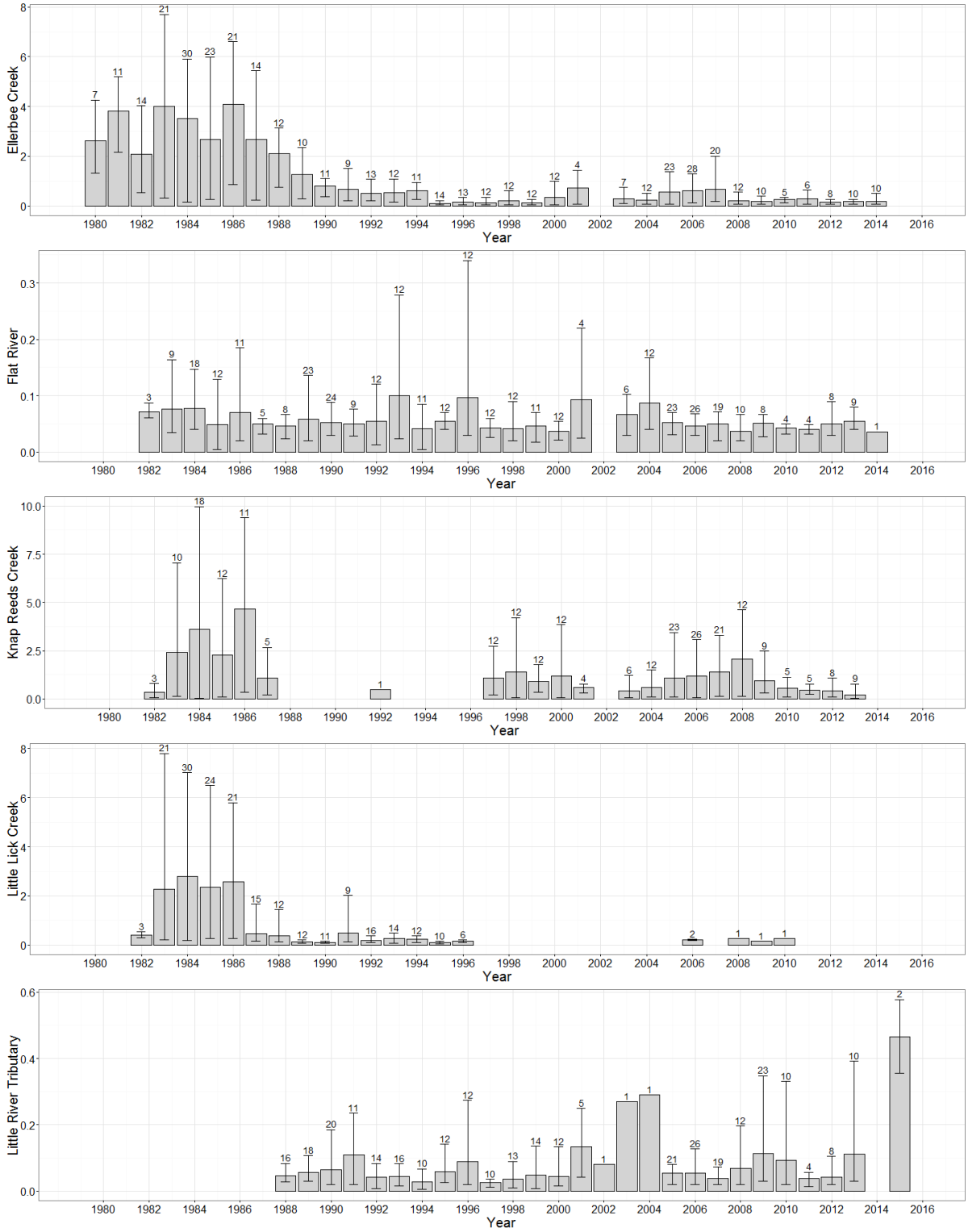
Organic Nitrogen Observed Watershed Concentrations (mg as N/L)



Total Nitrogen Observed Watershed Concentrations (mg as N/L)



Total Phosphorous Observed Watershed Concentrations (mg as P/L)



Appendix B. *Discharge and Withdrawal Model Parameter Estimates*

Estimates of model parameters for the flow linear regressions to impute missing discharge and withdrawal data are in the table below. The model parameters were the ER flow and year. The equation for the discharge and withdrawal flow is:

$$y = \beta_0 + Q_E * \beta_{Q_E} + Y * \beta_Y + \epsilon \qquad \text{Equation 10}$$

		Intercept		Eno River Flow, Q_E		Year, Y	
Site	Type	β_0	Std. Error	β_{Q_E}	Std. Error	β_Y	Std. Error
EC	Discharge	0.294	0.011	0.014	0.001	0.0017	0.0004
ER	Discharge	0.049	0.002	0.002	0.000	-0.0006	0.0001
ER	Withdrawal	0.087	0.006	-0.0006	0.0002	-0.0012	0.0002
FR	Withdrawal	0.022	0.150	-0.038	0.010	0.025	0.006
KR	Discharge	0.077	0.003	0.0032	0.0002	0.0015	0.0001
KR	Withdrawal	0.056	0.009	-0.0020	0.0006	0.0026	0.0003
LR	Withdrawal	1.601	0.153	-0.022	0.010	-0.039	0.006

Estimates of model parameters for the nutrient load linear regressions to impute missing discharge data. The model parameters were year, month, month squared, the flow from the discharge source, the ER flow, and a plant upgrade variable. Month was a number ranging from 1 in January to 12 in December. Month squared is the square of the month number so it would range from 1 in January to 144 in December. The flow from the discharge source if not observed was used from the above flow linear regressions. A plant upgrade variable was included in the models for EC to indicate where the plant was upgraded to remove nutrients at greater levels in early 1994. The general form of the equation is seen in Equation 11 where for each site and nutrient only the terms in tables on the next page are included as appropriate.

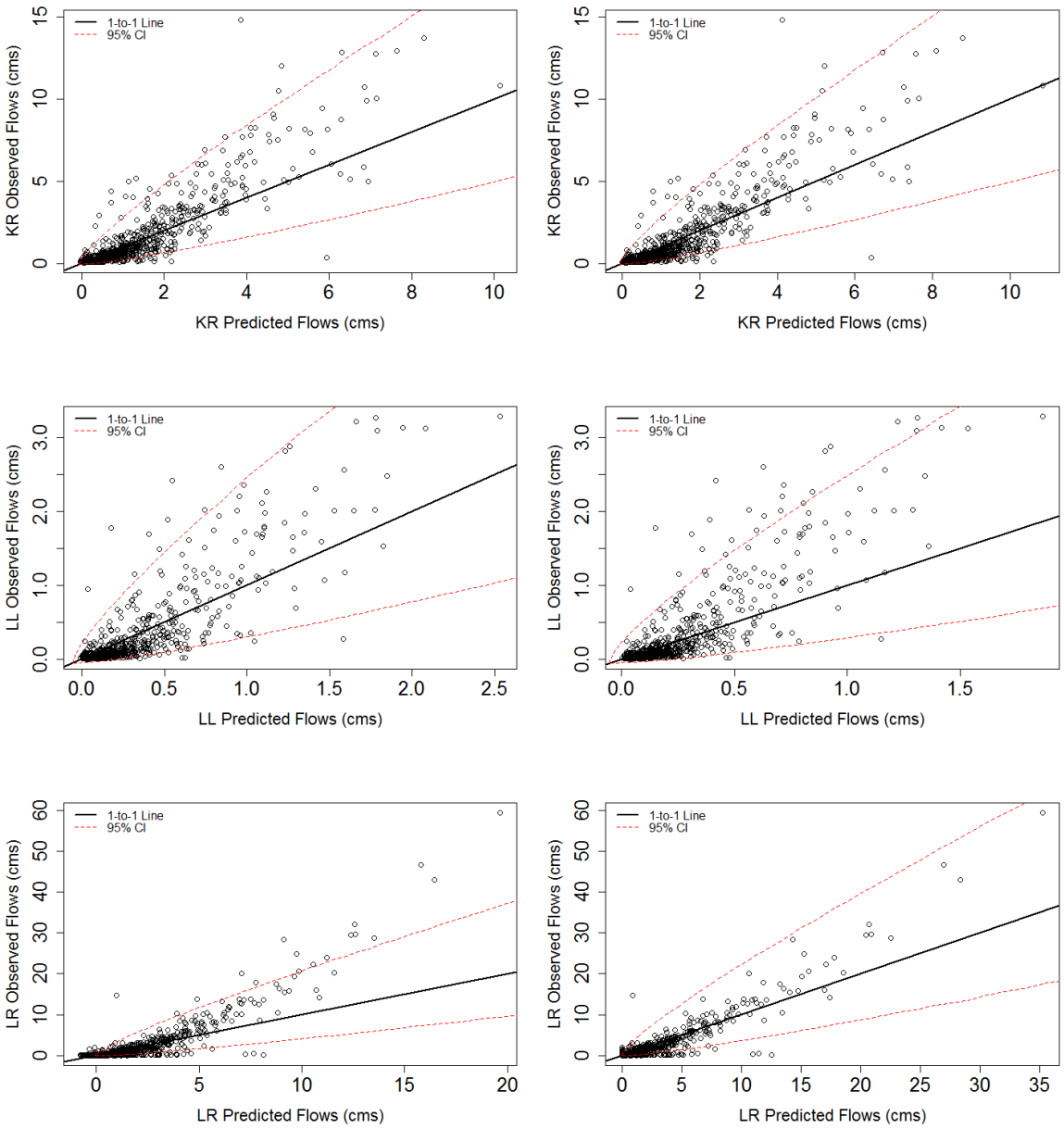
$$y = \beta_0 + Y * \beta_Y + M * \beta_M + M^2 * \beta_{M^2} + Q_W * \beta_{Q_W} + Q_E * \beta_{Q_E} + P * \beta_P \qquad \text{Equation 11}$$

Site	Nutrient	Intercept β_0	Intercept Std. Error	Year, Y β_Y	Year, Y Std. Error	Month, M β_M	Month, M Std. Error	Month Squared, M^2 β_{M^2}	Month Squared, M^2 Std. Error
EC	ON	37	4						
EC	TN	565	45	-3.61	0.71	-16.35	5.71	0.98	0.42
EC	TP	27	2						
ER	ON	-93	43	1.40	0.58	-0.09	0.02		
ER	TN	-1219	469	15.08	6.36	-8.46	2.42	0.59	0.18
ER	TP	7.34	0.81	-0.12	0.02	0.64	0.19	-0.06	0.01
KR	ON	17	5	-0.29	0.14				
KR	TN	96	15	-3.07	0.34				
KR	TP	9	5	-0.59	0.08	3.06	0.57	-0.23	0.04

Site	Nutrient	WWTP Flow, Q_W β_{Q_W}	WWTP Flow, Q_W Std. Error	Eno River Flow, Q_E β_{Q_E}	Eno River Flow, Q_E Std. Error	Plant Upgrade, P β_P	Plant Upgrade, P Std. Error
EC	ON			1.05	0.26	-13.5	4.4
EC	TN	416	119	-3.93	1.56	-422.8	21.3
EC	TP					-17.6	2.5
ER	ON	1046	492				
ER	TN	15976	5425				
ER	TP						
KR	ON						
KR	TN	460	56				
KR	TP	119	48	-0.59	0.20		

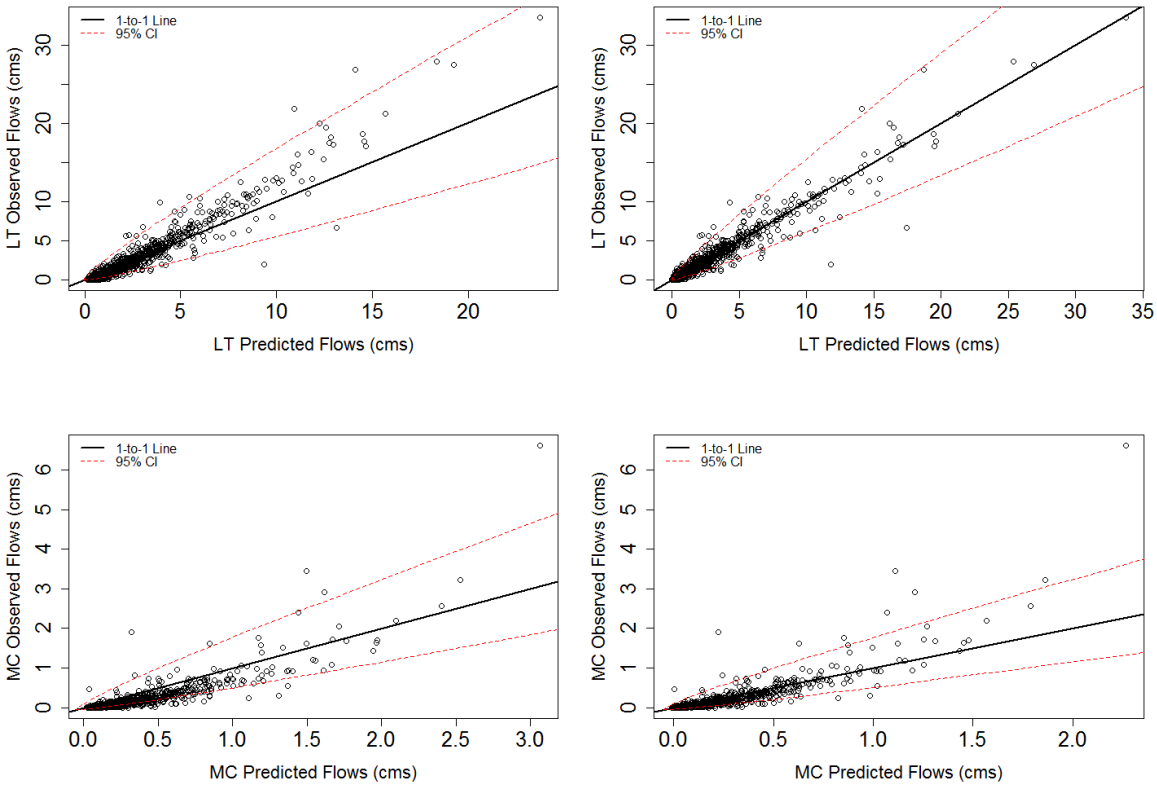
Appendix C. *Predicted versus Observed Flow Plots from Flow Model*

<u>Calibration</u> <u>Validation</u>



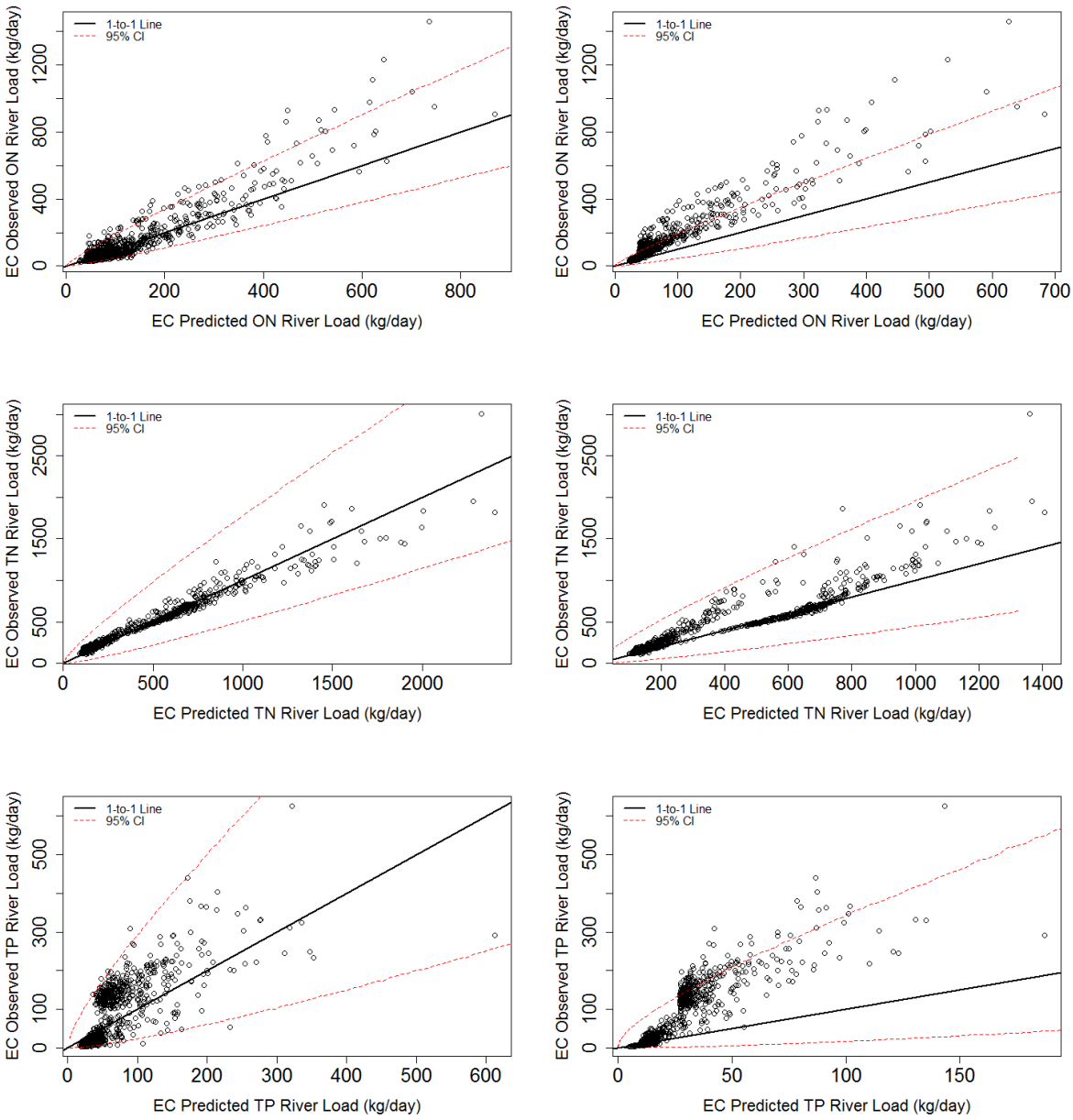
Calibration

Validation



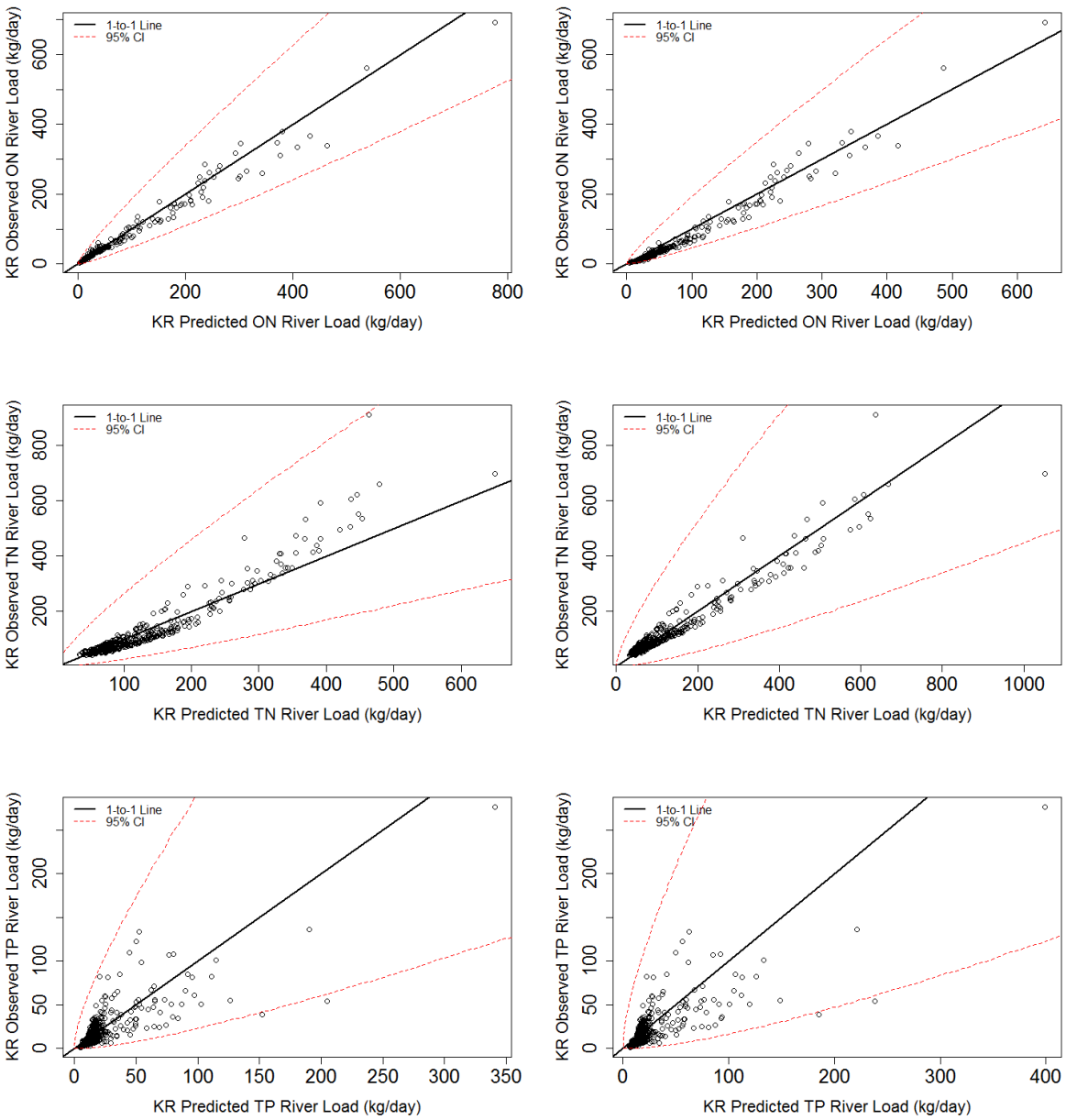
Appendix D. *Predicted versus Observed Load Plots from Nutrient Models*

Calibration | Validation



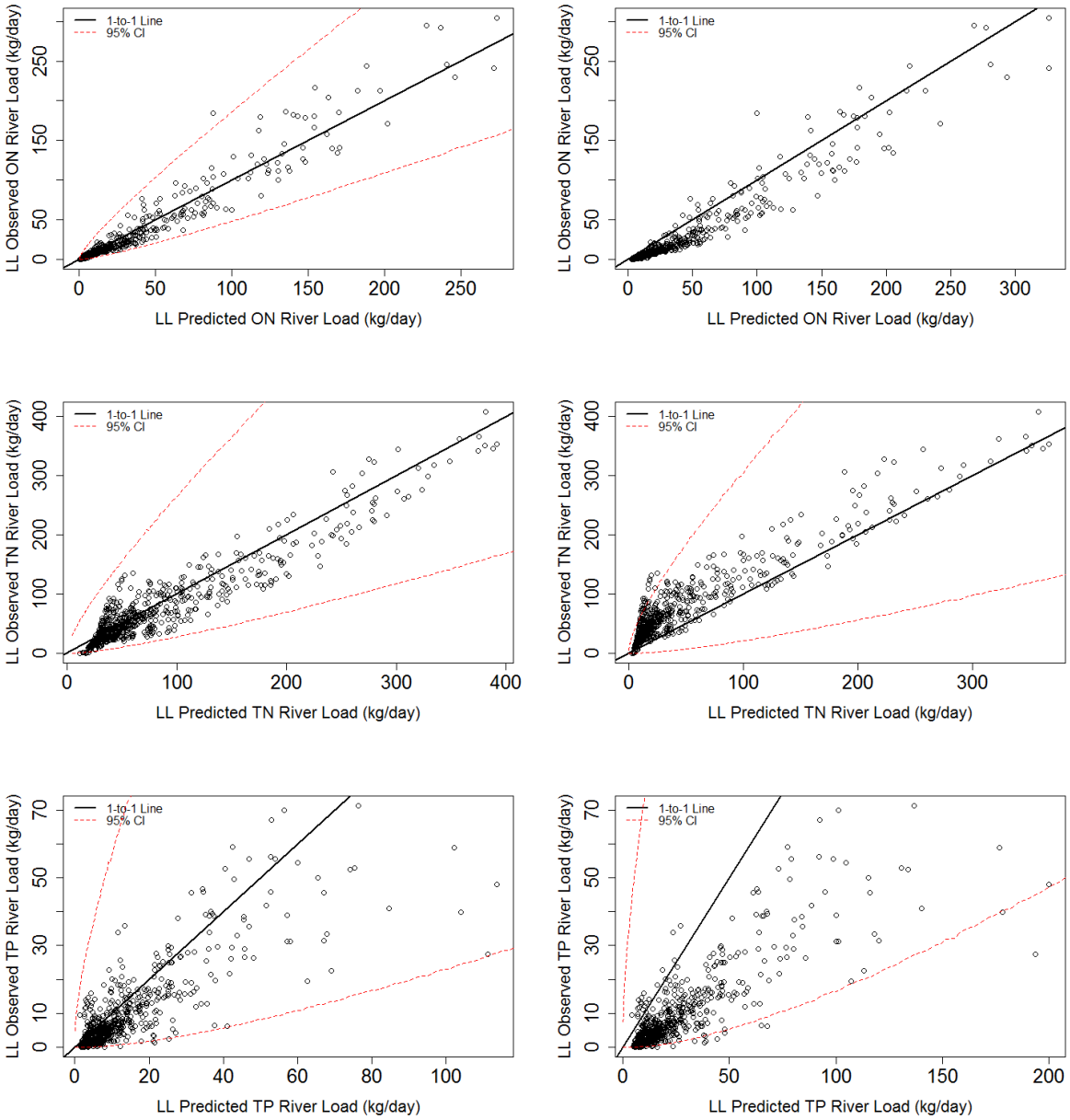
Calibration

Validation



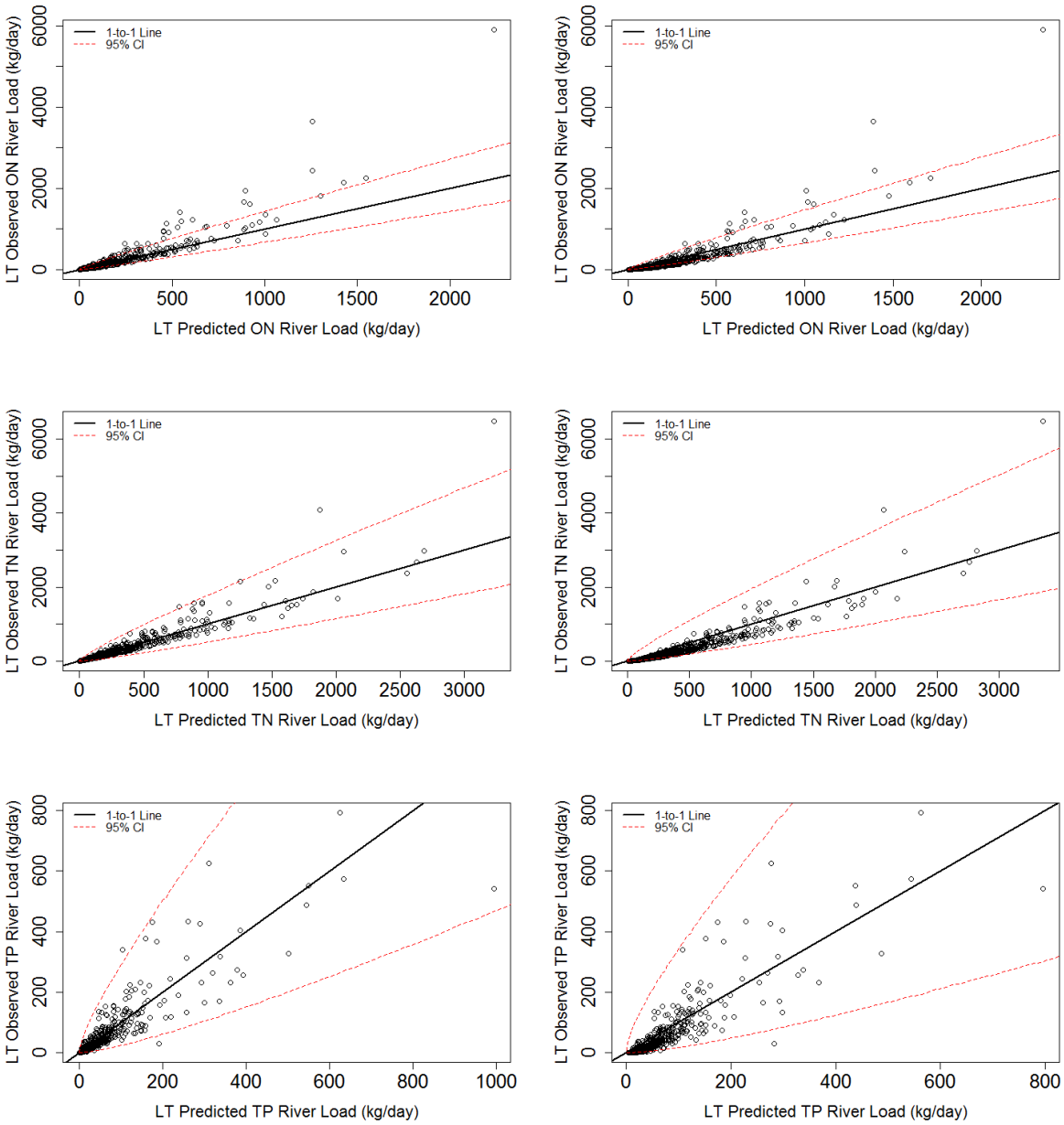
Calibration

Validation



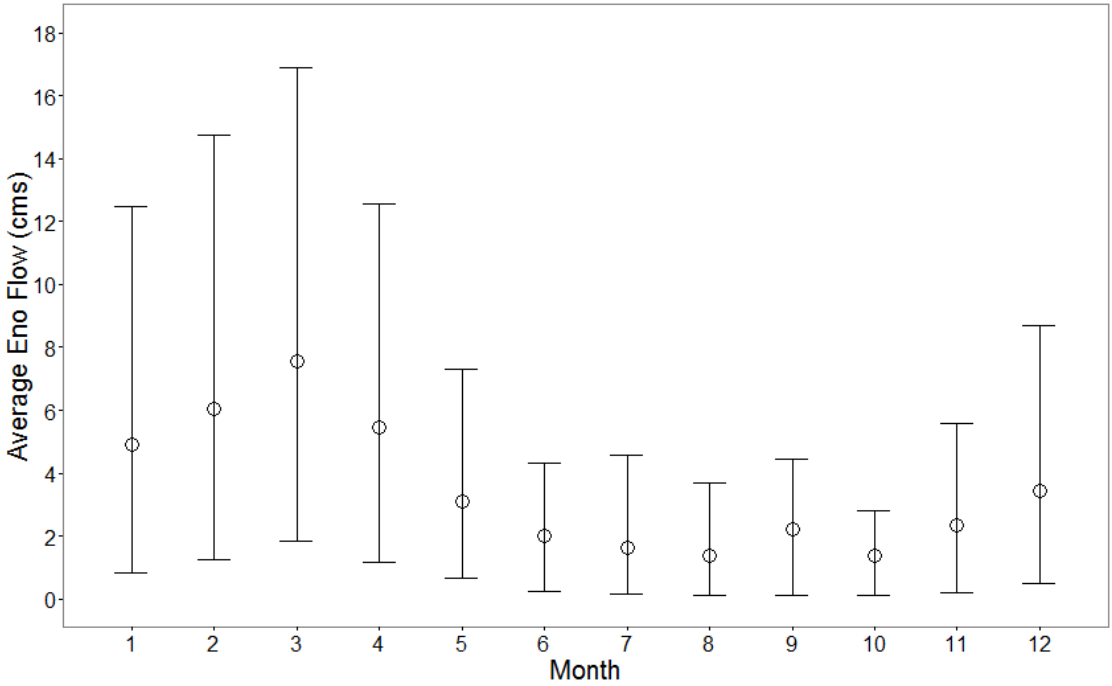
Calibration

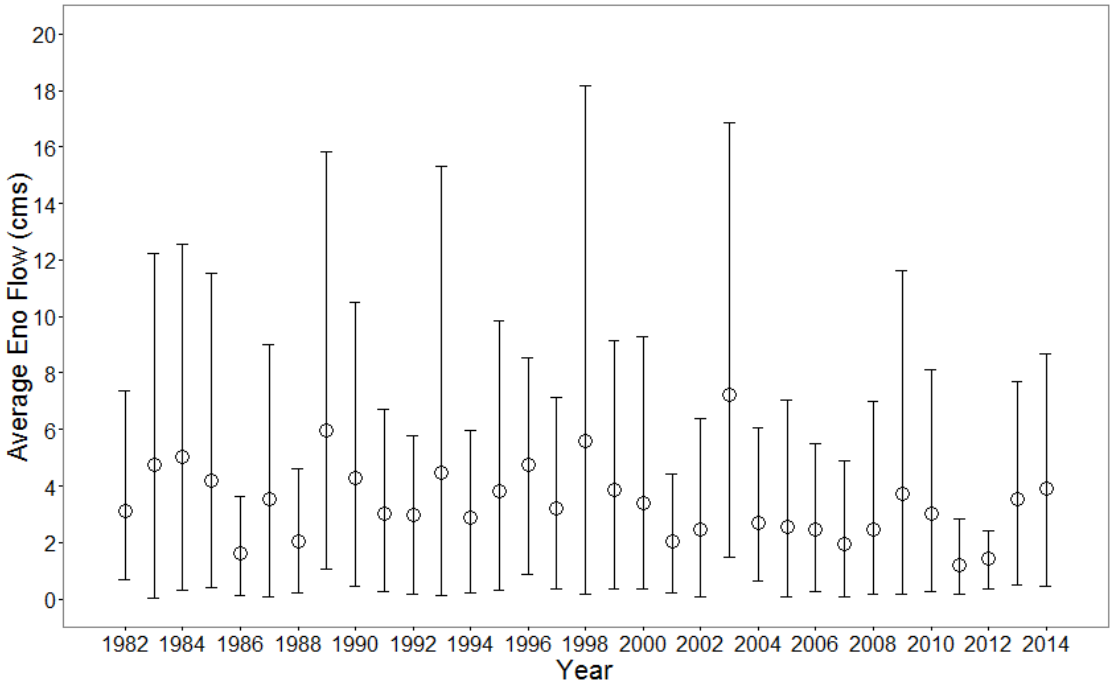
Validation

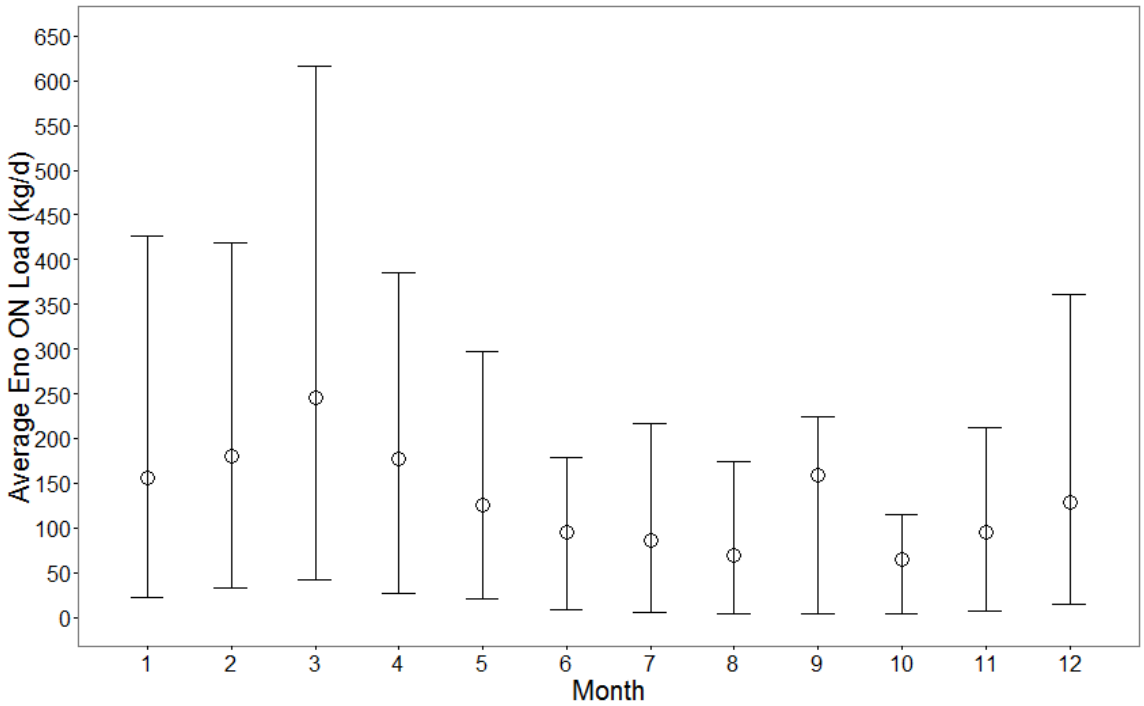


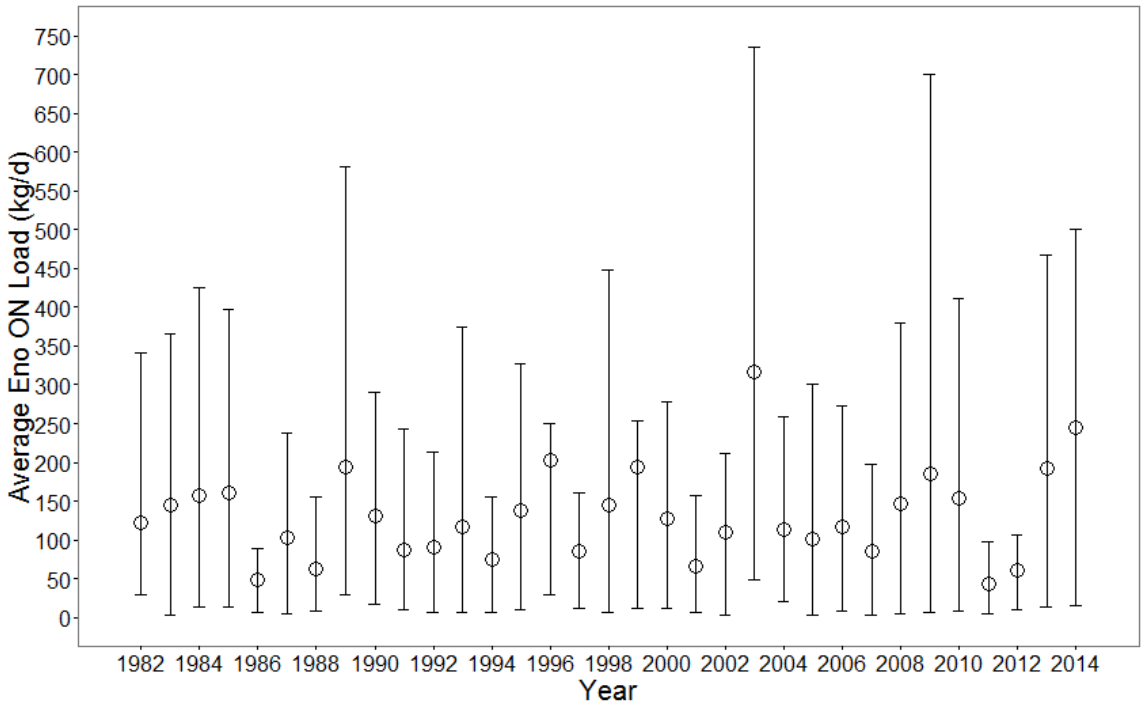
Appendix E. *Record of Flow and Nutrient Loads in the Eno River*

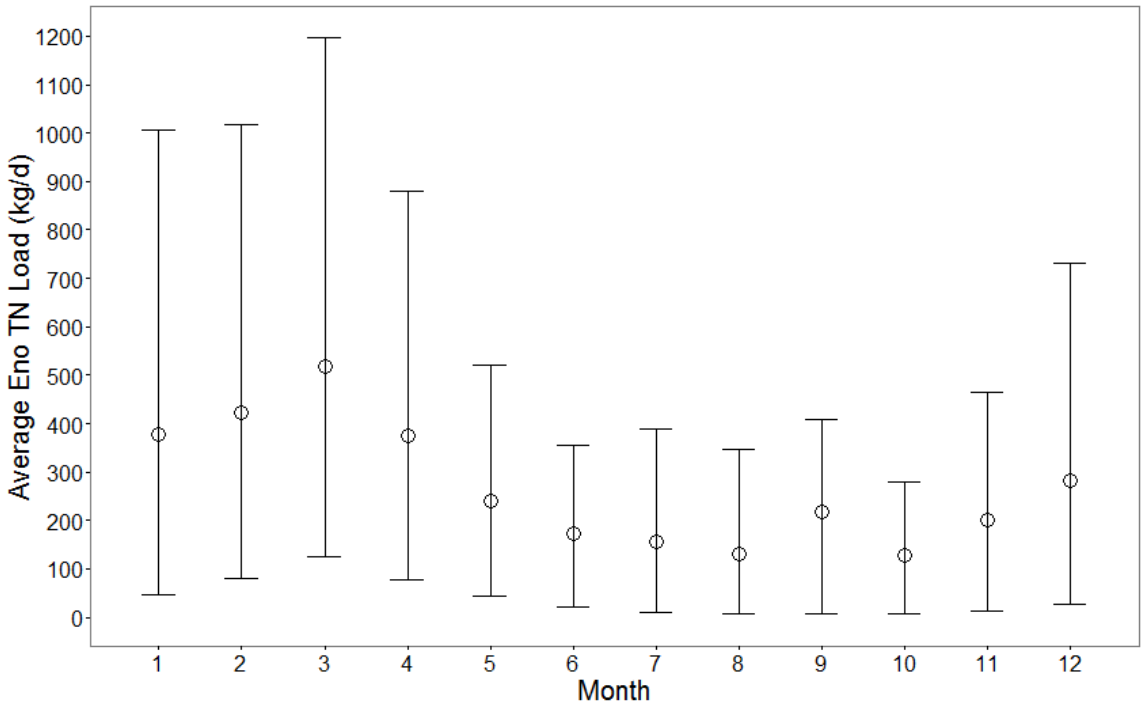
Points are the average value while bars indicate the 10% and 90% quantile value.

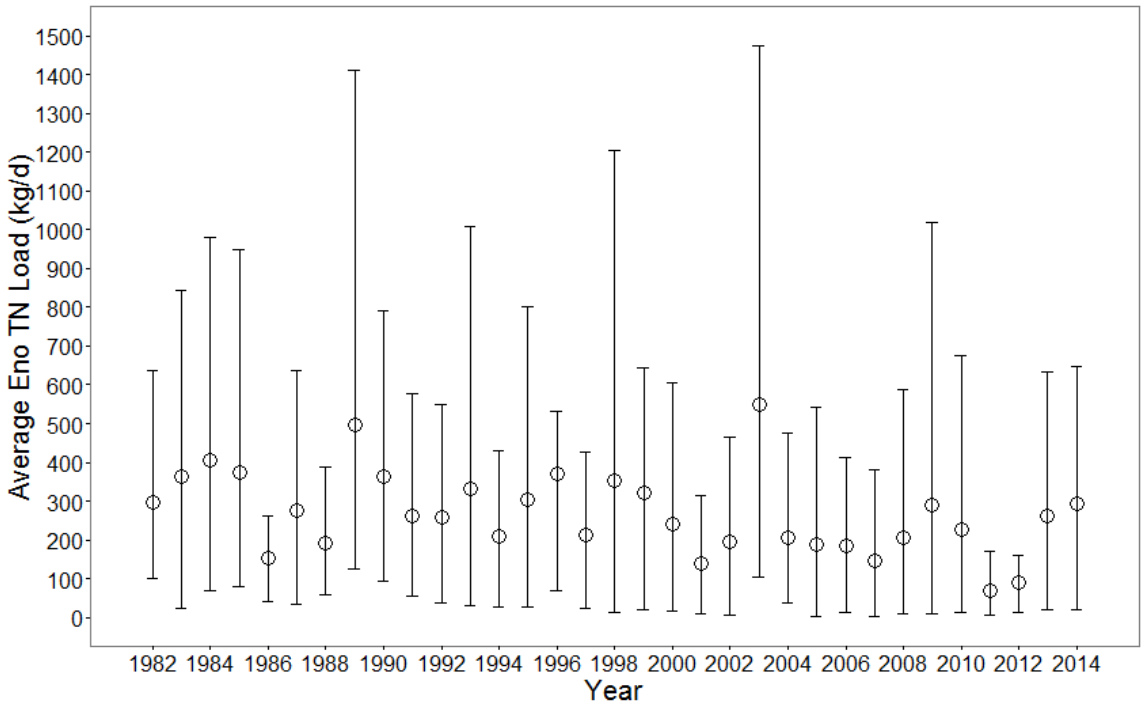


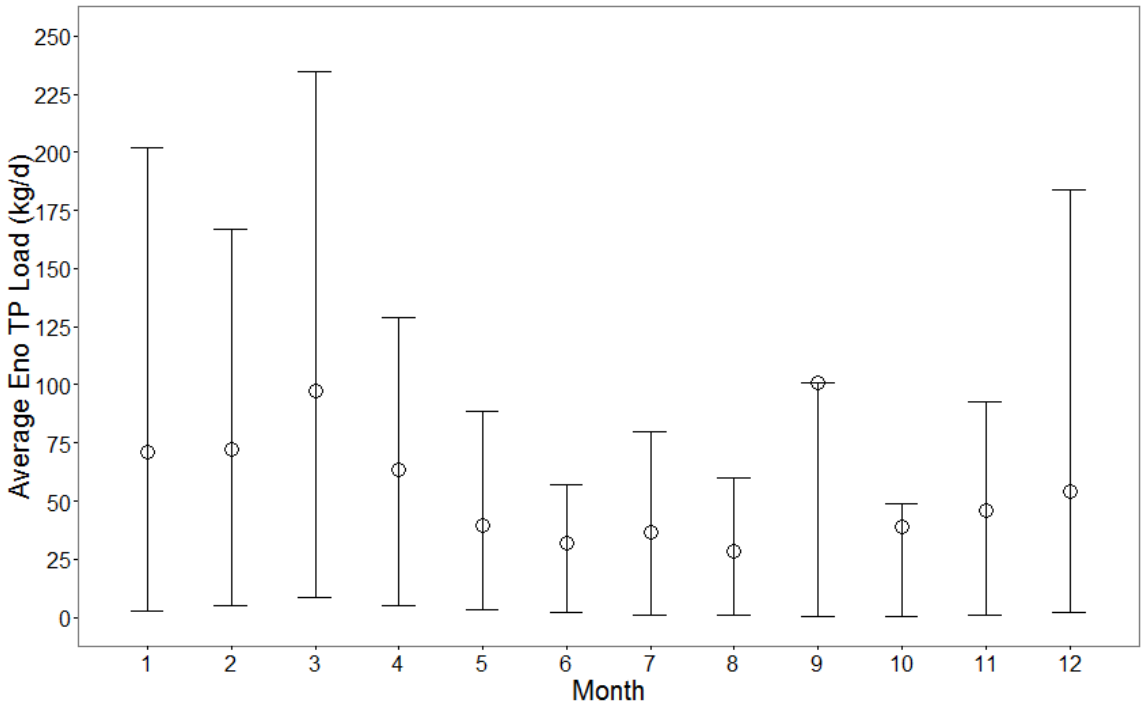


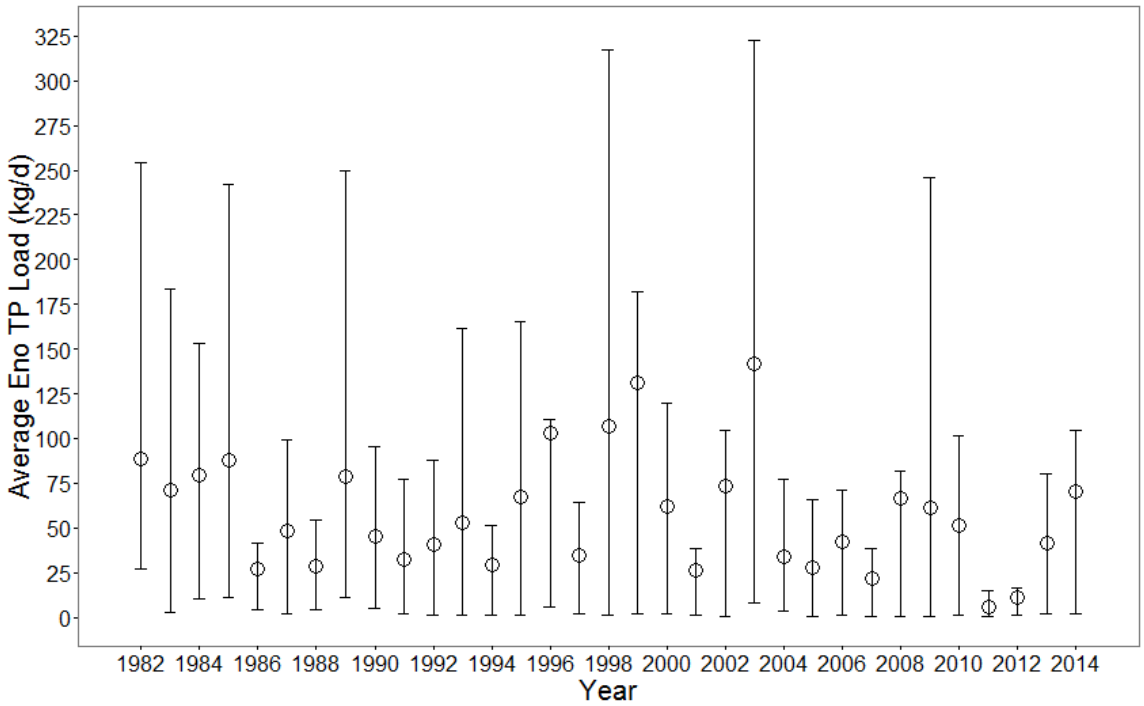

Average Eno ON Load (kg/d)
0
50
100
150
200
250
300
350
400
450
500
550
600
650
1
2
3
4
5
6
7
8
9
10
11
12
Month


Average Eno ON Load (kg/d)
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700
750
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year


Average Eno TN Load (kg/d)
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1
2
3
4
5
6
7
8
9
10
11
12
Month


Average Eno TN Load (kg/d)
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year


Average Eno TP Load (kg/d)
250
225
200
175
150
125
100
75
50
25
0
1
2
3
4
5
6
7
8
9
10
11
12
Month


Average Eno TP Load (kg/d)
325
300
275
250
225
200
175
150
125
100
75
50
25
0
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year