
Appendix E
**Watershed Model Development
and Calibration Report for
Juanita Creek Stormwater
Retrofit Study**
Ecology Grant: G0800618

August 2012



King County

Department of Natural Resources and Parks
Water and Land Resources Division

Science and Technical Support Section

King Street Center, KSC-NR-0600
201 South Jackson Street, Suite 600
Seattle, WA 98104

206-296-6519 TTY Relay: 711

www.kingcounty.gov/environment/wlr/science-section.aspx

Alternate Formats Available

206-296-6519 TTY Relay: 711

Appendix E

Watershed Model Development and Calibration Report for Juanita Creek Stormwater Retrofit Study

Ecology Grant: G0800618

Prepared for:

Stormwater Management Implementation Grant Program
Washington State Department of Ecology
Water Quality Program
Northwest Regional Office
Bellevue, Washington

Prepared by:

King County Science and Stormwater Services Sections
King County Water and Land Resources Division
Department of Natural Resources and Parks
Seattle, Washington



King County

Department of
Natural Resources and Parks

Water and Land Resources Division

Acknowledgements

This work was funded by Washington State Ecology through Stormwater Management Implementation Grant Program (Grant Number: **G0800618**), King County Stormwater Services Section, City of Kirkland Public Works Department, and Washington State Department of Transportation.

Citation

King County. 2011. Appendix E. Watershed Model Development and Calibration Report for Juanita Creek Stormwater Retrofit Study. Ecology Grant: G0800618. Prepared by Jeff Burkey. King County Department of Natural Resources and Parks. Water and Land Resources Division. Seattle, Washington.

Table of Contents

1.0.	Introduction.....	1
2.0.	Model development.....	2
2.1	Geographic Information System Layers.....	2
2.2	Channel Hydraulics.....	2
2.3	Survey Storm Sewer Network.....	3
2.4	Model Catchment Segmentation.....	4
2.5	Monitoring Locations.....	8
2.6	Local Weather Data.....	9
2.7	Hydraulic Response Units (HRUs).....	10
3.0.	Water Quantity Calibration.....	12
4.0.	Biotic Calibration.....	20
5.0.	Water Quality Calibration.....	24
5.1	Water Temperature.....	24
5.2	Total Suspended Solids.....	26
5.3	Dissolved Oxygen.....	30
5.4	Benthic Algae.....	32
5.5	Fecal Coliforms.....	33
5.6	Total Copper.....	36
5.7	Dissolved Copper.....	43
5.8	Nutrients.....	53
6.0.	Discussion.....	58
7.0.	References.....	60

Figures

Figure 1 Map of storm drain network with arrows depicting direction of flow. Yellow lines are derived catchment delineations using topography and stormwater network.....	4
Figure 2 Catchment map illustrating model segmentation. Light green are significant wetland areas, the blue lines and areas are modeled stream reaches and lakes.	6
Figure 3 Model schematic.....	7
Figure 4 Map of stream flow and water quality measurements.....	8
Figure 5 Map of weather monitoring stations used.	9
Figure 6 Scatter plots comparing observed to simulated mean daily flow rates.....	17
Figure 7 Scatter plots comparing observed to simulated daily maximum flow rates	18
Figure 8 Scatter plots comparing observed to simulated average monthly flow rates. Each of the 12 months is a different color dot.	19
Figure 9 Scatter plots with regressions lines for LPC, LPD, HPC, HPD, HPR, and QR.	21
Figure 10 Scatter plots with regression lines for TQmean, RBI, and PK2YR.....	22
Figure 11 Time series, scatter plot, and cumulative distribution plots of calibrated water temperature for the mainstem. Red is observed, blue is simulated.	25
Figure 12 Time series, scatter plot, and cumulative distribution plots of calibrated total suspended solids for the mainstem. Red is observed, blue is simulated.....	27
Figure 13 Time series, scatter plot, and cumulative distribution plots of calibrated total suspended solids for Billy Creek (2G). Red is observed, blue is simulated.	28
Figure 14 Time series, scatter plot, and cumulative distribution plots of calibrated total suspended solids for Totem Lake tributary (3G). Red is observed, blue is simulated..	29
Figure 15 Time series, scatter plot, and cumulative distribution plots of calibrated dissolved oxygen concentrations for the mainstem. Red is observed and blue is simulated.	31
Figure 16 Time series, scatter plot, and cumulative distribution plots of calibrated dissolved oxygen concentrations for the Totem Lake tributary (3G). Red is observed and blue is simulated.	32
Figure 17 Time series, scatter plot, and cumulative distribution plots of calibrated fecal colony forming units for the mainstem. Red is observed, blue is simulated.....	34
Figure 18 Time series, scatter plot, and cumulative distribution plots of calibrated fecal colony forming units for the outlet of wetland. Red is observed, blue is simulated.	35
Figure 19 Time series, scatter plot, and cumulative distribution plots of calibrated fecal colony forming units for the west headwater branch tributary. Red is observed, blue is simulated.	36

Figure 20 Time series, scatter plot, and cumulative distribution plots of calibrated total copper concentrations for the mainstem. Red colors are observed while blue is simulated. 38

Figure 21 Time series, scatter plot, and cumulative distribution plots of calibrated total copper concentrations for the wetland outlet (4GO). Red colors are observed while blue is simulated. Stream flows were not recorded at this location..... 39

Figure 22 Time series plot of total copper concentrations for outlet of wetland (4GO). Red colors are observed, blue is simulated..... 40

Figure 23 Time series, scatter plot, and cumulative distribution plots of calibrated total copper concentrations for the north branch (7G). Red colors are observed while blue is simulated..... 41

Figure 24 Time series plot of total copper concentrations for north branch tributary (7G). Red colors are observed, blue is simulated. 42

Figure 25 Time series, scatter plot, and cumulative distribution plots of calibrated for dissolved copper concentrations for the mainstem..... 44

Figure 26 Time series plot of dissolved copper concentrations for mainstem (1G). Red colors are observed, blue is simulated..... 45

Figure 27 Time series plot of dissolved copper concentrations for Billy Creek (2G). Red colors are observed, blue is simulated..... 46

Figure 28 Time series plot of dissolved copper concentrations for Totem Lake tributary (3G). Red colors are observed, blue is simulated..... 47

Figure 29 Time series plot of dissolved copper concentrations for outlet of wetland (4GO). Red colors are observed, blue is simulated. 48

Figure 30 Time series plot of dissolved copper concentrations for inlet to wetland (4GI). Red colors are observed, blue is simulated. 49

Figure 31 Time series plot of dissolved copper concentrations for west branch tributary (5G). Red colors are observed, blue is simulated..... 50

Figure 32 Time series plot of dissolved copper concentrations for east branch tributary (6G). Red colors are observed, blue is simulated..... 51

Figure 33 Time series plot of dissolved copper concentrations for north branch tributary (7G). Red colors are observed, blue is simulated..... 52

Figure 34 Time series, scatter plot, and cumulative distribution plots of calibrated for ammonia-n concentrations for the mainstem (1G). Red is observed and blue is simulated. 54

Figure 35 Time series, scatter plot, and cumulative distribution plots of calibrated for nitrate concentrations for the mainstem (1G). Red is observed and blue is simulated. 55

Figure 36 Time series, scatter plot, and cumulative distribution plots of calibrated for orthophosphorus concentrations for the mainstem (1G). Red is observed and blue is simulated. 56

Tables

Table 1. Individual catchment areas in acres..... 5

Table 2 Conversion of land use to HRUs for pervious and impervious land segments (PERLNDs and IMPLNDs)..... 10

Table 3 Hydrologic Response Units. 11

Table 4 List of calibrated HSPF parameters 13

Table 5 Summary statistics for calibration of hourly time increment flow rates..... 14

Table 6 Summary of model accuracy for flow rates using quantiles characterizing various magnitudes of flow rates. 15

Table 7 Summary of annual 7-day minimums and instantaneous maximums..... 15

Table 8 Summary of flashiness metrics used in this study and resultant *p*-value using Mann-Whitney *U*-test comparing observed and simulated flow rates for mainstem Juanita Creek (station 27a). 16

Table 9 Regression coefficients (a,b) for predicting BIBI (*y*) from hydrologic flashiness metrics. 20

Table 10 Simulated versus observed BIBI scores. 22

Table 11 Summary statistics for calibrated water temperature. 24

Table 12 Summary statistics for calibrated total suspended solids..... 26

Table 13 Summary statistics for calibrated dissolved oxygen. 30

Table 14 Summary statistics for calibrated fecal coliforms. 33

Table 15 Summary statistics of calibration for total and dissolved copper..... 36

Table 16 Summary statistics of calibration for total and dissolved copper..... 43

Table 17 Summary of Mann-Whitney U-Test of calibration for nitrogen and phosphorus species..... 53

Table 18 Summary statistics for calibration of ammonia-N. 57

Table 19 Summary statistics for calibration of nitrates. 57

Table 20 Summary statistics for orthophosphate..... 57

1.0. INTRODUCTION

A hydrologic watershed model was developed for the Juanita Creek drainage basin to evaluate various future stormwater management and mitigation scenarios. Juanita Creek drainage basin is predominately located within the City of Kirkland and drains to Juanita Bay on the east shore of Lake Washington. The model simulations will provide guidance on techniques and their relative effectiveness towards meeting the long-term objective of fully restoring beneficial uses to Juanita Creek.

The Juanita Creek hydrology model is based on Hydrologic Simulation Program- FORTRAN (HSPF). The model simulates water quantity and water quality at multiple catchments throughout the drainage. The model output is used to calculate flow, water quality, geomorphology, and biology metrics to evaluate the projected improvements under different mitigation scenarios. This appendix documents the development and calibration of the watershed model. This is presented to provide context to the level of applicability and relevancy when evaluating simulation results (e.g. model accuracy, inferring conclusions, etc.) and informing policy makers.

2.0. MODEL DEVELOPMENT

Development of the model included assembly of several types of data including:

- land use,
- land cover,
- surficial geology,
- channel/wetland/lake hydraulics,
- measured atmospheric conditions and stream flows, and
- sampled water quality.

The water quantity mechanics are modeled using three flow pathways representative of surface runoff, shallow subsurface (interflow), and shallow active ground water resurfacing in the stream within the same catchment. Water quality elements are modeled using “build up/wash off” for constituents defined to accumulate (build up) at constant rates creating the potential of wash off, as well as user defined seasonally variable interflow and groundwater concentrations. The rate of wash off varies depending on antecedent conditions leading up to a storm, as well as the rainfall intensity and duration. Further detail on the individual parameters is given in subsequent sections of this appendix.

2.1 Geographic Information System Layers

- Several data layers were used to develop the watershed model land use and land cover was identified based on a coverage derived from King County 2002 imagery.
- Topography was identified based on a digital elevation map derived from King County LiDAR (Light Detection And Ranging) imagery
- Soil characteristics were identified and generalized based on the King County surficial geology map.
- Catchment delineations were developed using topographic conditions generated from the afore mentioned LiDAR

2.2 Channel Hydraulics

Stream channel width, depth, slope, roughness, and shape were estimated using various methods depending on the data available and the types of features controlling the behavior of the stream system. Relationships between water depth, surface area, channel volumes, and flow rates were defined based on the most hydraulically controlling feature in a stream reach per catchment. These control points were either a result of culverts, channel roughness, or combinations of both. The transient surface area and storage volumes were determined based on the length of the channel through the catchment and an assumed typical cross-section. Regional stormwater ponds and large storage area (e.g. lakes or wetlands) volumes were based on either topographic contours using King County LiDAR or bathometric data obtained from City of Kirkland.

The distributed storage volumes of the existing storm sewer pipes/vaults/ponds in the basin were integrated into the assumed stream reach representing the channel for a given

catchment. While these volumes can be relatively small compared to the stream reach, the added storage could have an effect on frequent small storms, which are more sensitive to some of the metrics used in analysis. Where HEC-RAS modeling was performed, those outputs were used to define the primary stream channel stage-area-storage-discharge relationships.

2.3 Survey Storm Sewer Network

A detailed survey of the storm drainage network including all pipes, underground vaults, ponds, and open channels was completed (Figure 1). The information was used to refine catchment delineations originally derived from landscape topography to more accurately identify interceptions and redirection of surface runoff. Further detail on these data can be obtained from City of Kirkland, Washington.

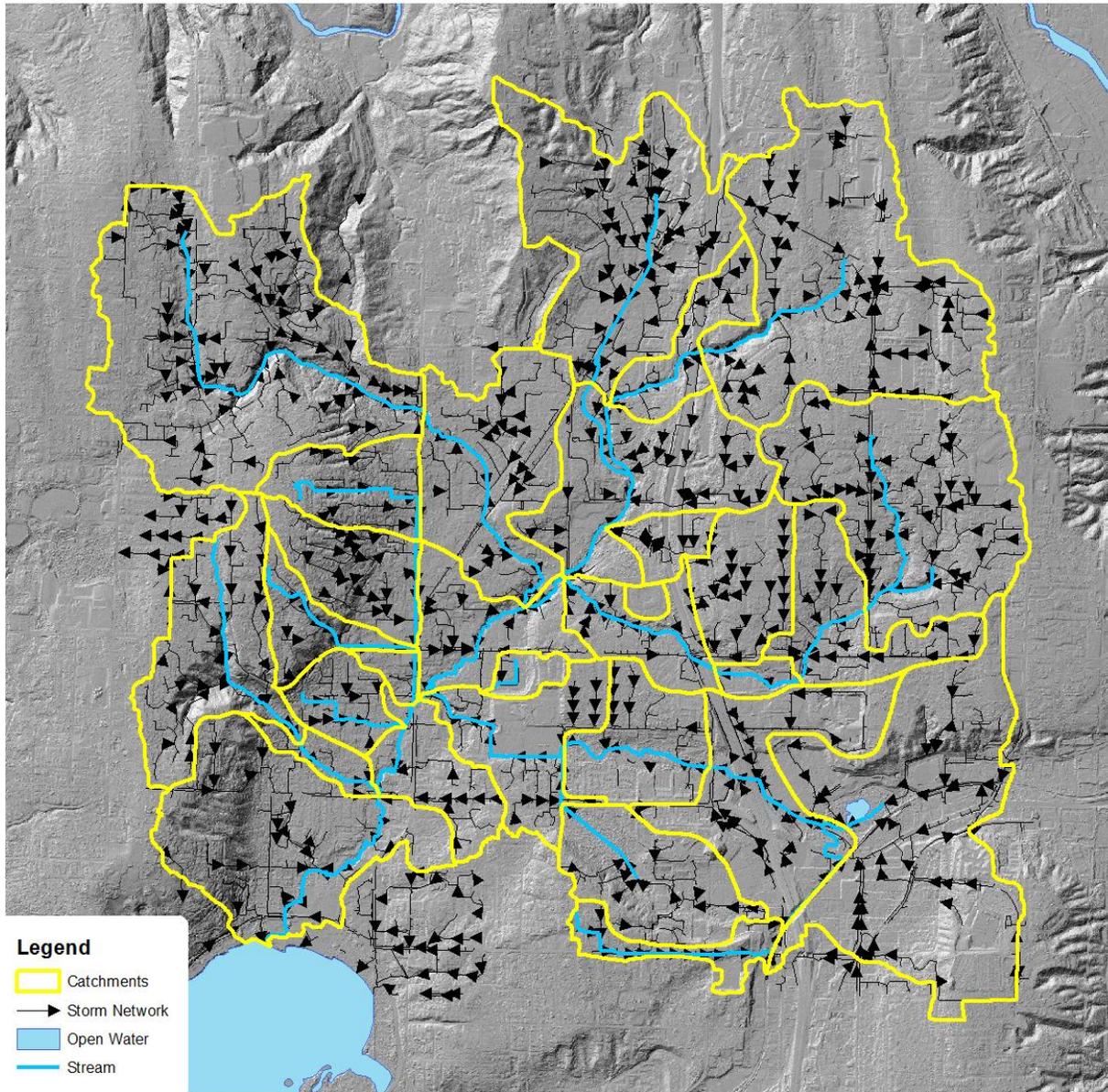


Figure 1 Map of storm drain network with arrows depicting direction of flow. Yellow lines are derived catchment delineations using topography and stormwater network.

2.4 Model Catchment Segmentation

The 6.8-square mile Juanita Creek drainage was segmented into 30 catchments with similar meteorological conditions, topographical features, land use practices, and/or are a region of interest (e.g., non-point source loads need to be quantified). Once the catchments and channel segments have been defined, these catchments must then be further refined to: 1) develop the model categories to represent; 2) define the physical parameters (e.g., elevation, slopes, channel length) for HSPF using available data; and 3) establish initial calibration parameters for HSPF based on past applications within the region and past

experience with the model. Figure 2 presents a map of the catchments and Figure 3 illustrating the linkages of the catchments and channel reaches. Given the densely piped storm network and fast travel times (measured in minutes), catchments were kept relatively small in size. The median catchment area is 99 acres, with catchments ranging from 9 to 484 acres based on landscape conditions covering 4,343 acres (Table 1).

Table 1. Individual catchment areas in acres.

Catchment	Acres	Catchment	Acres
WA3001	203	WA3016	27
WA3002	20	WA3017	484
WA3003	134	WA3018	199
WA3004	202	WA3019	307
WA3005	77	WA3020	417
WA3006	101	WA3021	46
WA3007	119	WA3022	409
WA3008	81	WA3023	178
WA3009	318	WA3024	264
WA3010	93	WA3025	97
WA3011	54	WA3026	122
WA3012	84	WA3027	34
WA3013	69	WA3028	18
WA3014	53	WA3029	9
WA3015	16	WA3030	108

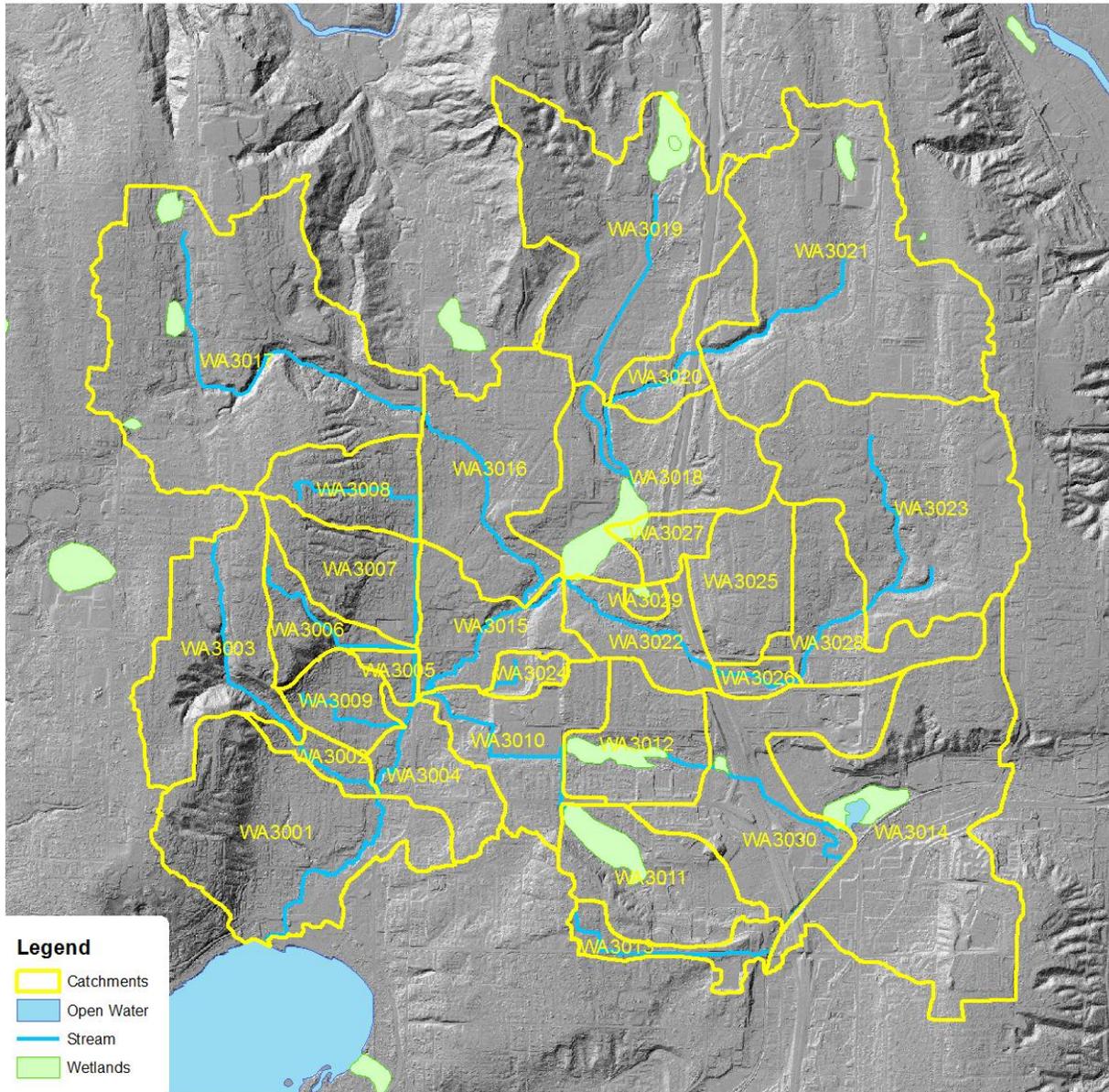


Figure 2 Catchment map illustrating model segmentation. Light green are significant wetland areas, the blue lines and areas are modeled stream reaches and lakes.

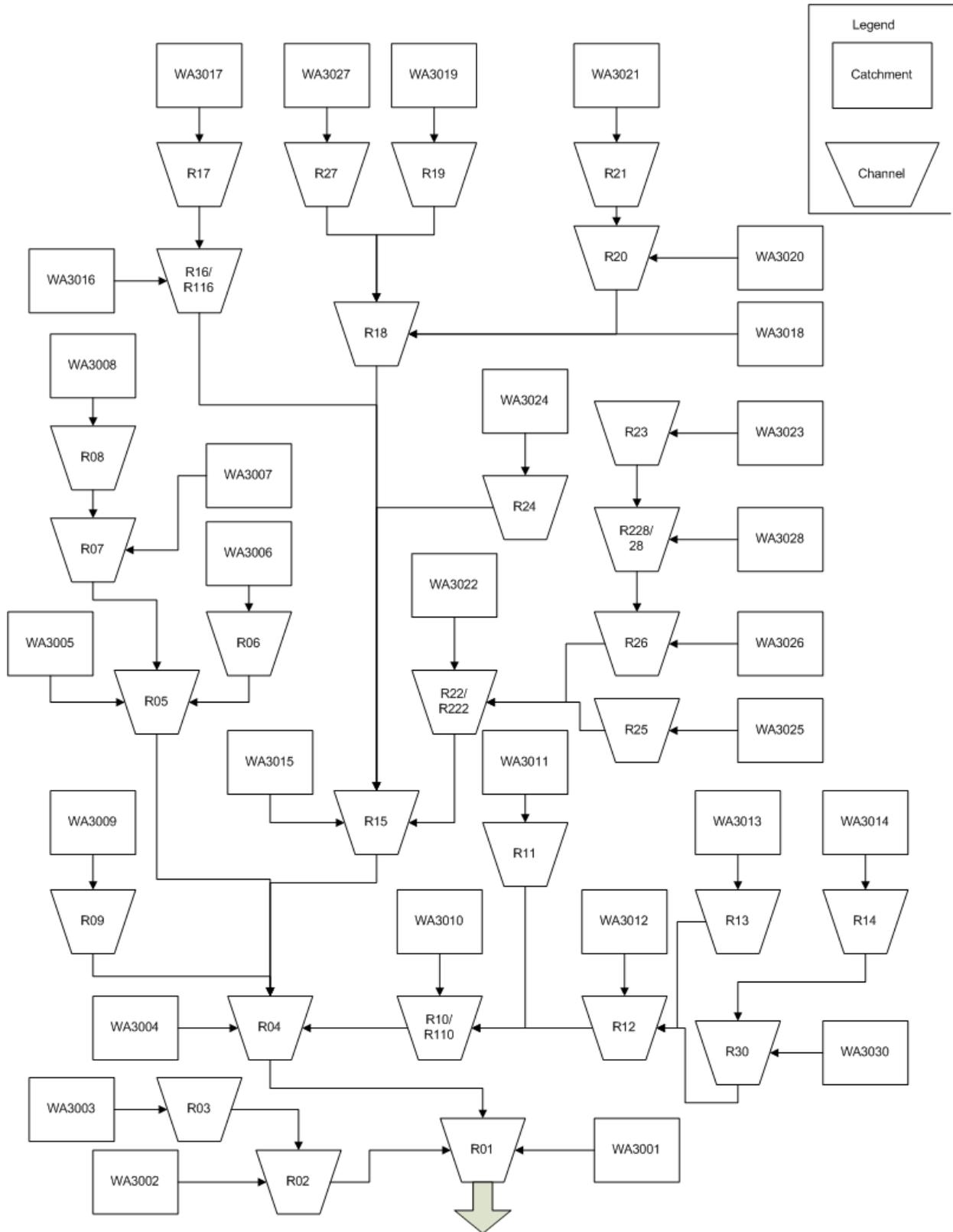


Figure 3 Model schematic

2.5 Monitoring Locations

Detailed stream flow and water quality monitoring was conducted between October 2008 and March 2010. Stream flow gauges were installed in the five major tributaries (stations 27i, 27C, 27DN, 27H, and 27J) and in the mainstem near the mouth (station 27A). Water quality samples were either collected at or near the same location (stations 2G, 3G, 5G, 6G, 7G, and 1G). Additional water quality samples were taken near the inlet (station 4GI) and outlet (station 4GO) of a large wetland downstream of Totem Lake where the majority of commercial development in the basin area exists (Figure 4). Further detail on the monitoring conducted as part of this study can be found in a monitoring report located in appendix A.

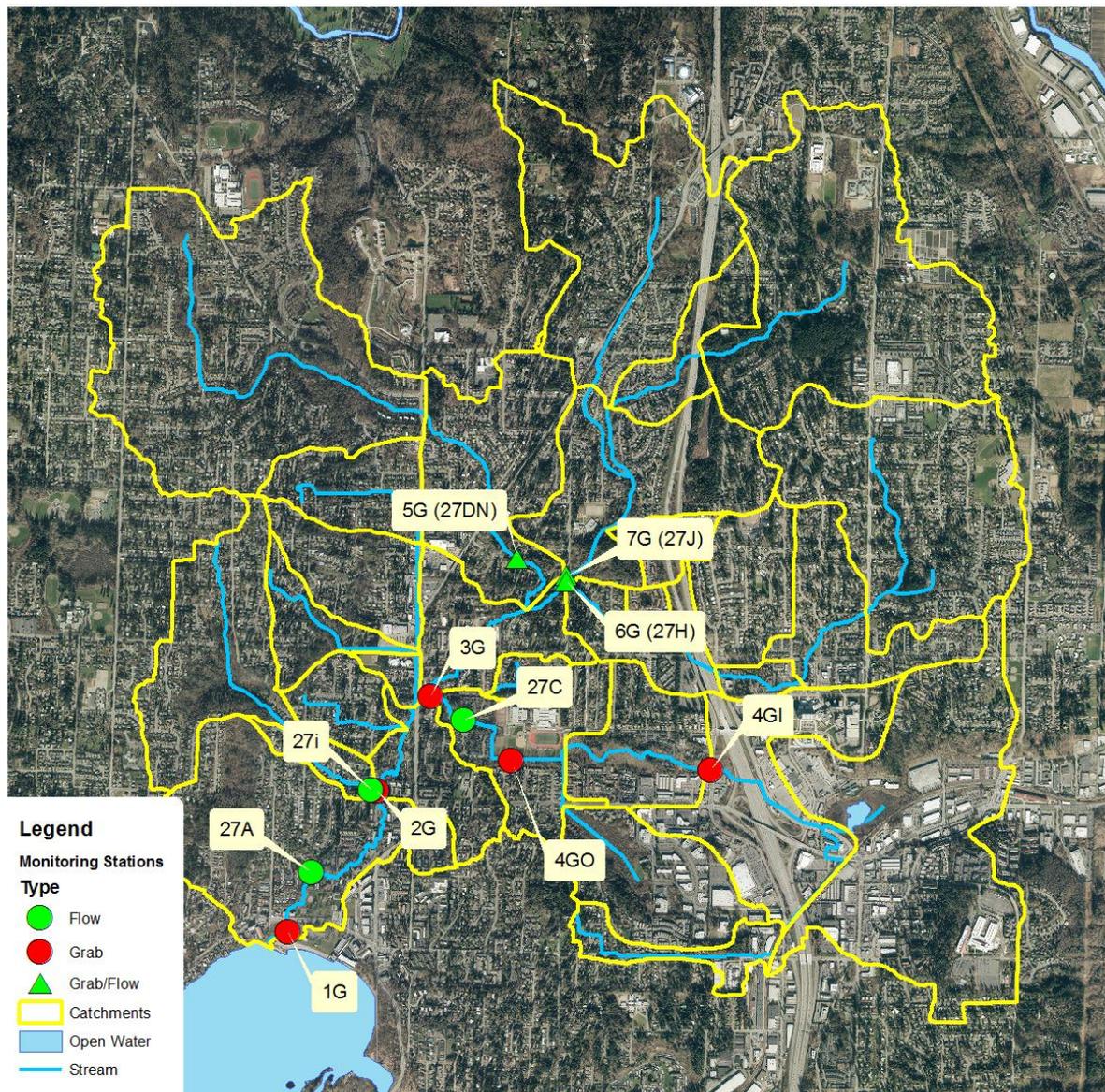


Figure 4 Map of stream flow and water quality measurements.

2.6 Local Weather Data

Precipitation, air temperature, dew point, wind speed and cloud cover were derived from National Weather Service Sea-Tac monitoring station. Solar radiation was obtained from NOAA Integrated Surface Irradiance Study (ISIS) network Sand Point station. Lastly, computed evapotranspiration was obtained from Washington State University Agricultural Weather Network, Puyallup station.

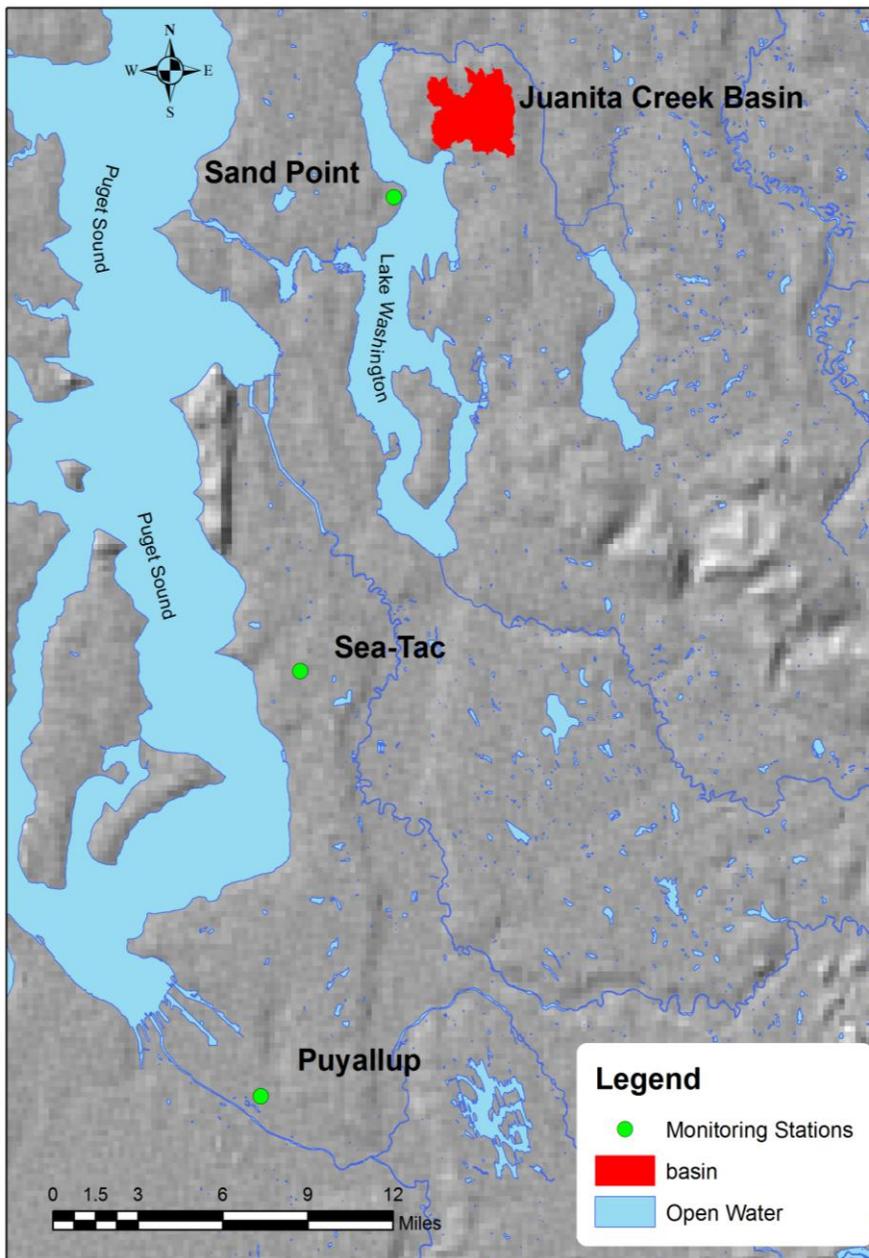


Figure 5 Map of weather monitoring stations used.

2.7 Hydraulic Response Units (HRUs)

For modeling purposes, a distinction is made between total and effective impervious area. Total impervious area includes all surfaces that do not infiltrate runoff. Roofs, paved streets, sidewalks, driveways, and parking lots are all part of the total impervious area. Effective impervious area (EIA) is defined as the area where there is no opportunity for surface runoff from an impervious site to infiltrate into the soil before it reaches a conveyance system (pipe, ditch, stream, etc.). Because it is extremely expensive and time consuming to look at every impervious surface in a watershed to determine whether or not it is an effective impervious area, average EIA values are used instead. Each average EIA value is based on the land use (forest, low density residential, high density residential, commercial, etc.) and previous experience in other Puget Sound lowland watersheds. The following percentages in Table 2 are representative values used in development of the Juanita Creek model, and are similar to other calibrated model schemes in King County.

Table 2 Conversion of land use to HRUs for pervious and impervious land segments (PERLNDs and IMPLNDs).

Land cover \ GIS	Forest	Pasture / Grass	Grass-Lt Urban	Grass-Dens. Urban	EIA-Lt. Urban	EIA-Dens. Urban	EIA-Roads	Wetlands
Mixed Forest	100%							
Forest	100%							
Scrub/Shrub	50%	50%			0%			
Dry Ground		50%	50%		0%			
Light Urban (< 75%)		5%		70%	25%			
Dense Urban (> 75%)				25%		75%		
Bare Ground				70%	30%			
Wetlands								100%
Open Water ¹								
Roads				15%			85%	

Integrating the slope and surficial geology into the above listed land covers (Table 2), generates a list of 37 unique hydrologic response units (HRUs) used in the model (Table 3). Not all HRUs may be present depending on presence of a given element during the integration process.

Table 3 Hydrologic Response Units.

HRU #	Soil	Land Cover	Slope	
11	Till	Forest	Flat	
12			Mild	
13			Moderate	
14			Steep	
21		Pasture	Flat	
22			Mild	
23			Moderate	
24			Steep	
31		Residential Forest	Flat	
32			Mild	
33			Moderate	
34			Steep	
41		Light Density Grass	Flat	
42			Mild	
43			Moderate	
44			Steep	
61		High Density Grass	Flat	
62			Mild	
63			Moderate	
64			Steep	
71		Outwash	Forest	n/a
72			Pasture	
73			Residential Forest	
74			Light Density Grass	
75	Moderate Density Grass			
76	High Density Grass			
81	Saturated	Forest	n/a	
82		Pasture		
83		Residential Forest		
84		Light Density Grass		
85		Moderate Density Grass		
86		High Density Grass		
87		Wetlands		
91	Impervious Surfaces	Low Density Residential	n/a	
92		High Density Residential		
93		Commercial/Industrial		
94		Roads		

3.0. WATER QUANTITY CALIBRATION

An iterative procedure of parameter evaluation and refinement was used to determine parameter values to use in the watershed model. Calibration was based on fifteen years (1995-2010) of simulation for the mainstem and slightly more than one year for the tributaries to evaluate parameters under a variety of climatic, soil moisture, and water quality conditions.

Calibration includes the comparison of both monthly and annual values, and individual storm events, whenever sufficient data are available for these comparisons. In addition, when a continuous or large number of observed record are available, simulated and observed values are analyzed on a frequency basis and their resulting empirical cumulative distributions (e.g., duration curves) compared to assess the model behavior and agreement over the full range of observations.

A weight of evidence approach is most widely used and accepted when models are examined and judged for acceptance as no single procedure or statistic is widely accepted as measuring, nor capable of establishing, acceptable model performance. Therefore, the calibration relied on numerous statistical tests (e.g., correlation tests, model fitness, test of distributions) and graphical plots (e.g., scatter, time series, frequency) to determine the model's ability to represent the system. Parameters that are used to evaluate exceedances of specific thresholds are compared using more rigorous statistical measures (correlation tests and model fitness), whereas parameters used for annual loading rates are tested for similar distributions using Mann-Whitney U-Test.

The following four characteristics of the watershed hydrology were evaluated (in the order shown): (1) annual water balance, (2) seasonal and monthly flow volumes, (3) baseflow, and (4) storm events. Simulated and observed values for reach characteristic are examined and critical parameters are adjusted to attain acceptable levels of agreement (discussed further below).

The critical parameters that govern the annual water balance are as follows:

- LZSN - lower zone soil moisture storage (inches).
- LZETP - vegetation evapotranspiration index (dimensionless).
- INFILT- infiltration index for division of surface and subsurface flow (inches/hour).
- UZSN - upper zone soil moisture storage (inches).
- DEEPPFR- fraction of groundwater inflow to deep recharge (dimensionless).

Changes in LZSN and LZETP affect evapotranspiration by making more or less moisture available to evaporate or transpire. Both LZSN and INFILT also have a major impact on percolation and are important in obtaining an annual water balance. In addition, on extremely small watersheds (less than 200 to 500 acres) that contribute runoff only during and immediately following storm events, the UZSN parameter can also affect annual runoff

volumes because of its impact on individual storm events (described below). While there was no assumed loss of groundwater via DEEPFR (i.e. DEEPFR = 0), there was some intra-basin transfers of groundwater among the catchments connected to contiguous outwash soils but with topographic divides to achieve mass balance.

The portion of stream baseflow is adjusted in conjunction with the seasonal/monthly flow calibration (previous step) because moving runoff volume between seasons often means transferring the surface runoff from storm events in wet seasons to low-flow periods during dry seasons. By adjusting INFILT, runoff can be shifted to either increase or decrease groundwater or baseflow conditions. The shape of the groundwater recession; i.e., the change in baseflow discharge is controlled by the following parameters:

AGWRC- groundwater recession rate (1 / day).

KVARY- index for nonlinear groundwater recession.

AGWRC is calculated as the rate of baseflow (i.e. groundwater discharge to the stream) on one day divided by the baseflow on the previous day; thus AGWRC is the parameter that controls the rate of outflow from the groundwater storage. These values are adjusted as needed through calibration. The KVARY index allows users to impose a nonlinear recession so that the slope can be adjusted as a function of the groundwater gradient. KVARY ranges were based on soil types to account for changes in recession rates between wet and dry seasons. Parameters associated with impervious surfaces (HRUs 91 through 94) are not differentiated for water quantity and are characterized with a different set of parameters based on flow length (150 ft), slope (0.01 ft/ft), surface roughness ($n = 0.15$) and surface storage (0.10 inches). The list of parameter values are summarized in Table 4 below.

Table 4 List of calibrated HSPF parameters

HRU #	LZSN	INFILT	LSUR	SLSUR	KVARY	AGWRC
11	4.6	0.493	350	0.028	0.405	0.997
12	3.8	0.462	300	0.072	0.405	0.997
13	3.2	0.431	250	0.116	0.405	0.997
14	2.6	0.4	200	0.195	0.405	0.997
21	3.2	0.308	350	0.026	0.405	0.996
22	3.2	0.277	300	0.07	0.405	0.996
23	3.6	0.224	250	0.116	0.45	0.997
24	3.6	0.196	200	0.186	0.45	0.997
31	4.6	0.493	350	0.028	0.405	0.997
32	3.8	0.431	300	0.072	0.405	0.997
33	4.3	0.336	250	0.116	0.45	0.998
34	4.3	0.28	200	0.195	0.45	0.998
41	3.2	0.308	350	0.028	0.405	0.995
42	3.2	0.277	300	0.07	0.405	0.995
43	3.6	0.224	250	0.117	0.45	0.996
44	3.6	0.196	200	0.18	0.45	0.996
61	2.9	0.224	350	0.03	0.45	0.996

HRU #	LZSN	INFILT	LSUR	SLSUR	KVARY	AGWRC
62	2.9	0.196	300	0.071	0.45	0.996
63	2.9	0.168	250	0.114	0.45	0.996
64	2.9	0.14	200	0.172	0.45	0.996
71	5.8	3.36	300	0.089	0.27	0.995
72	5.8	3.36	300	0.06	0.27	0.995
73	5.8	0.678	300	0.089	0.27	0.995
74	5.8	0.678	300	0.077	0.27	0.995
75	7.2	0.462	300	0.067	0.3	0.996
76	7.2	0.323	300	0.067	0.3	0.996
81	2.3	4.4	150	0.048	0.45	0.997
82	2.6	3	150	0.043	0.5	0.998
83	2.6	4.4	150	0.048	0.5	0.998
84	2.3	3	150	0.043	0.45	0.997
85	2.6	2.4	150	0.046	0.5	0.998
86	2.6	2.4	150	0.075	0.5	0.998
87	2.6	3	150	0.043	0.5	0.998

Table 5 represents model accuracy based on continuous hourly time steps for the available period of record per calibration point. Values closer to unity for r-square (i.e. explanation of variance) and the Nash-Sutcliffe (i.e. model accuracy taking into account “signal to noise” ratio) denote a more accurate model for estimating flow rates.

Table 5 Summary statistics for calibration of hourly time increment flow rates.

Statistic	Mainstem (27a)	Billy Creek (27i)	Totem Lake Trib. (27c)	West Branch (27dn)	East Branch (27h)	North Branch (27j)
Pearson Coefficient	0.86	0.76	0.88	0.8	0.91	0.91
Mean Error (cfs)	-0.2	0.06	0.31	0.08	0.1	0.23
RMSE (cfs)	8.1	0.45	1.99	1.35	0.74	1.49
r-square	0.74	0.58	0.78	0.64	0.83	0.82
Mean Absolute Error (cfs)	4.19	0.2	1.05	0.64	0.34	0.69
Nash-Sutcliffe	0.69	0.44	0.61	0.52	0.81	0.75
Skill Score ¹	0.45	0.25	0.37	0.31	0.56	0.5

¹Skill Score = 1 – RMSE/Std. Dev.

Table 6 summarizes how well the model simulates a broader spectrum of flow rates grouped into five thresholds for analysis. One subbasin (Totem Lake tributary) is poorly calibrated to annual low flow conditions such that simulated flow rates are about twice observed with a relative percent difference of 187% (e.g. simulated might be 2 cfs while observed may be 1 cfs).

Table 6 Summary of model accuracy for flow rates using quantiles characterizing various magnitudes of flow rates.

Statistic	Mainstem (27a)	Billy Creek (27i)	Totem Lake Trib. (27c)	West Branch (27dn)	East Branch (27h)	North Branch (27j)
Mean	-2%	18%	13%	6%	8%	10%
90-Percentile	5%	22%	11%	27%	-2%	26%
75-Percentile	-2%	12%	-15%	-8%	3%	10%
50-Percentile	-19%	26%	-2%	-14%	20%	-5%
25-Percentile	-23%	-3%	38%	-35%	32%	-8%
10-Percentile	-13%	2%	187%	-13%	25%	-26%

The differences between simulated and actual annual 7-day minimum flow rates, instantaneous maximum flows were assessed (Table 7). This analysis shows the model characterizes observed conditions quite well except for low flow conditions as previously mentioned for Totem Lake (27c) and now East Branch (27h). However, these types of statistics based on annual events is tenuous at best, given the short period of available data for all calibration points except near the mouth (27a).

Table 7 Summary of annual 7-day minimums and instantaneous maximums.

Metric	Statistic	Mainstem (27a)	Billy Creek (27i)	Totem Lake Trib. (27c)	West Branch (27dn)	East Branch (27h)	North Branch (27j)
Annual 7-Day Low Flow	Difference (cfs)	-0.06	0.05	0.79	0.06	0.31	0.23
	RPD	-1%	26%	90%	9%	55%	17%
Instantaneous Daily Maximums	Difference (cfs)	1.9	0.02	0.8	0.03	0.29	0.68
	RPD	11%	3%	24%	1%	12%	16%
Instantaneous Annual Maximums	Difference (cfs)	1.99	-0.51	1.4	-3.76	1.82	-7.39
	RPD	1%	-8%	4%	-19%	11%	-18%

In addition to the typical methods of testing model accuracy, an assessment was also performed on the hydrologic flashiness metrics used in this study (Table 8). However, these metrics depend on annual summaries and only the mainstem (27a) had sufficient amount of data for evaluation. Using the Mann-Whitney *U*-test, two of the metrics individually fail the alternative hypothesis of equivalence with *p*-values less than 0.05 (HPC and RBI); however, combining all metrics together does pass the test (*p*-value \approx 0.70).

Table 8 Summary of flashiness metrics used in this study and resultant *p*-value using Mann-Whitney *U*-test comparing observed and simulated flow rates for mainstem Juanita Creek (station 27a).

Metric	Name	Description	<i>p</i> value
LPC	Low Pulse Count	Number of times each calendar year that discrete low flow pulses occurred	0.08
LPD	Low Pulse Duration	Annual average duration of low flow pulses during a calendar year	0.70
HPC	High Pulse Count	Number of days each water year that discrete high flow pulses occur	0.03 ²
HPD	High Pulse Duration	Annual average duration of high flow pulses during a water year	0.39
HPR	High Pulse Range	Range in days between the start of the first high flow pulse and the end of the last high flow pulse during a water year	0.12
QR	Flow Reversals	The number of times that the flow rate changed from an increase to a decrease or vice versa during a water year. Flow changes of less than 2% are not considered	0.09
TQ _{mean}	TQ _{mean}	The fraction of time during a water year that the daily average flow rate is greater than the annual average flow rate of that year	0.54
RBI	R-B Index	Richards-Baker Index – A dimensionless index of flow oscillations relative to total flow, based on daily average discharge measured during a water year	0.00 ²
P2YR ¹	Peak 2-yr:Winter Baseflow	Ratio of the estimated 2-year peak flow to winter baseflow (i.e., mean flow for October through April)	n/a

¹Relationship between metric and BIBI still in development as part of EPA WRIA 9 grant.

²Simulated metric is statistically different from observed, simulation fails test.

To understand model accuracy beyond the statistics previously provided, scatter plots are included to visually inspect for added value and illustrate possible aspects not encapsulated statistically. For reference, a 1:1 line is drawn illustrating a perfect fit of simulated to observed. Mean daily flows are illustrated in Figure 6, while daily maximum flows are shown in Figure 7. Additionally, visualizing model predictions of seasonality vis-à-vis mean monthly flow rates (Figure 8) highlights any volumetric bias that otherwise may cancel out looking at annual mass balances. For those series of graphs, each month is color coded to more easily see a consistent bias for any given month.

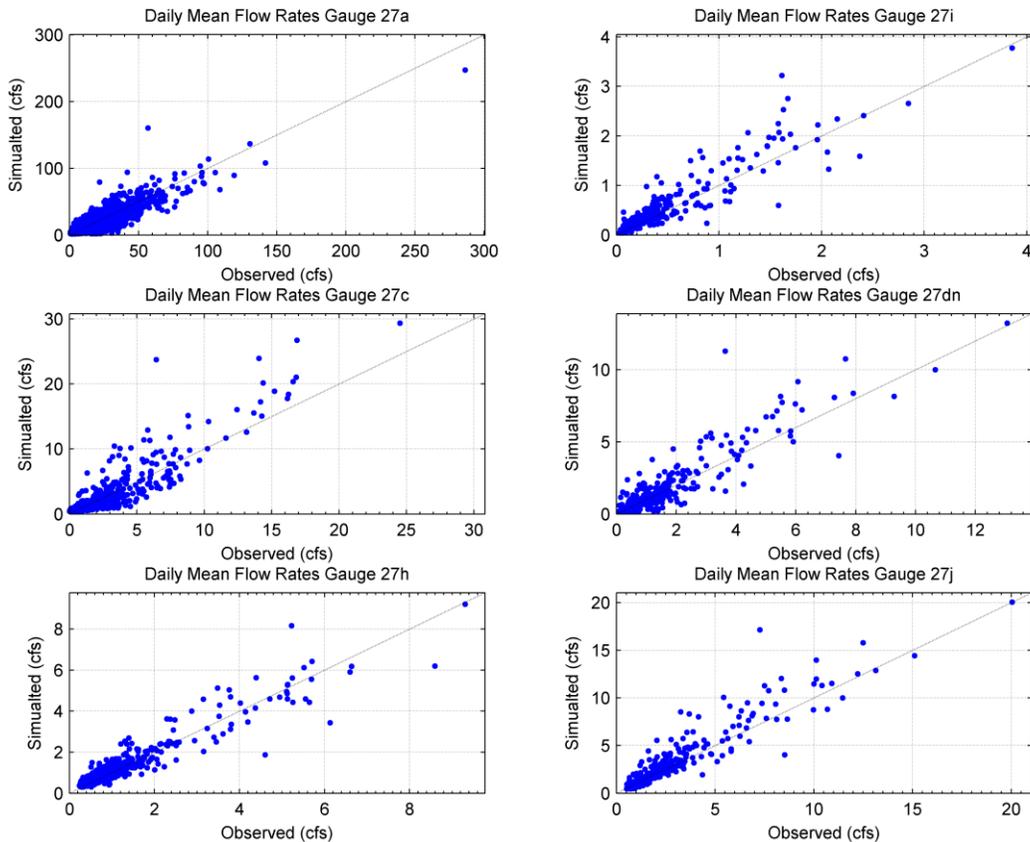


Figure 6 Scatter plots comparing observed to simulated mean daily flow rates

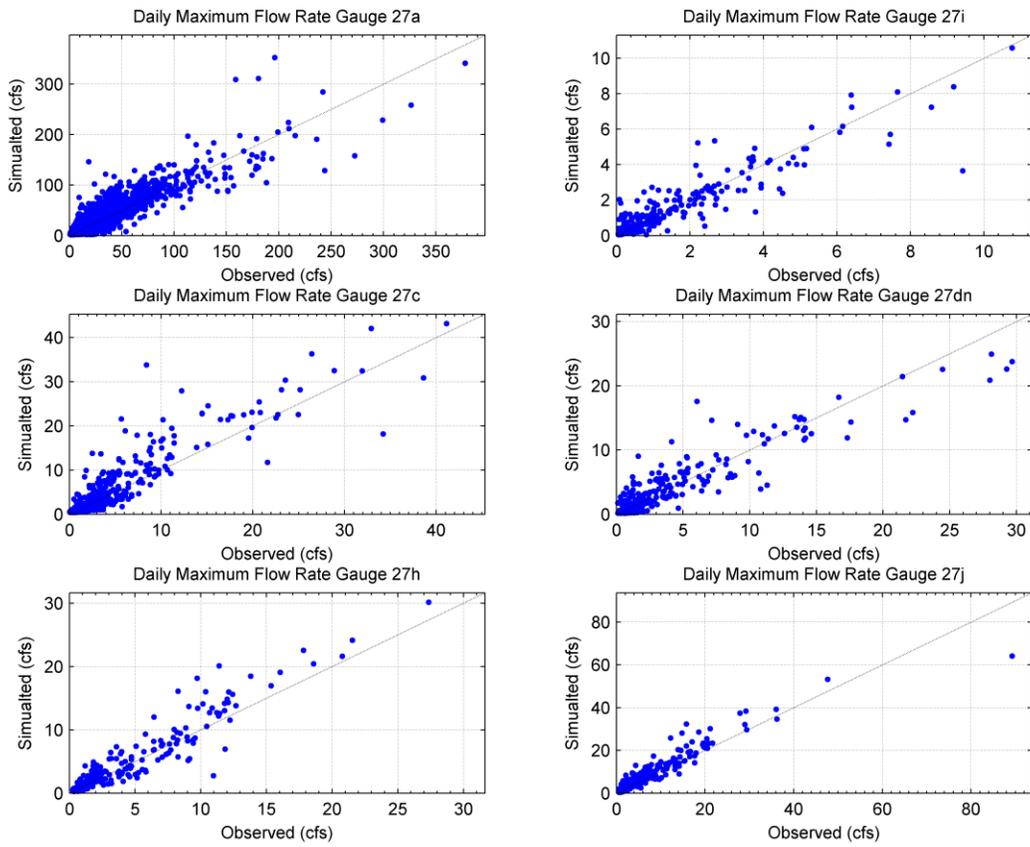


Figure 7 Scatter plots comparing observed to simulated daily maximum flow rates

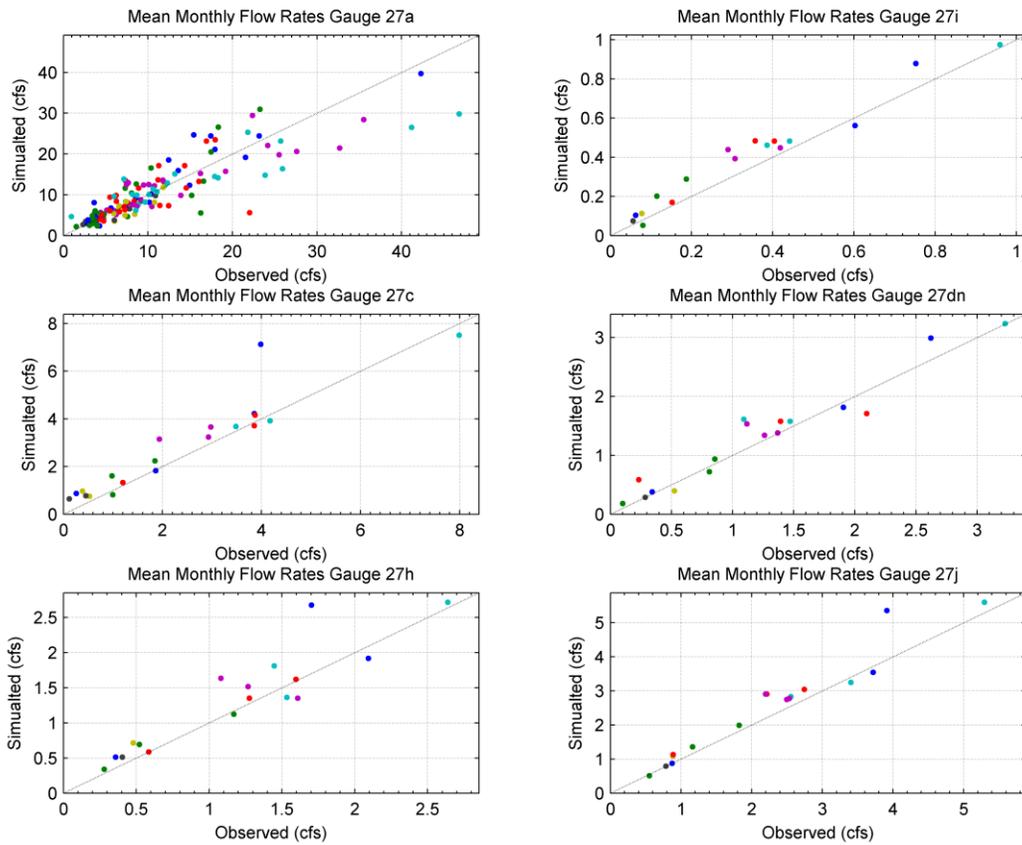


Figure 8 Scatter plots comparing observed to simulated average monthly flow rates. Each of the 12 months is a different color dot.

4.0. BIOTIC CALIBRATION

Biotic calibration for this study is based on relationships between hydrology and Benthic Index of Biotic Integrity (BIBI), but includes no parameter adjustments to the watershed model. These relationships are derived from a previous study identifying significant correlations between hydrologic flashiness metrics (previously defined in Table 8) and BIBI data collected at 16 different locations within King County (DeGasperi, et al., 2009). Model accuracy is dependent on the correlation between predicted outcomes from the regressions and observed. For this study, defined regressions are characterized into three types: log-linear (Equation 1), linear (Equation 2), and exponential (Equation 3). Summarized in Table 9 are the coefficients of the regressions and its corresponding explanatory power as measured using R^2 . Scatter plots characterizing the data as well as the regression predictions are illustrated in Figure 9 and Figure 10.

Table 9 Regression coefficients (a,b) for predicting BIBI (y) from hydrologic flashiness metrics.

Metric (x)	Equation Used	a	b	R^2
LPC	1	45.331	-22.466	0.44
LPD	1	-5.1273	23.214	0.59
HPC	1	53.05	-30.106	0.71
HPD	1	8.9753	23.498	0.64
HPR	2	44.167	-0.1148	0.73
QR	2	66.994	-0.7664	0.42
TQ _{mean}	2	-21.493	147.3	0.47
RBI	2	38.616	-51.851	0.49
P2YR	3	57.277	-0.311	0.22

Equation 1 Regression used for LPC, LPD, HPC, and HPD.

$$y = a + b \log_{10} x$$

Equation 2 Regression used for HPR, QR, TQ_{mean}, and RBI

$$y = a + bx$$

Equation 3 Regression used for P2YR.

$$y = ax^b$$

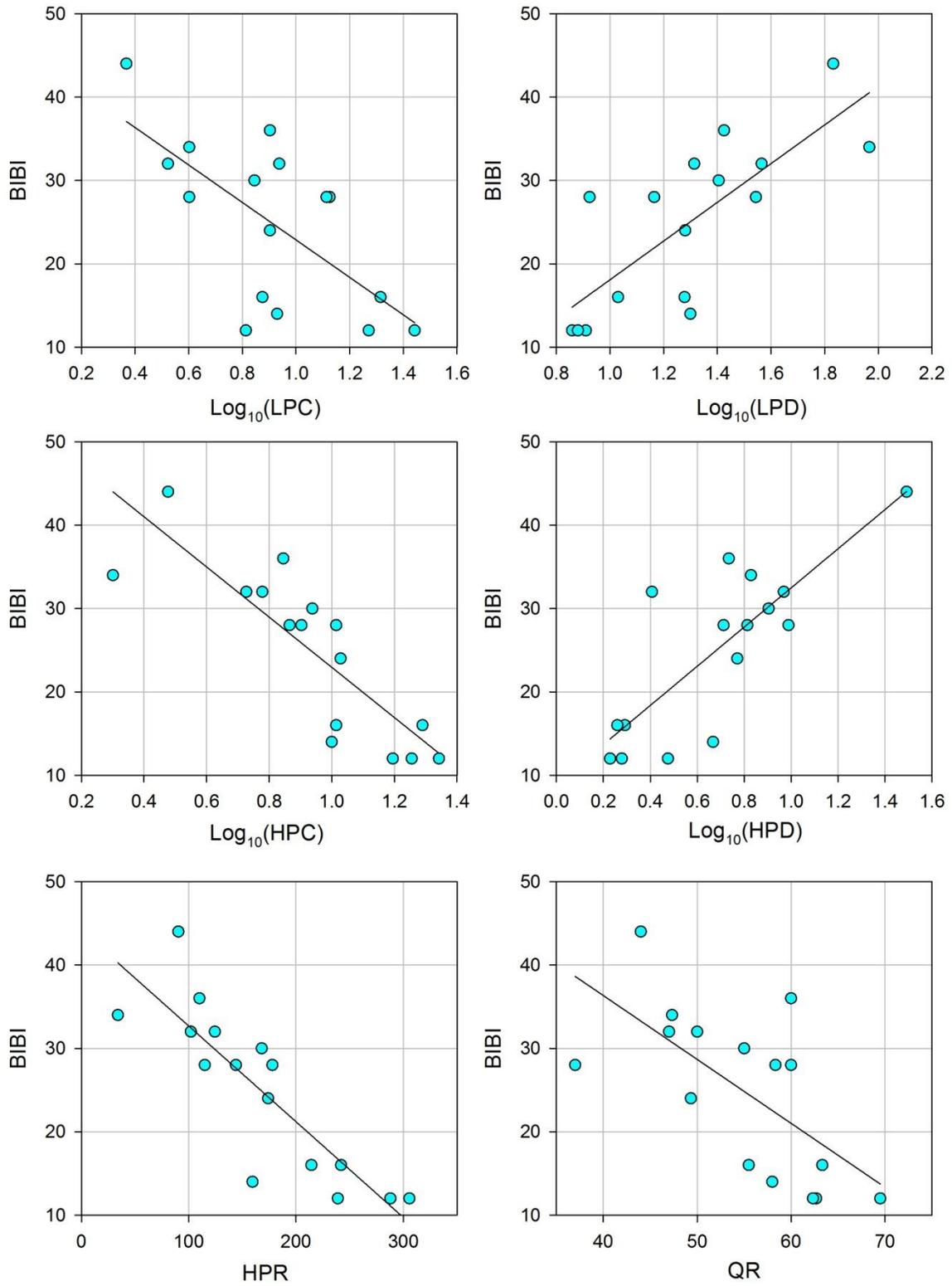


Figure 9 Scatter plots with regressions lines for LPC, LPD, HPC, HPD, HPR, and QR.

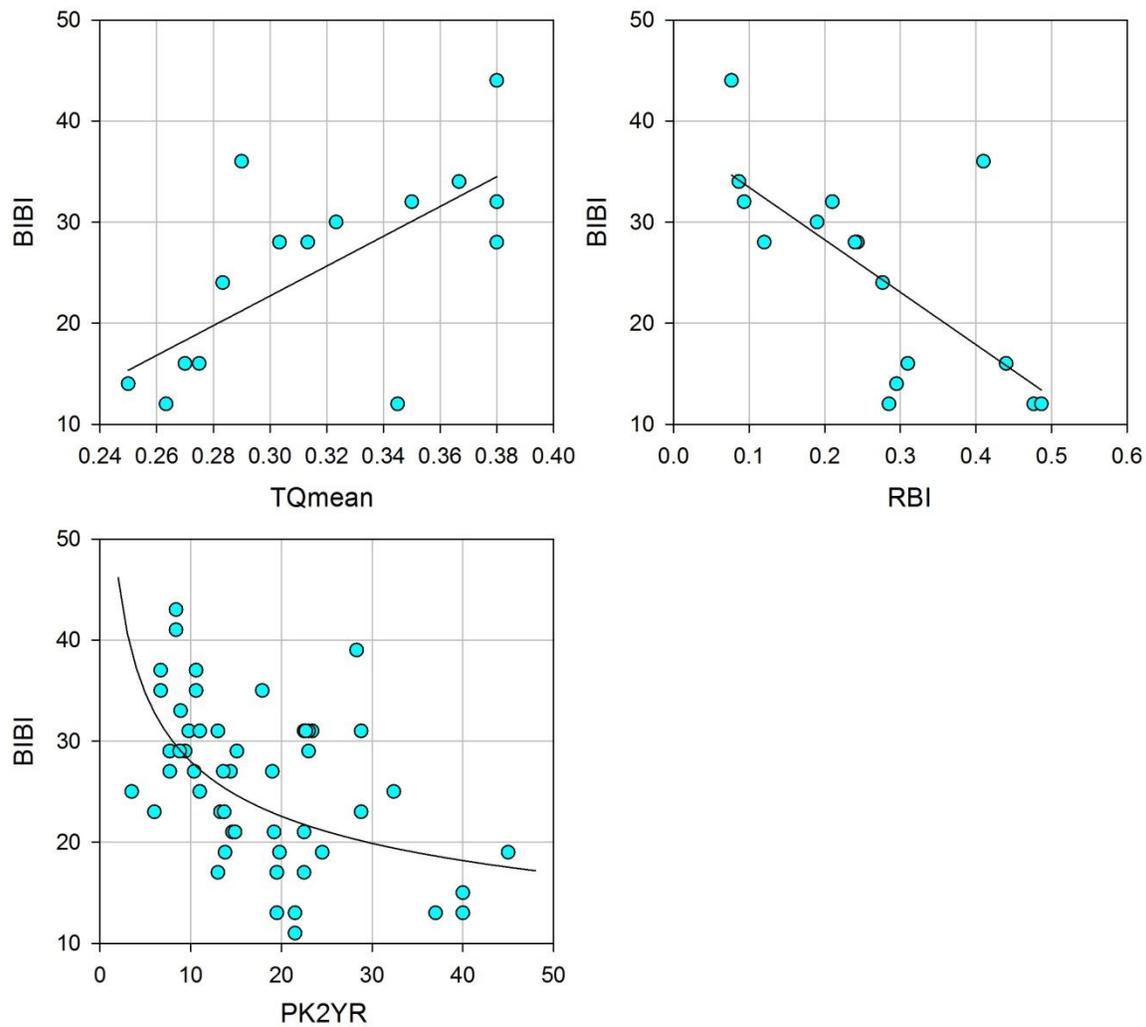


Figure 10 Scatter plots with regression lines for TQmean, RBI, and PK2YR

Simulated scores representing existing conditions (i.e. 2002 land use), on average, were three points higher (BIBI = 17) than averaged observed (BIBI = 14.4) over a similar time period (Table 10) at four locations within the Juanita Creek study area. This level of accuracy, with 18% relative error, is far greater than the identified accuracy among the defined relationships as previously shown in Figure 9 and Figure 10. Thus this application is judged adequate for use in this study (in conjunction with prior stated caveats in previous sections and in the main report this appendix is supporting).

Table 10 Simulated versus observed BIBI scores.

Site Code	Date Range	Observed	Simulated	RPD
		Mean	Mean	$\frac{(S - O)}{O} * 100$
JuanitaKirk1	2002 - 2008	14	16	14%
JuanitaKirk2	2005 - 2008	14	16	14%

JuanitaKirk3	2002 - 2008	13	18	38%
JuanitaKirk4	2002 - 2008	15	18	20%
E1186	2006 - 2009	16	17	6%
Average		14.4	17.0	18%

5.0. WATER QUALITY CALIBRATION

The hydrologic model was calibrated for the following water quality parameters

- Water temperature
- Total suspended solids (TSS)
- Dissolved oxygen (DO)
- Benthic algae
- Fecal coliforms
- Total copper
- Dissolved copper
- Ammonia nitrogen
- Total nitrogen
- Ortho phosphorus
- Total phosphorus

Each parameter was specified whether its loadings are primarily a function of sediment, surface runoff, subsurface (interflow and/or groundwater) runoff, or a combination. For the impervious land surfaces, only sediment loads and surface runoff are used to determine the loadings. All constituent loadings except fecal coliforms are assigned units of pounds. The mass units for fecal coliforms are 10^9 CFUs (colony forming units or organisms).

5.1 Water Temperature

Water Temperature is modeled by performing an energy balance in each stream segment. Heat and energy inputs to the stream are determined from the temperature of nonpoint, point, and boundary inflows; and from meteorological data

Model accuracy for water temperature is assessed by comparing hourly model output to discrete in-situ measurements for calibration points (2G, 3G, 4GO, and 4GI) and continuous measurements recorded at (1G, 5G, 6G, and 7G). Juanita Creek mainstem (1G) has a longer period of record available and was used for this calibration (2007 – 2010). Final calibration shows the model under simulating water temperatures between 2 and 4 degrees Fahrenheit which is approximately 3 to 9 percent error for all but one location (4GO). The wetland outlet was slightly over simulated averaging 2% above observed. The variance is well characterized at all eight locations (Table 11) as conveyed using r-square statistics.

Table 11 Summary statistics for calibrated water temperature.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Water Temperature	RMSE	3.31	3.64	2.78	2.62	3.26	3.99	5.40	2.75
	ME	-2.78	-2.53	-1.85	1.32	-2.74	-3.13	-4.65	-1.76

Parameter (deg-F)	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
	RPD	-5%	-5%	-4%	2%	-5%	-6%	-9%	-3%
	r-square	0.95	0.94	0.95	0.92	0.96	0.92	0.92	0.91

Visual representation allows for the added interpretation of model accuracy not expressed in the statistics above. In Figure 11 below, a set of four graphs representing Juanita Creek mainstem (1G) are illustrated summarizing a time series plot of water temperature (upper left), scatter plot (upper right), cumulative distribution (lower left) and for reference the time series of stream flow for observed and simulated (lower right). Observed are shown in red while simulated in blue. As revealed in the statistics above, the mainstem is generally under simulated by a few percent and is most visible in the cumulative distribution plot (lower left), but overall tracks exceptionally well over multiple seasons and years.

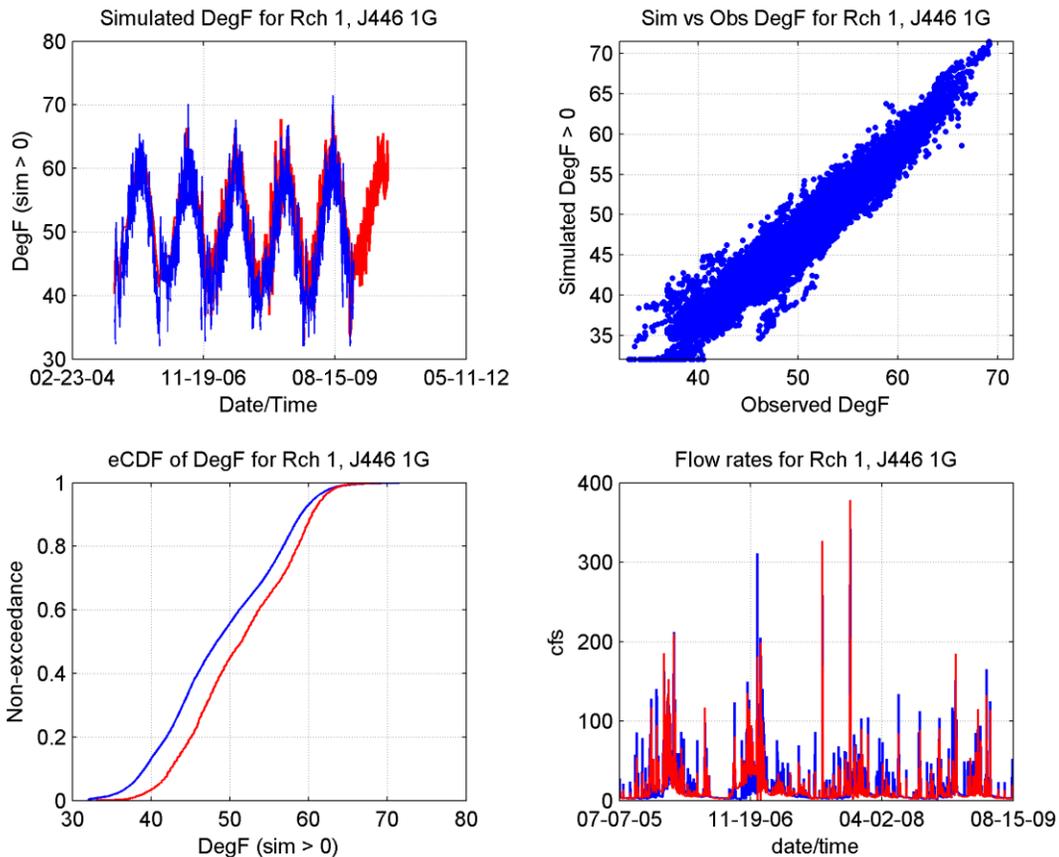


Figure 11 Time series, scatter plot, and cumulative distribution plots of calibrated water temperature for the mainstem. Red is observed, blue is simulated.

5.2 Total Suspended Solids

Sediment is simulated in-stream as three separate size fractions, referred to as sand, silt, and clay. The sediment loadings (in units of tons) are generated using the surface storage and surface runoff results from the hydrologic simulation.

Simulating TSS does not take into account any episodic events that are discrete in nature (e.g. bank failure) and not easily predictable. The goal for TSS calibration is to reasonably simulate annual mass loadings. However, since copper and phosphorus are modeled as adsorbed to solids, characterizing instantaneous concentrations is also important. Statistical comparison of simulated and actual TSS concentrations is presented in Table 12. During parameter adjustment, the mainstem was the focus for minimizing error and improving model predictions.

Table 12 Summary statistics for calibrated total suspended solids.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Total Suspended Solids (mg/L)	RMSE	67.95	358.76	47.81	22.92	11.80	188.76	65.34	67.97
	ME	-24.87	-73.73	10.05	13.99	6.27	-98.77	-16.19	-12.88
	RPD	-43%	-50%	48%	240%	66%	-71%	-31%	-24%
	r-square	0.66	0.04	0.00	0.02	0.36	0.11	0.34	0.06

Simulated mainstem concentrations of suspended solids compare well for most of the observed conditions (approximately 70% of observed concentrations are 10 mg/L or below). Between 10 and 300 mg/L, the model under-simulates concentrations. One large event was sampled, and the model generally reflects a similar magnitude in concentrations (Figure 12).

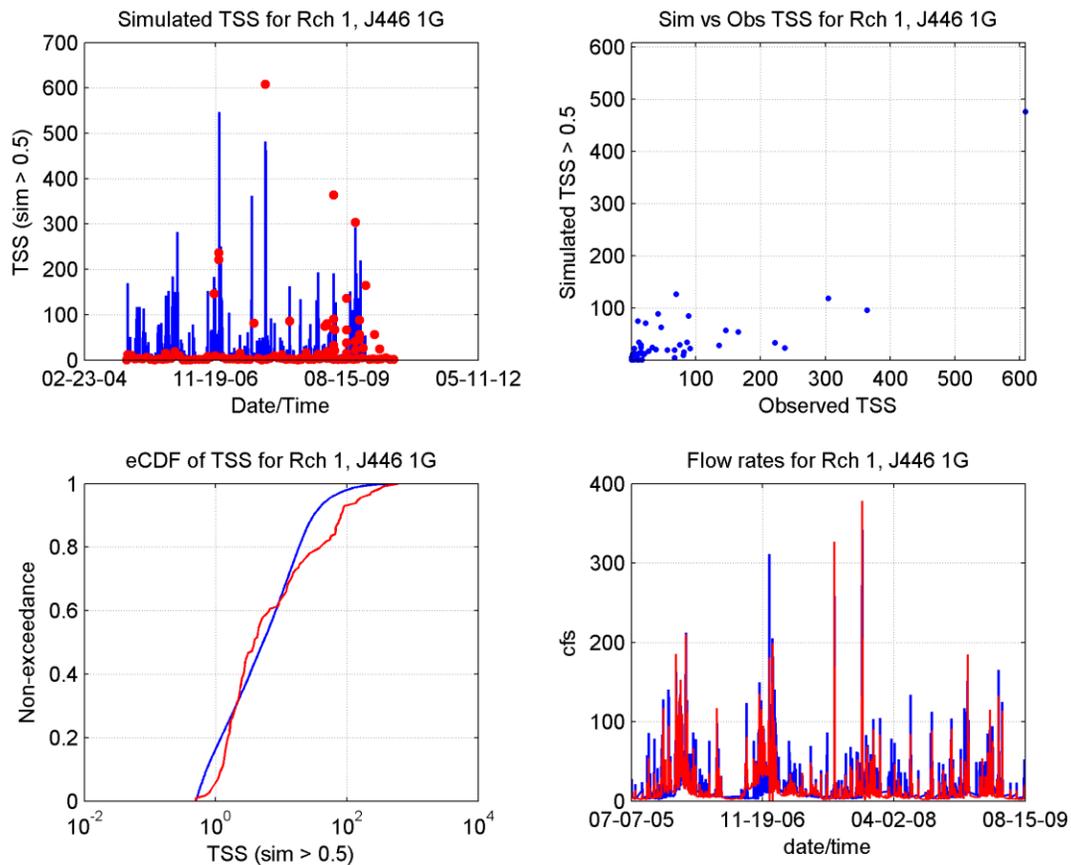


Figure 12 Time series, scatter plot, and cumulative distribution plots of calibrated total suspended solids for the mainstem. Red is observed, blue is simulated.

As reported in Table 12 above, some of the calibration points have statistically poor model accuracy of simulated instantaneous concentrations. However, those statistics are largely affected by observed concentrations most likely a result of some discrete release of fines not included in the model design or calibration, and/or due to very few data points affected by a few outliers. The large observed concentration may be a bank failure, or possible flushing of catch basins in the storm drainage network; either way those source mechanisms are not represented in the model. As an example using Billy Creek (2G), the four graphs in Figure 13 reflect variable concentrations over the period of available data except for the before mentioned condition. Except for Totem Lake tributary (i.e. 3G and 4GO), the other calibration points were similar in results as presented in above Figure 12 and Figure 13 below. Comparing TSS concentrations in the Totem lake tributary again show poor explanation of the variance among instantaneous observed concentrations, but do reflect the general variability of concentrations over time as shown in the time series graph and cumulative distribution graph in Figure 14.

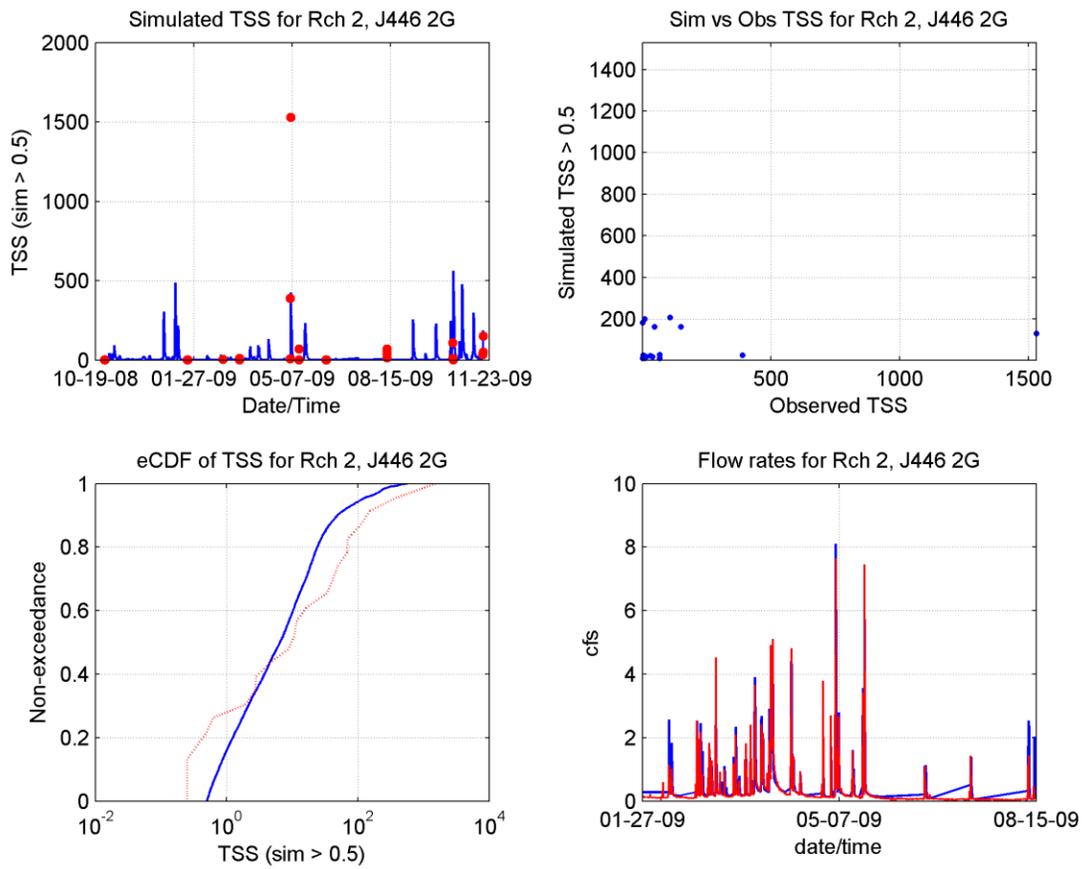


Figure 13 Time series, scatter plot, and cumulative distribution plots of calibrated total suspended solids for Billy Creek (2G). Red is observed, blue is simulated.

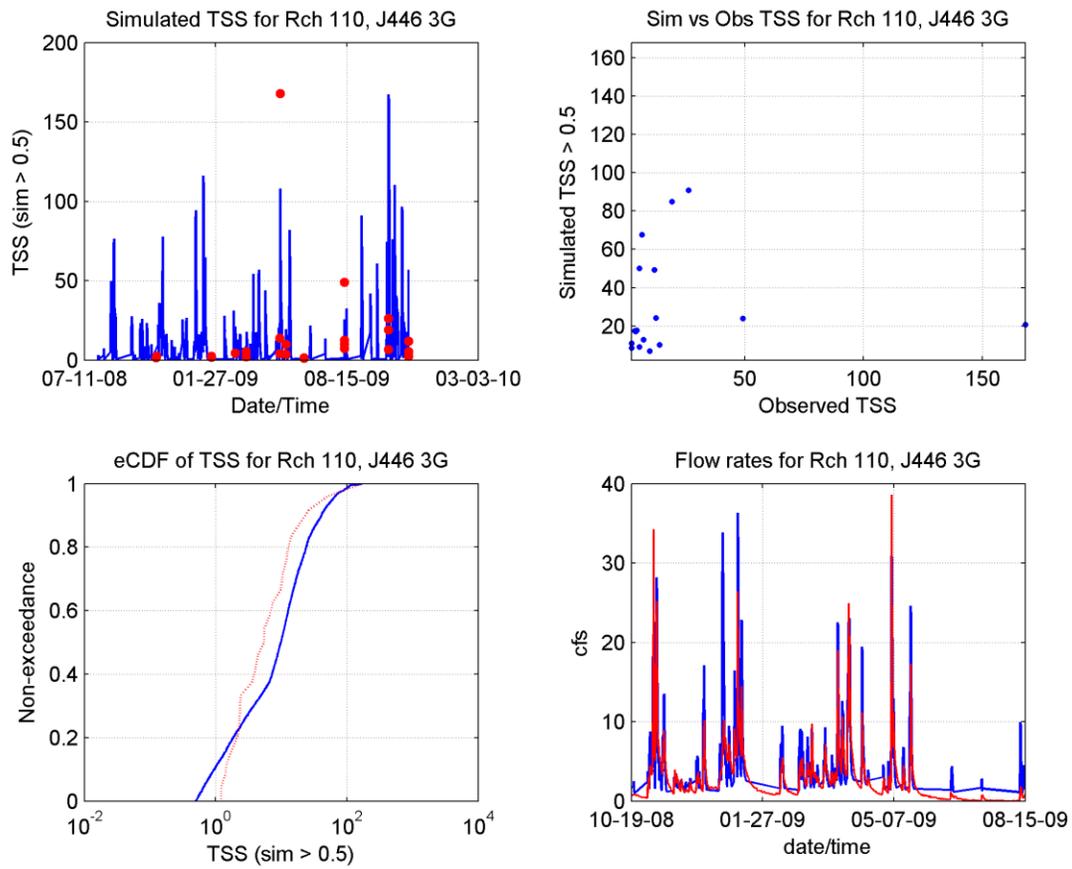


Figure 14 Time series, scatter plot, and cumulative distribution plots of calibrated total suspended solids for Totem Lake tributary (3G). Red is observed, blue is simulated.

5.3 Dissolved Oxygen

Dissolved Oxygen is simulated in the stream by defining the oxygen loads in nonpoint and point runoff, and representing reaeration and biological/chemical processes in the stream. As shown in Table 13, model accuracy has good agreement among calibration points that are more representative of faster moving stream reaches with r-squares ranging from 0.59 to 0.79 (with the exception of Billy Creek—2G), and less accurate for the Totem Lake subbasin (r-squares of 0.42). Billy Creek (2G) dissolved oxygen was measured in a plunge pool below a 3-ft drop out of a culvert. This supersaturated condition of air entrainment in the water column is not represented in the model design or calibration process.

Table 13 Summary statistics for calibrated dissolved oxygen.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Dissolved Oxygen (mg/L)	RMSE	0.82	1.55	1.27	3.58	1.59	1.51	1.79	0.71
	ME	-0.45	-1.22	-0.11	2.97	1.14	-1.26	-1.57	-0.06
	RPD	-4%	-10%	-1%	70%	14%	-11%	-14%	-1%
	r-square	0.69	0.43	0.42	0.42	0.79	0.71	0.59	0.59

Further evaluation is illustrated using the same four types of graphs in the following two figures, Figure 15 representing mainstem conditions (1G) and, as an example, illustrating the least accurate calibration location (3G) downstream of Totem Lake and a large wetland (Figure 16). Contrary to the lesser r-square, the model does reflect the seasonal fluctuations of dissolved oxygen over the period of available data except for the late fall/early winter in 2009 where observed concentrations are substantially lower than the previous year.

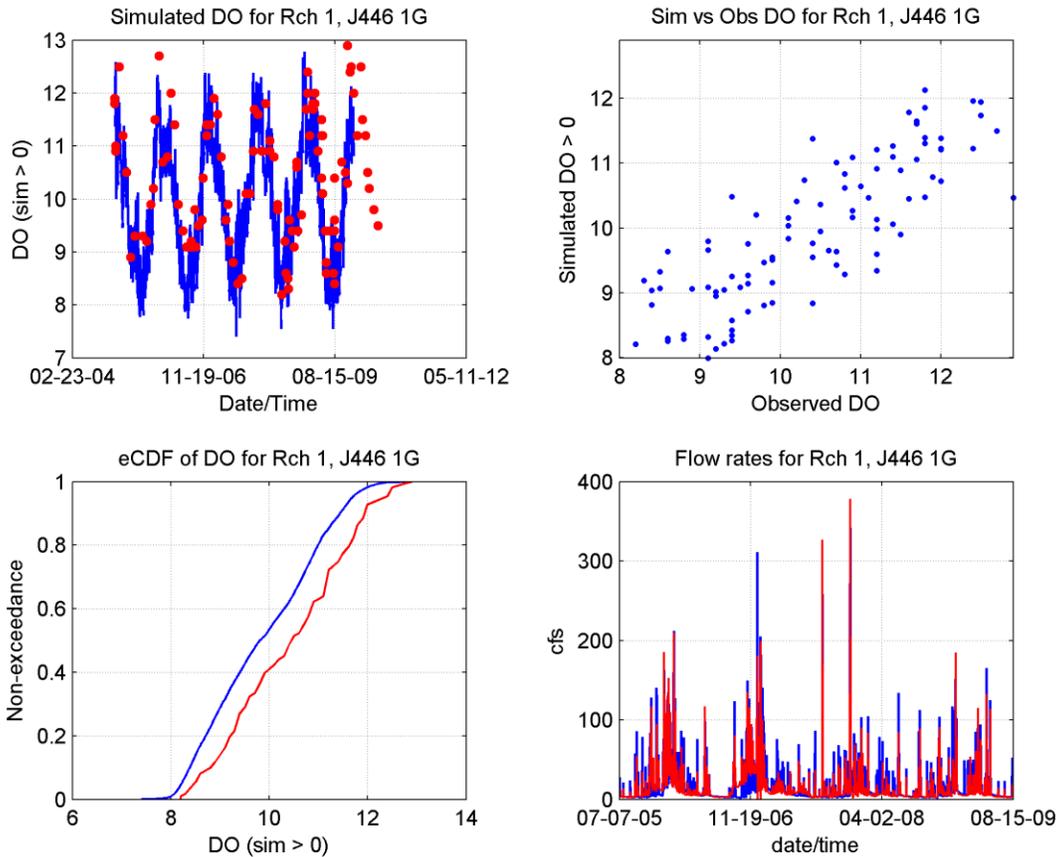


Figure 15 Time series, scatter plot, and cumulative distribution plots of calibrated dissolved oxygen concentrations for the mainstem. Red is observed and blue is simulated.

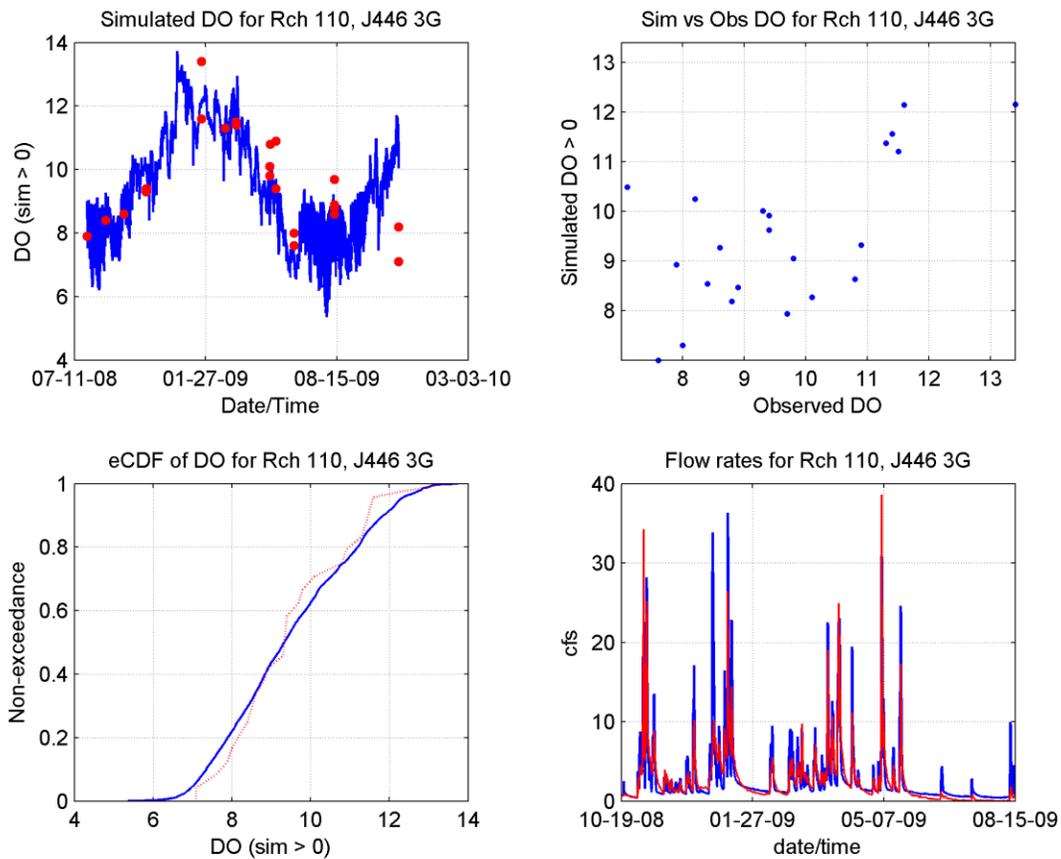


Figure 16 Time series, scatter plot, and cumulative distribution plots of calibrated dissolved oxygen concentrations for the Totem Lake tributary (3G). Red is observed and blue is simulated.

5.4 Benthic Algae

Benthic Algae (or periphyton) is stationary, living organic matter that is modeled using the same Monod growth kinetics that is used to simulate phytoplankton. The material grows, respire, and dies in response to light, nutrient (N and P) availability, and temperature. It takes up and releases nutrients and releases organic matter (detritus) during respiration, sloughing, and death. To adequately characterize the diurnal swings in dissolved oxygen, benthic algae densities, growth and death rates were adjusted sometimes specific to a given catchment reach containing ponds, lakes, and/or wetlands; otherwise the values were consistent among all reaches. However, no data was collected for this parameter to evaluate model accuracy.

5.5 Fecal Coliforms

Fecal coliform concentrations are extremely variable and difficult to predict. One reason for this is that many of the larger loadings of bacterial material probably occur not only during storms, but also during somewhat random but “catastrophic” events, such as failure or illicit sewer connections of waste disposal facilities, which can produce large, unpredictable concentrations. Therefore, efforts were made to attain general agreement between the simulated concentrations by adjusting loading rates, both surface and subsurface runoff-associated by land use. Because of the difficulty in matching actual observed values, the explanatory regression coefficient (i.e. r-square) is used more as guidance than a test of acceptability but still necessary for evaluation given metrics used in scenario analyses are dependent on absolute thresholds of concentrations. Due to the high concentrations and variability, calibrated loading rates for this study should not be used for any other basin.

Model accuracy simulating fecal concentrations is substantially less than the other parameters receiving similar scrutiny. The variance in regressions (r-square) range from 0.0 to 0.26 (Table 14). Only two locations were above r-squares of 0.10, mainstem (1G) and inlet to the large wetland (4GI). Focusing on the mainstem (1G) and the two least accurate tributary calibrations (4GO and 5G), the four graphs in each figure illustrate relevancy comparing instantaneous observed concentrations to simulated over a period of record. Figure 17 illustrates the quality of simulation for the mainstem of Juanita Creek with concentrations within the same order of magnitude except for a large simulated spike early in the model time span. The two least accurate calibrations (4GO and 5G) with r-squares equal to 0.0 visually compare well to measured except during the largest observed concentrations in excess of 4000 cfu/100ml in both cases (Figure 18 and Figure 19). The west branch tributary was observed to have two high concentration events, both substantially under simulated. The other headwater tributary (6G) was similar to 5G in response and (2G, 3G, 4GI, and 7G) were more characteristic of 4GO accuracy.

Table 14 Summary statistics for calibrated fecal coliforms.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Fecals (cfu/100 ml)	RMSE	1500	2466	724	1508	910	6265	5815	2274
	ME	-402	-607	101	232	-283	-2405	-2784	-517
	RPD	-49%	-40%	23%	33%	-46%	-70%	-76%	-37%
	r-square	0.18	0.02	0.08	0.00	0.26	0.00	0.01	0.01

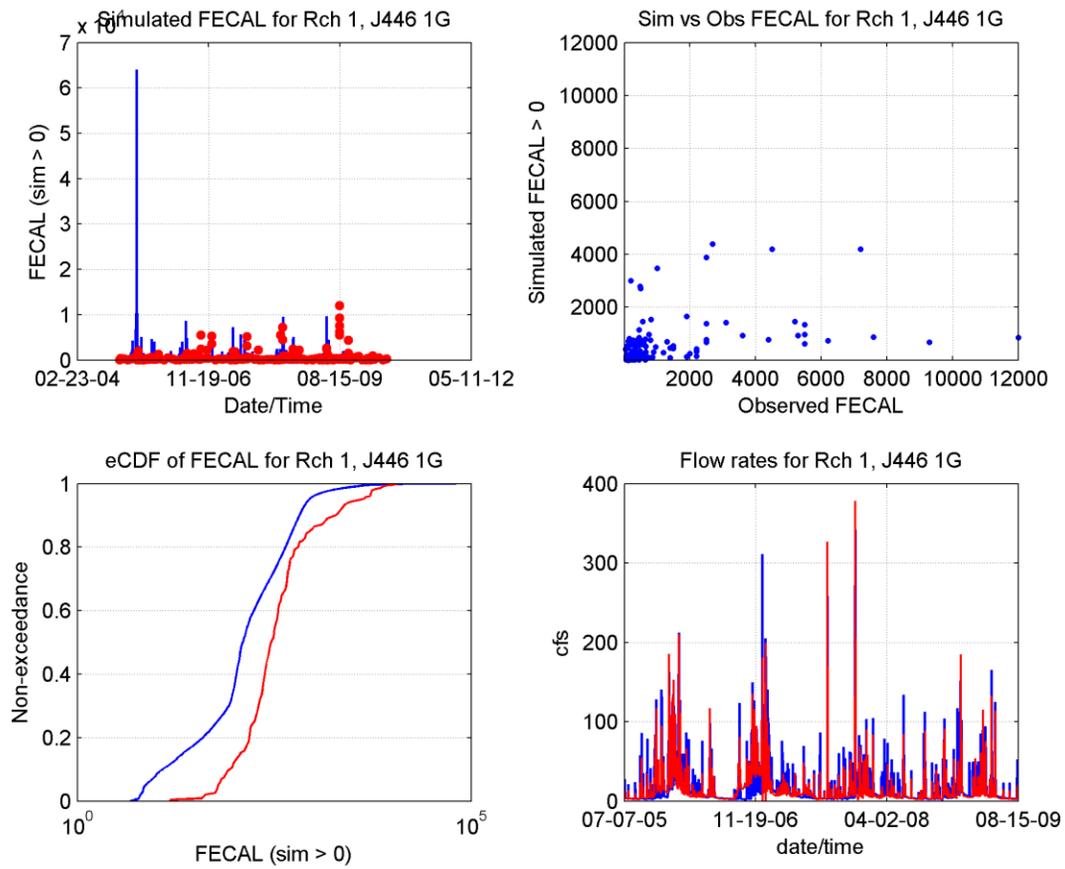


Figure 17 Time series, scatter plot, and cumulative distribution plots of calibrated fecal colony forming units for the mainstem. Red is observed, blue is simulated.

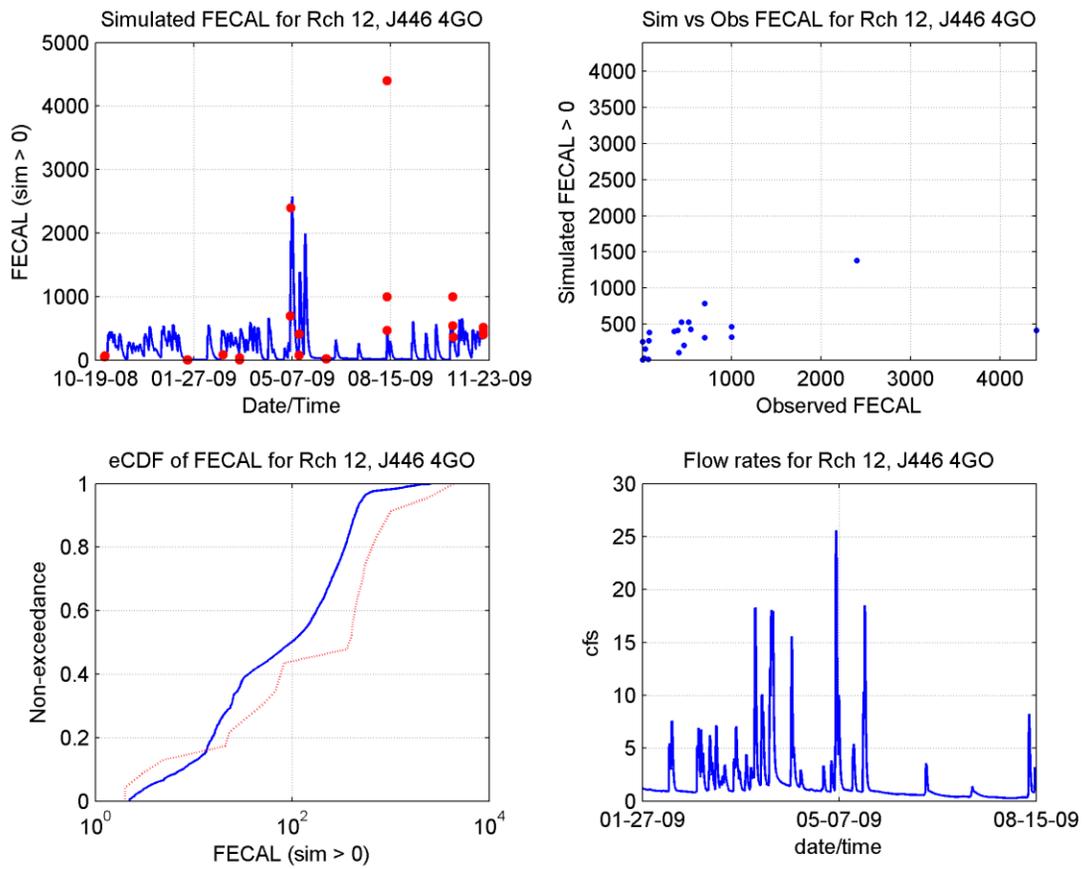


Figure 18 Time series, scatter plot, and cumulative distribution plots of calibrated fecal colony forming units for the outlet of wetland. Red is observed, blue is simulated.

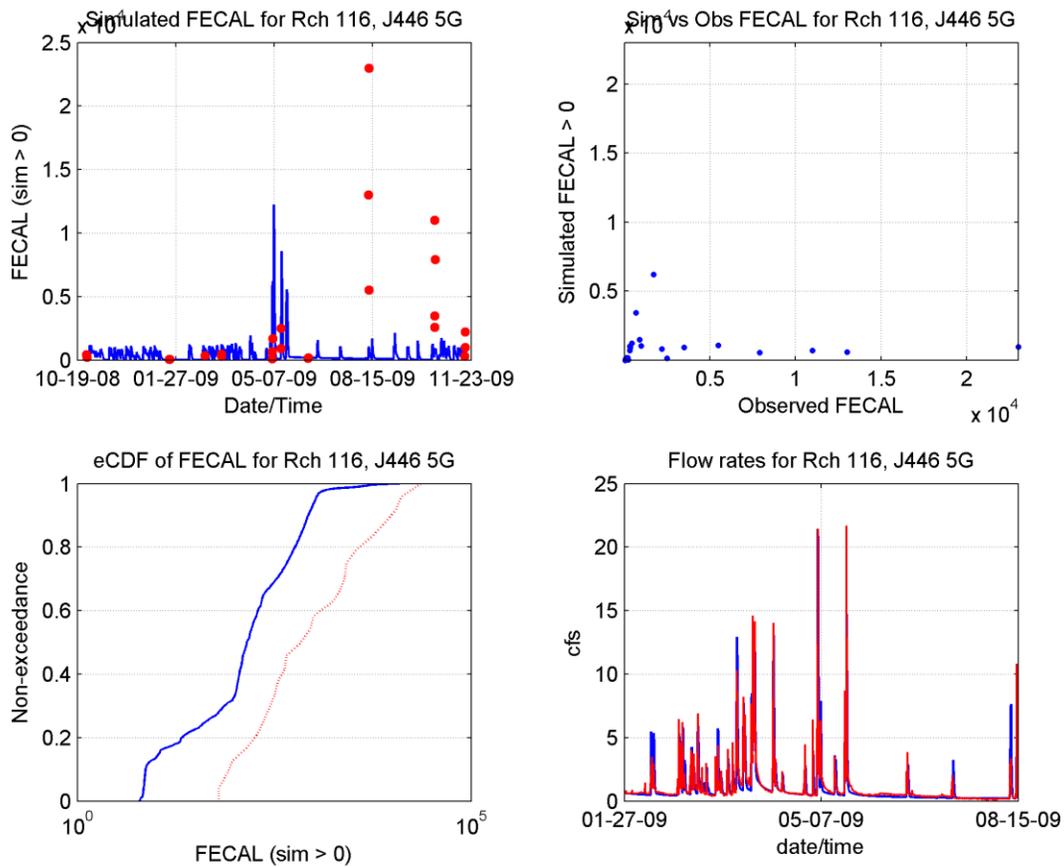


Figure 19 Time series, scatter plot, and cumulative distribution plots of calibrated fecal colony forming units for the west headwater branch tributary. Red is observed, blue is simulated.

5.6 Total Copper

Total copper concentrations were calibrated by adjusting the land use-specific interflow and groundwater concentrations and the surface parameters (potency factors) to achieve a statistical fit with the available data. Copper is sediment-associated, so all surface loading was modeled in the sorbed (i.e. attached) phase.

Total copper r-square values range from .02 to .52 with the wetland outlet (4G0) and north branch (7G) having the lowest r-squares of .02 and .14, respectively (Table 15). However, reviewing the four types of graphs in Figure 20, the model’s ability to simulate total copper is better than statistically reported due to substantially under simulated concentrations during one suspiciously large observed event—the Totem lake tributary (3G) r-square value was similarly affected.

Table 15 Summary statistics of calibration for total and dissolved copper.

Parameter	Statistic	Mainstem (1G)	Billy Creek	Totem Lake	Wetland outlet	Wetland inlet	West Branch	East Branch	North Branch

			(2G)	Trib. (3G)	(4GO)	(4GI)	(5G)	(6G)	(7G)
Total Copper (ug/L)	RMSE	10.18	8.00	3.25	6.74	1.71	7.62	8.12	8.08
	ME	0.00	2.10	2.13	1.86	0.94	-0.53	0.66	2.55
	RPD	0%	29%	100%	32%	50%	-6%	9%	38%
	r-square	0.45	0.52	0.26	0.02	0.42	0.35	0.20	0.14

While the wetland outlet concentrations were the least accurately simulated (r-square 0.02), the four graphs in Figure 21 illustrate accuracy seemingly better than reported. Furthermore, the majority of the error can be seen associated with the defined concentrations in the active groundwater component of the model (Figure 22), which is user specified. Adjusting those would require a separate set of parameterization counter to the overall model calibration method—no unique submodels.

The north branch calibration point (7G) shows a divergence between under and over simulating events (scatter plot in Figure 23). Further reviewing the time series, there appears to be a shift in timing between like magnitudes (Figure 24). The cause is unknown, but peculiar. The other simulated calibration points were similar to or better in model accuracy using visual inspection.

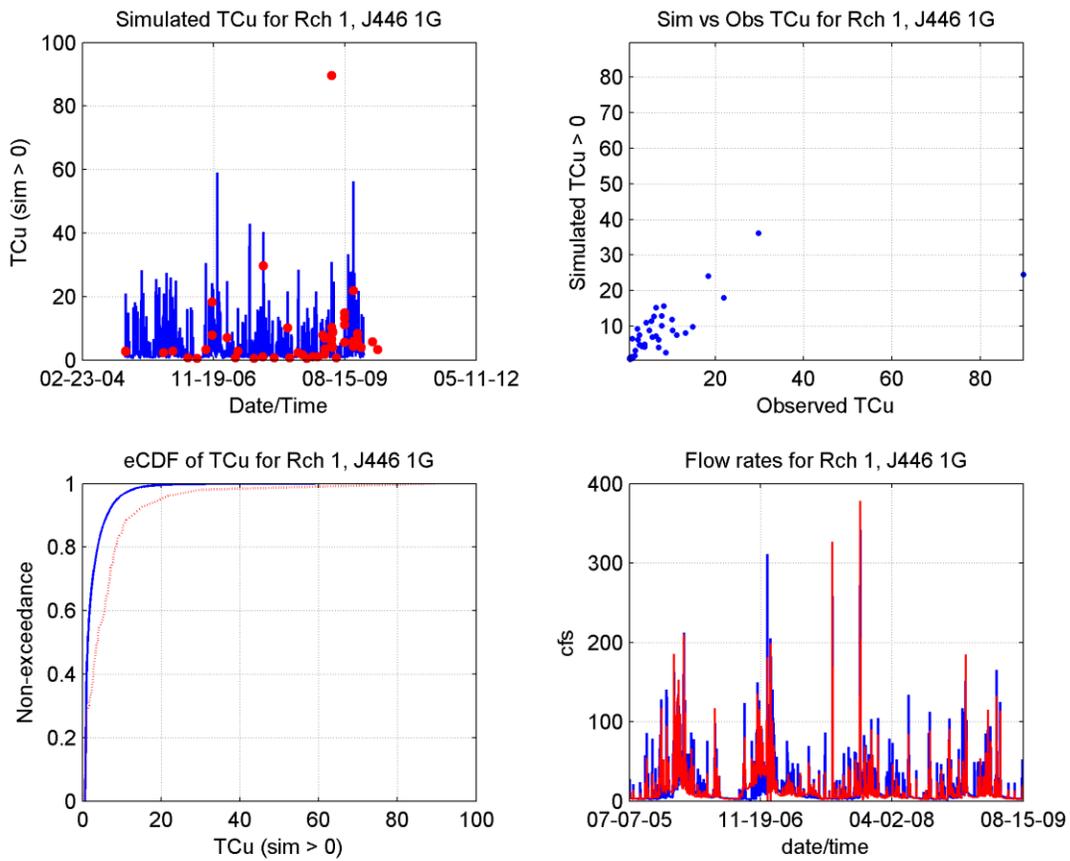


Figure 20 Time series, scatter plot, and cumulative distribution plots of calibrated total copper concentrations for the mainstem. Red colors are observed while blue is simulated.

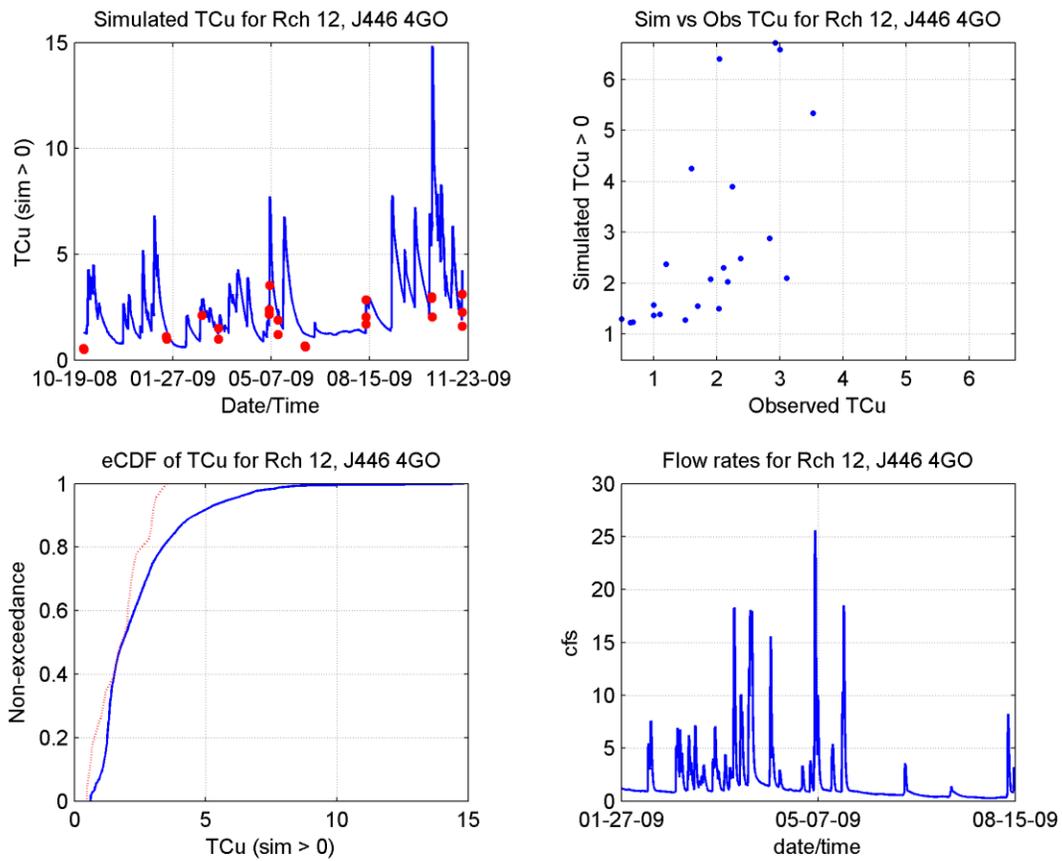


Figure 21 Time series, scatter plot, and cumulative distribution plots of calibrated total copper concentrations for the wetland outlet (4GO). Red colors are observed while blue is simulated. Stream flows were not recorded at this location.

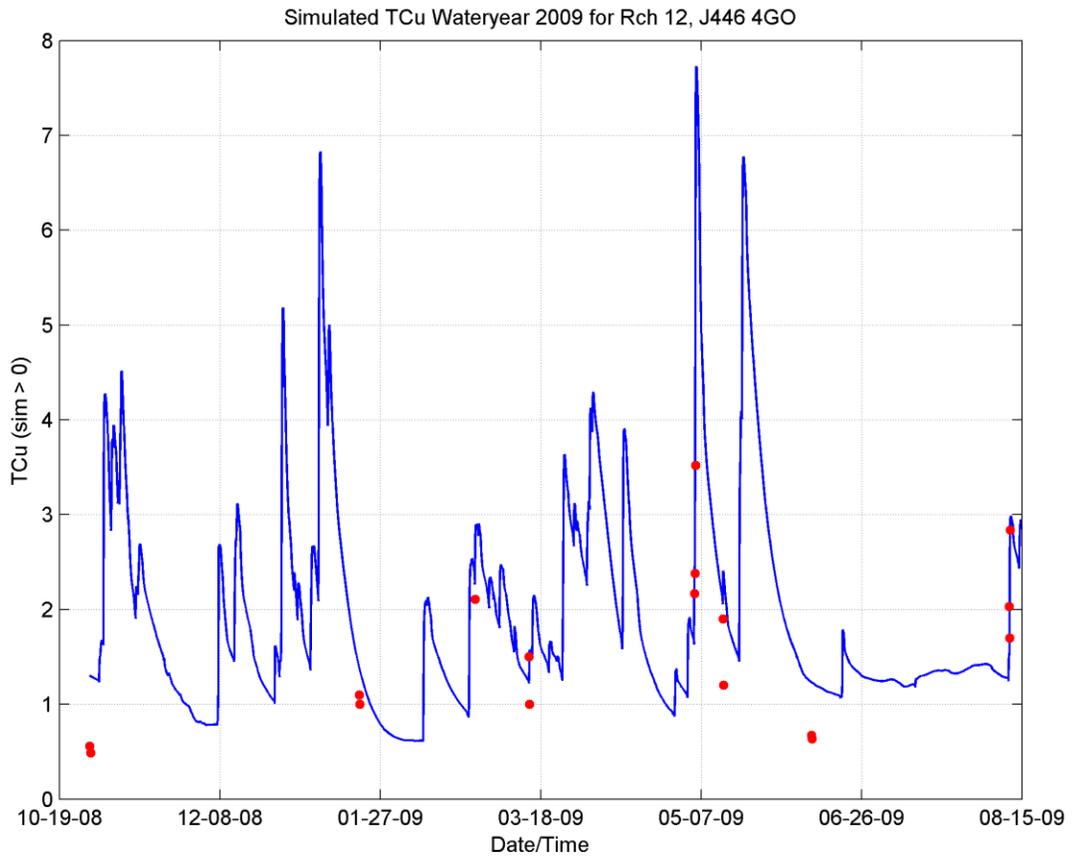


Figure 22 Time series plot of total copper concentrations for outlet of wetland (4GO). Red colors are observed, blue is simulated.

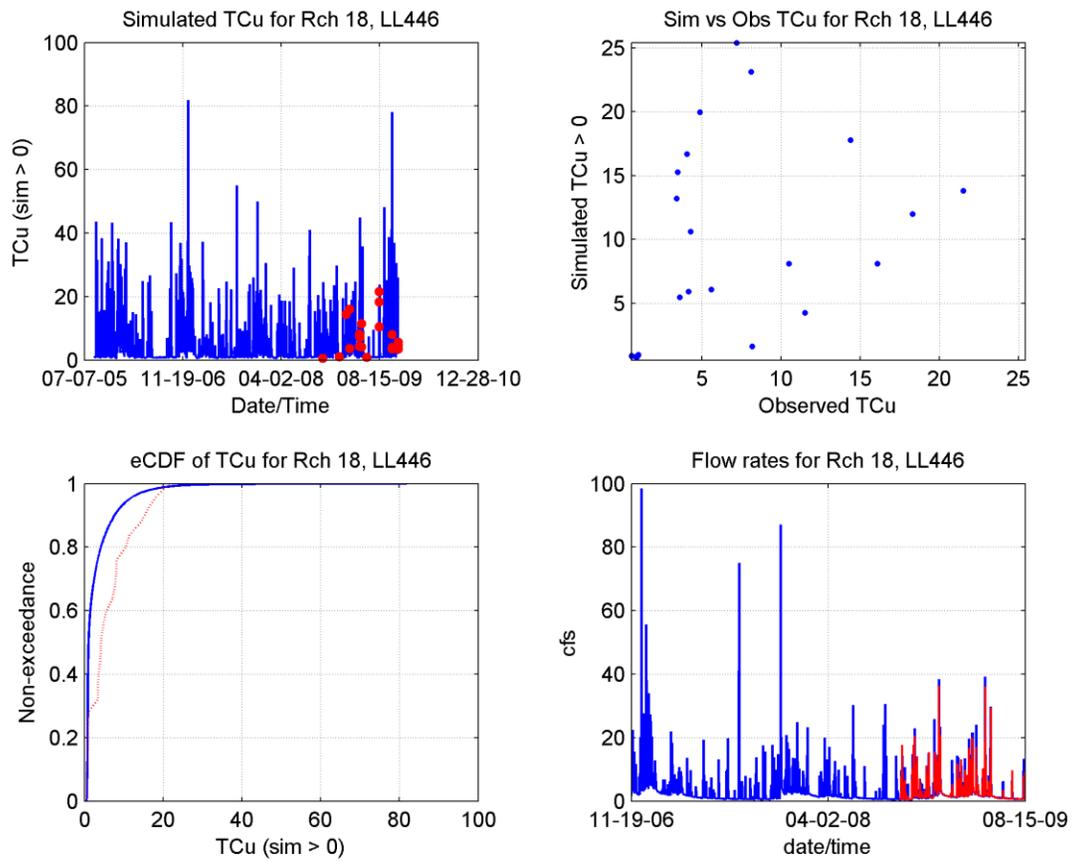


Figure 23 Time series, scatter plot, and cumulative distribution plots of calibrated total copper concentrations for the north branch (7G). Red colors are observed while blue is simulated.

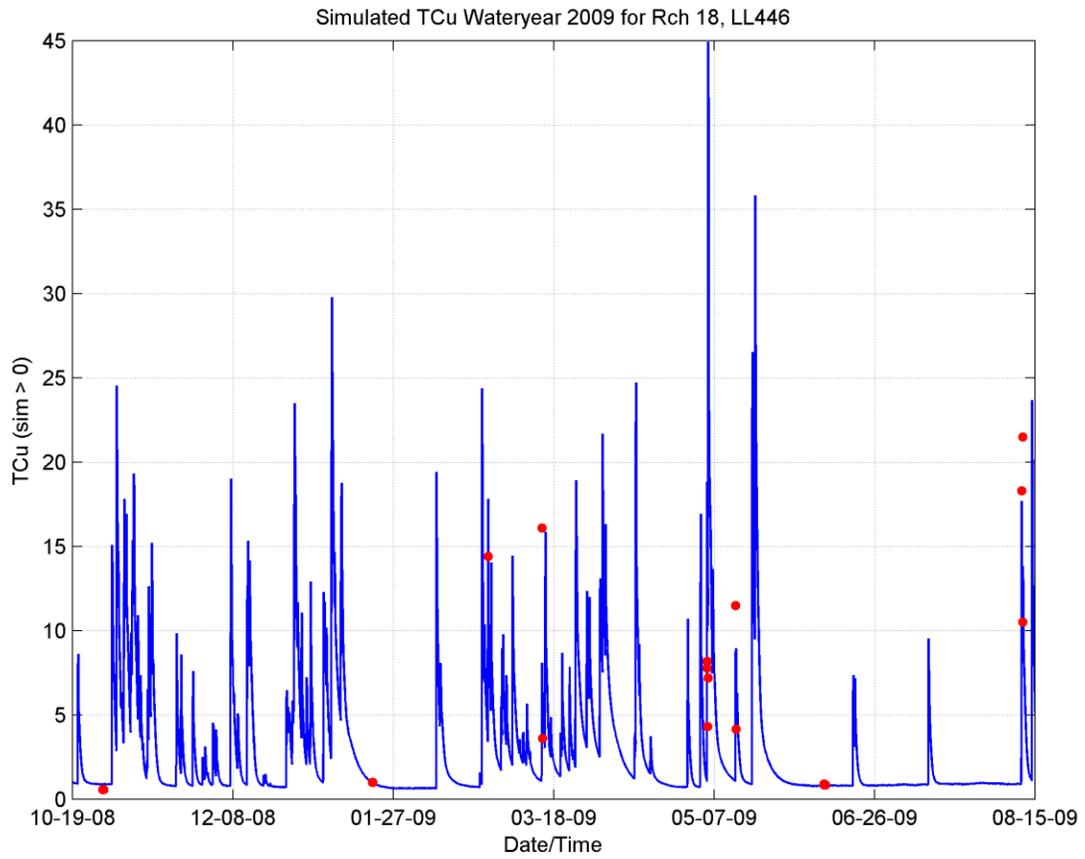


Figure 24 Time series plot of total copper concentrations for north branch tributary (7G). Red colors are observed, blue is simulated.

5.7 Dissolved Copper

The level of model accuracy is generally better modeling total copper as opposed to dissolved copper which is dependent on other time varying environmental factors such as hardness and concentration of suspended solids. Again, since metrics used to evaluate modeled scenarios relies on acute and chronic concentrations, the same higher level of statistical scrutiny is applied to simulated results on instantaneous concentrations.

Dissolved copper r-square values ranged from 0.01 to 0.49 (Table 13), where the 0.49 value here was higher than the total copper simulation for the same calibration point (0.26). Similar to the model capabilities of simulating total copper, dissolved copper has relatively poor statistical accuracy but overall reflects the variability in concentrations (Figure 25). However, visual inspection of time series plots (Figure 26 through Figure 33) illustrates comparable results between simulated and observed when allowing for some shifts in timing beyond the instantaneous data values used testing model accuracy, with one exception—wetland outflow concentrations at 4GO. Sequestering of dissolved copper in the wetland is not well characterized in the model as can be seen with persistent over simulating concentrations (Figure 29). This is likely due to sequestering from vegetative uptake not present in the fate/transport of the model design.

Table 16 Summary statistics of calibration for total and dissolved copper.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Dissolved Copper (ug/L)	RMSE	2.03	2.49	0.90	2.97	3.24	3.65	2.62	1.91
	ME	0.81	0.80	0.53	-0.70	2.57	0.27	-0.21	-0.46
	RPD	39%	36%	44%	-19%	189%	9%	-7%	-18%
	r-square	0.15	0.01	0.49	0.03	0.12	0.02	0.04	0.07

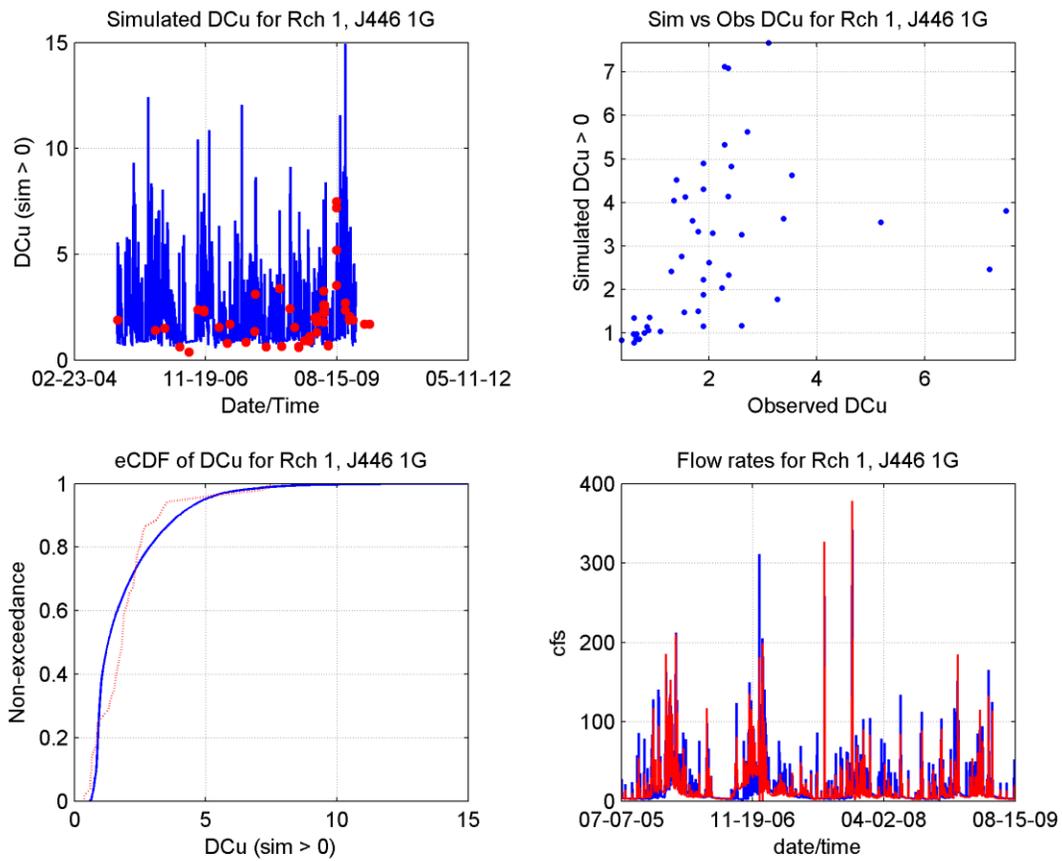


Figure 25 Time series, scatter plot, and cumulative distribution plots of calibrated for dissolved copper concentrations for the mainstem.

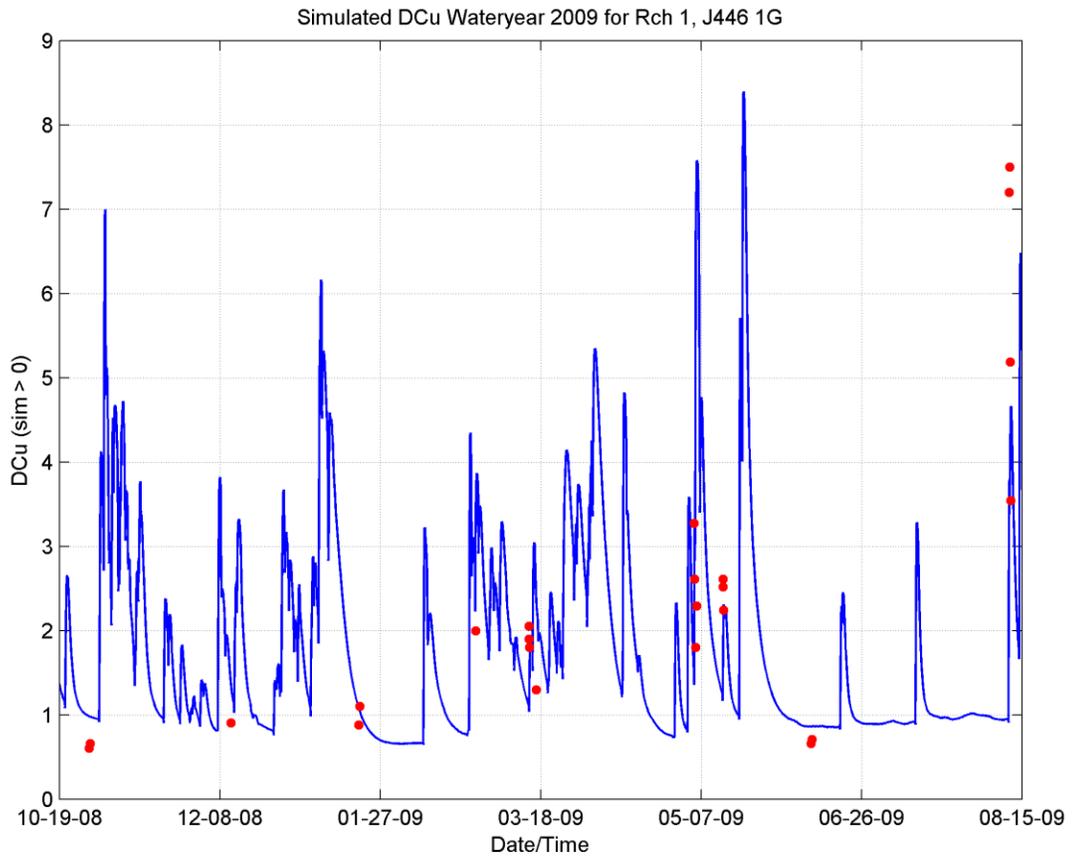


Figure 26 Time series plot of dissolved copper concentrations for mainstem (1G). Red colors are observed, blue is simulated.

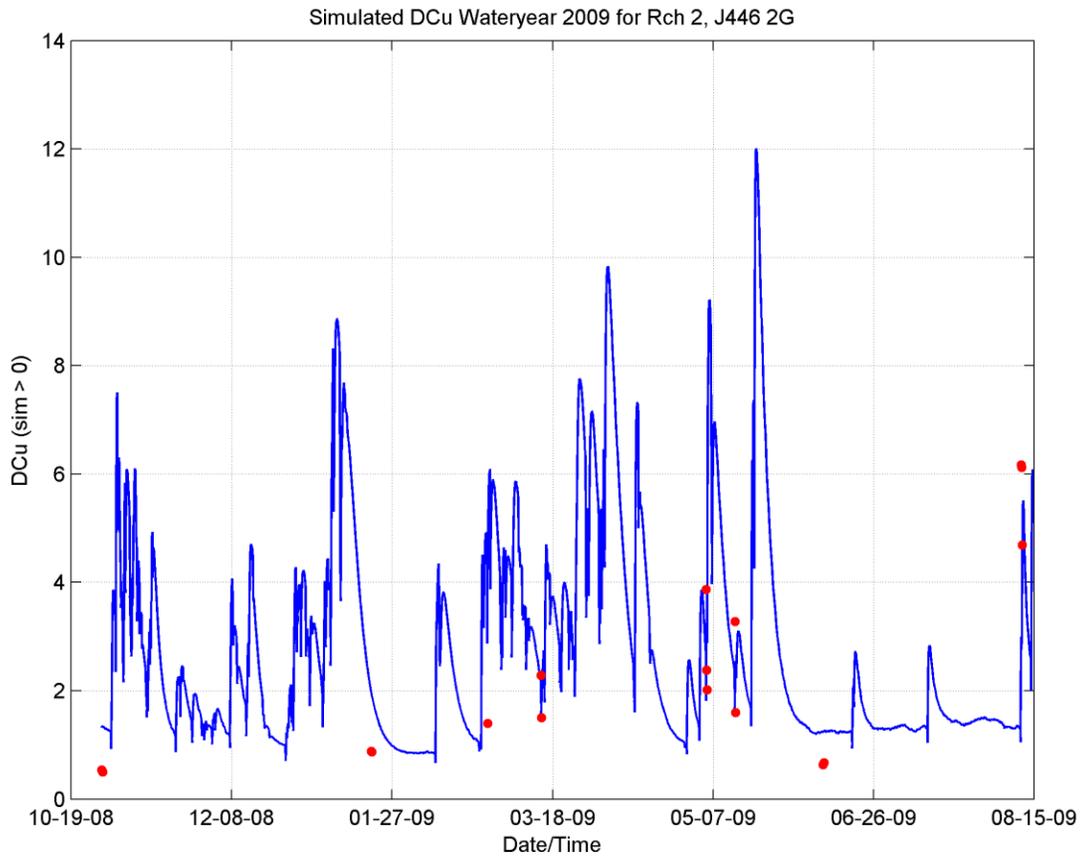


Figure 27 Time series plot of dissolved copper concentrations for Billy Creek (2G). Red colors are observed, blue is simulated.

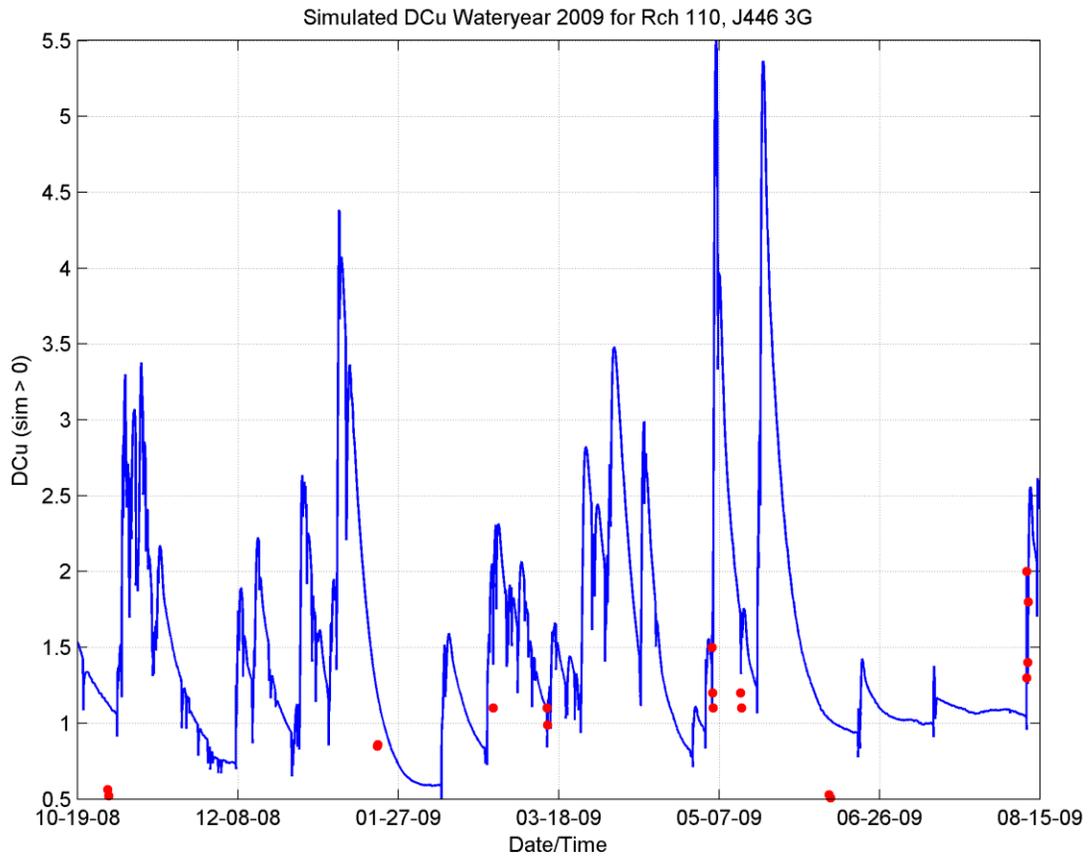


Figure 28 Time series plot of dissolved copper concentrations for Totem Lake tributary (3G). Red colors are observed, blue is simulated.

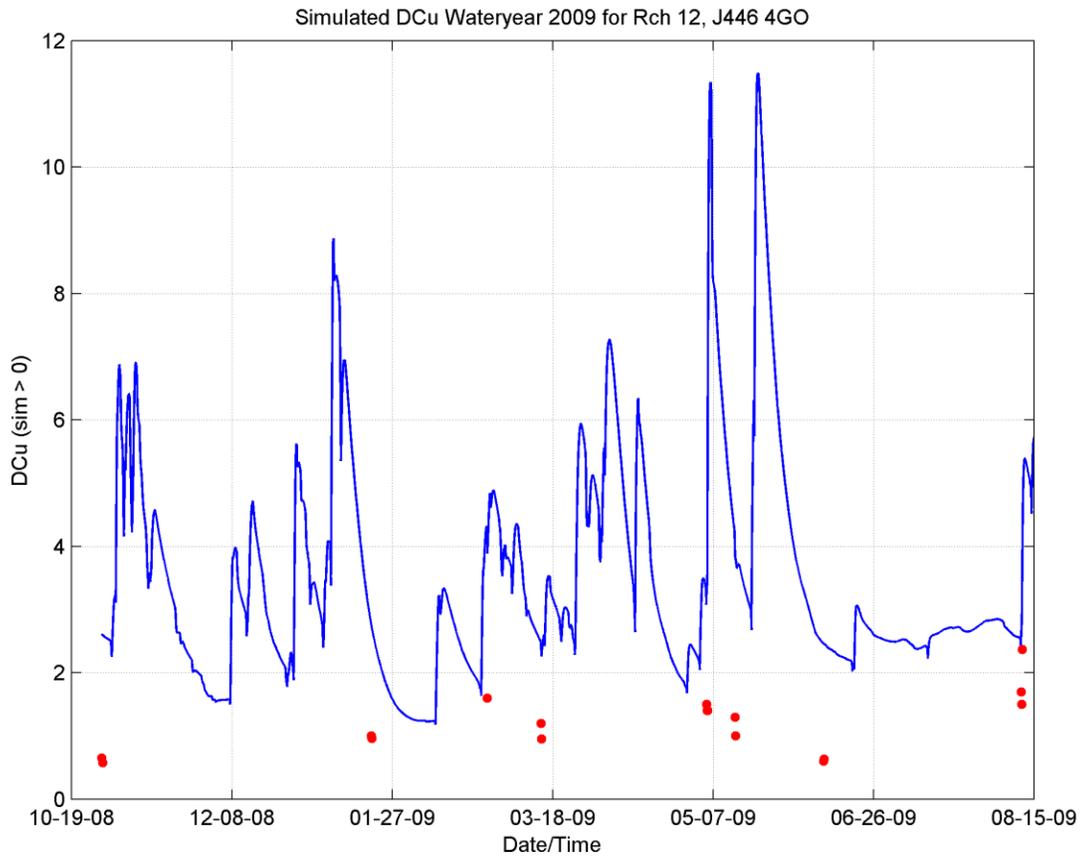


Figure 29 Time series plot of dissolved copper concentrations for outlet of wetland (4GO). Red colors are observed, blue is simulated.

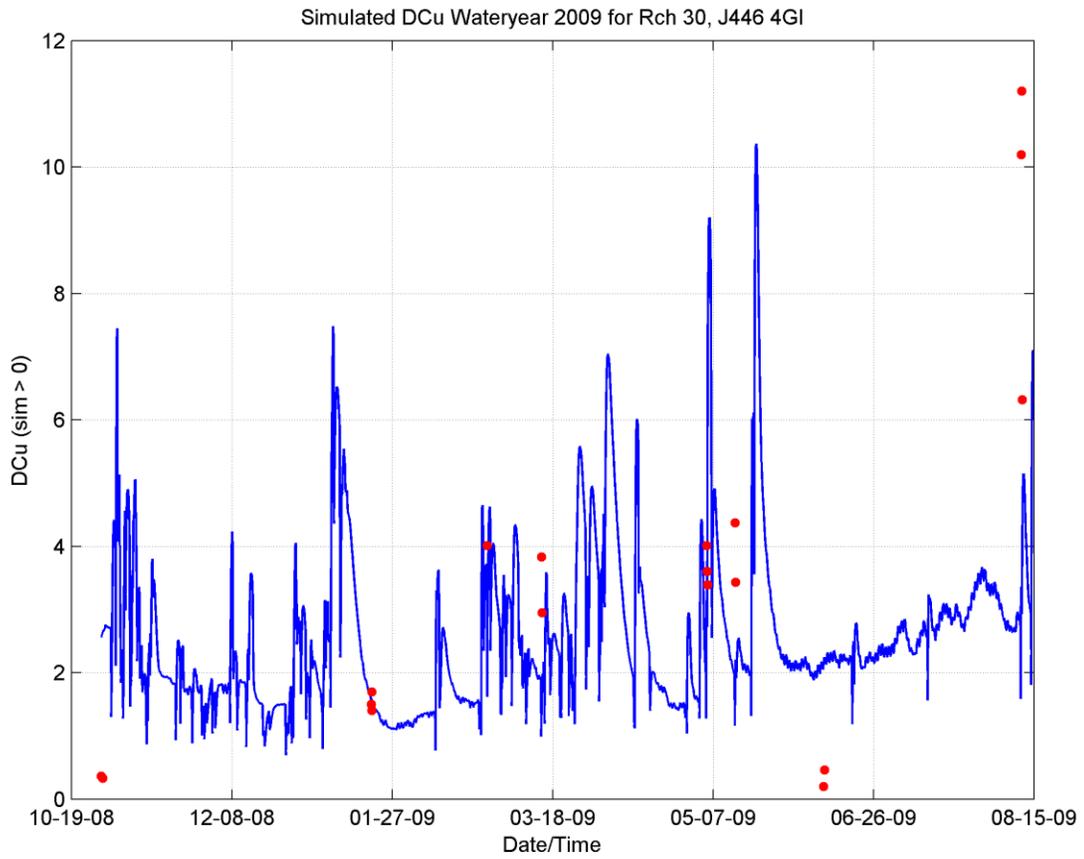


Figure 30 Time series plot of dissolved copper concentrations for inlet to wetland (4GI). Red colors are observed, blue is simulated.

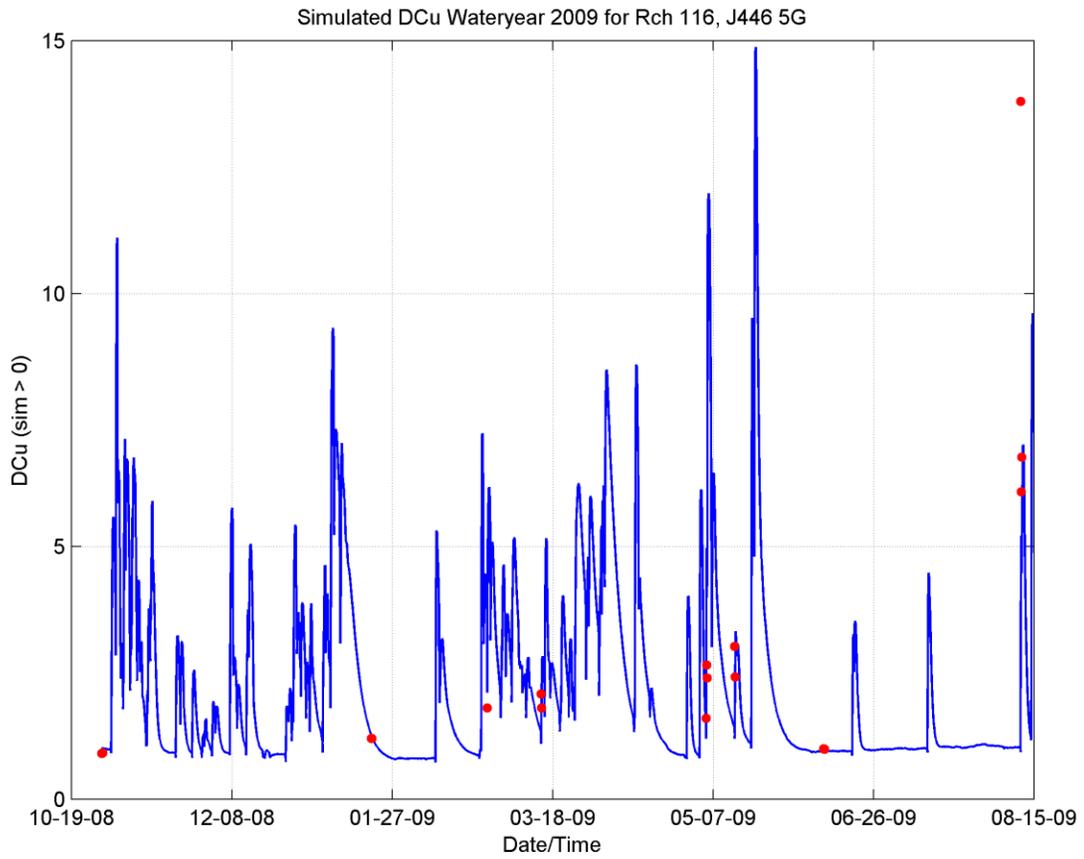


Figure 31 Time series plot of dissolved copper concentrations for west branch tributary (5G). Red colors are observed, blue is simulated.

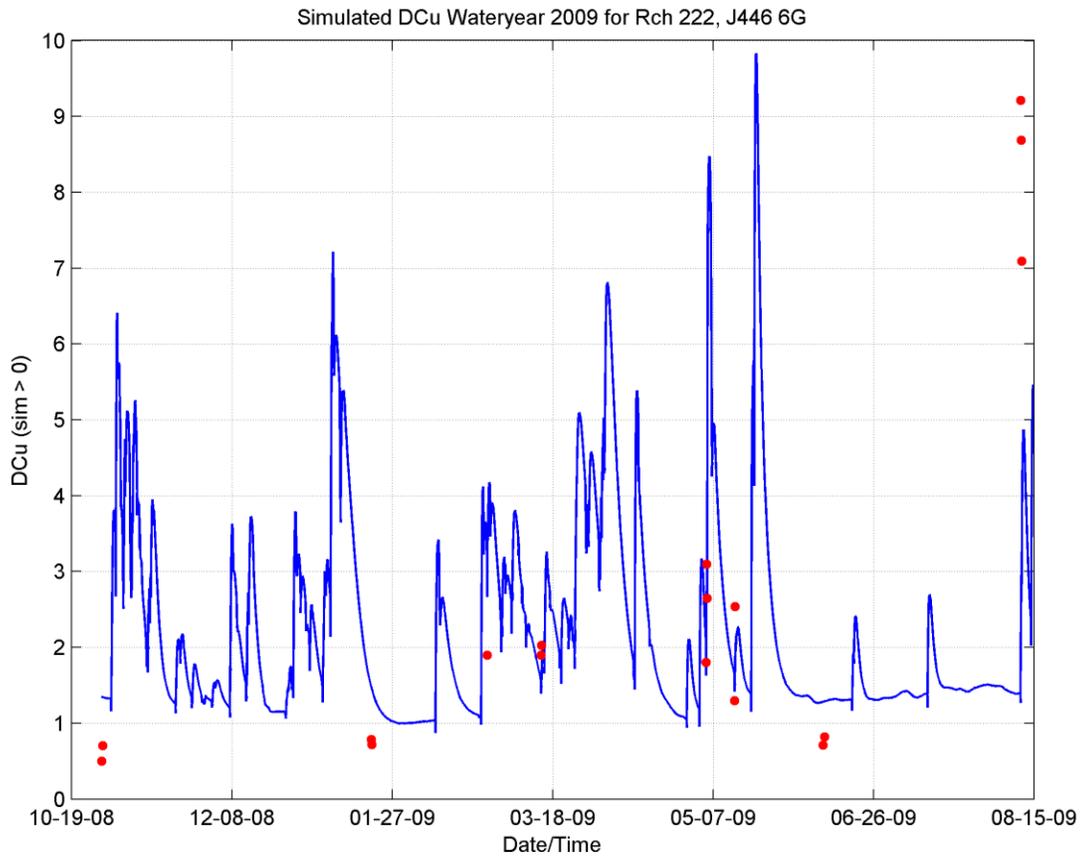


Figure 32 Time series plot of dissolved copper concentrations for east branch tributary (6G). Red colors are observed, blue is simulated.

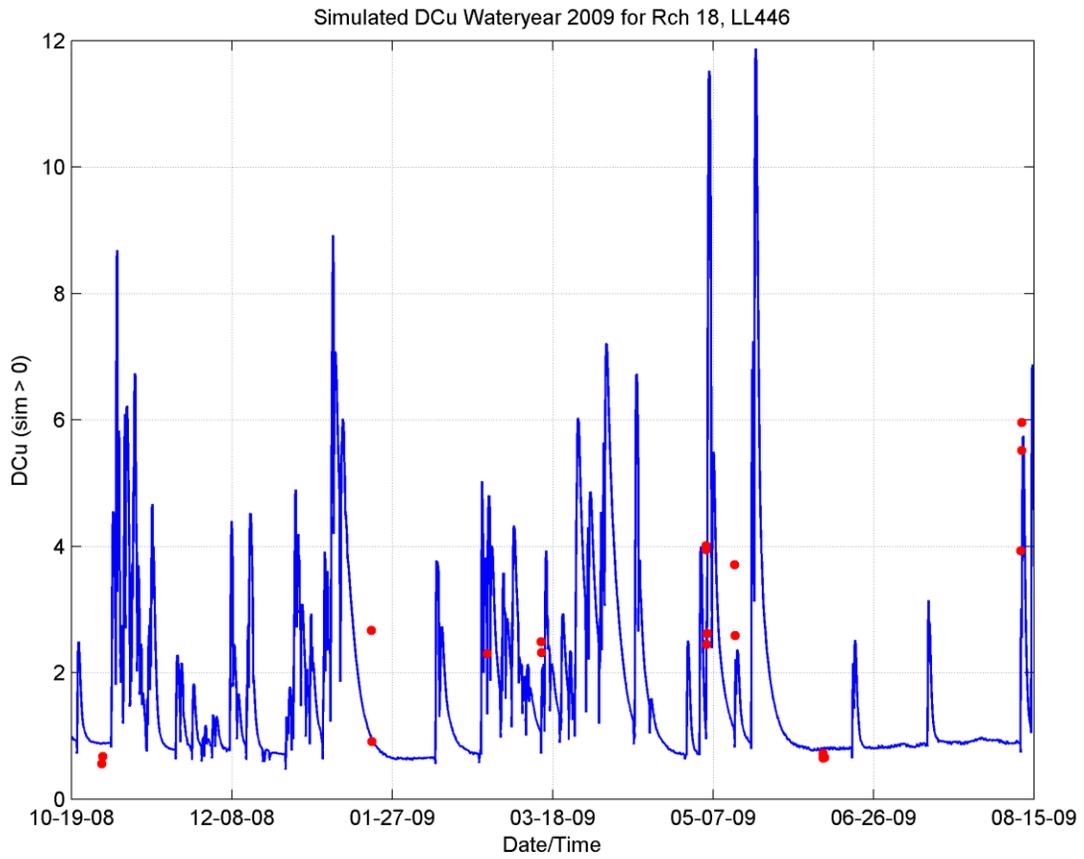


Figure 33 Time series plot of dissolved copper concentrations for north branch tributary (7G). Red colors are observed, blue is simulated.

5.8 Nutrients

Metrics used for evaluating nutrient loads in defined scenarios are based on annual loading rates for Juanita Creek mainstem. Therefore the test for model acceptance is based on performing a non-parametric test of distributions using the Mann-Whitney *U*-Test with a p-value greater than 0.10 as acceptable similarity in distributions and summarized in Table 17 below. Further detail about the calibration parameters are in the follow subsections.

Table 17 Summary of Mann-Whitney U-Test of calibration for nitrogen and phosphorus species.

Mann-Whitney U-test for Juanita Creek mainstem (1G)		
Parameter	p-value	Test
Nitrates	0.01	Fail*
Ammonia-N	0.24	Pass
Orthophosphorus	0.41	Pass
Total Phosphorus	0.31	Pass

*Passes two of the three headwater tributaries (.75, .57,.08)

Ammonia-N

Ammonia-N is modeled by generating nonpoint loadings from surface runoff, interflow, and groundwater. Ammonia is assumed to be exclusively in dissolved form, and not associated with sediment.

Modeled mainstem ammonia concentrations do not reflect the high ammonia concentration measured during discrete storm events. Two observed events have concentrations three times and ten times all other observed events during the five years of reported values (Figure 34). Those two events appear to have substantial duration associated with them suggesting a systemic (possibly seasonal) source. Without further understanding the causality of those, replicating them in the model would be tenuous and lead to excessively high annual loading rates (Horner, et al. 1994); thus, were ignored during parameter adjustments but not removed from the dataset used testing model accuracy. Simulated mainstem ammonia concentrations pass the U-test with a p-value of 0.24. See Table 18 at the end of this section for other statistics.

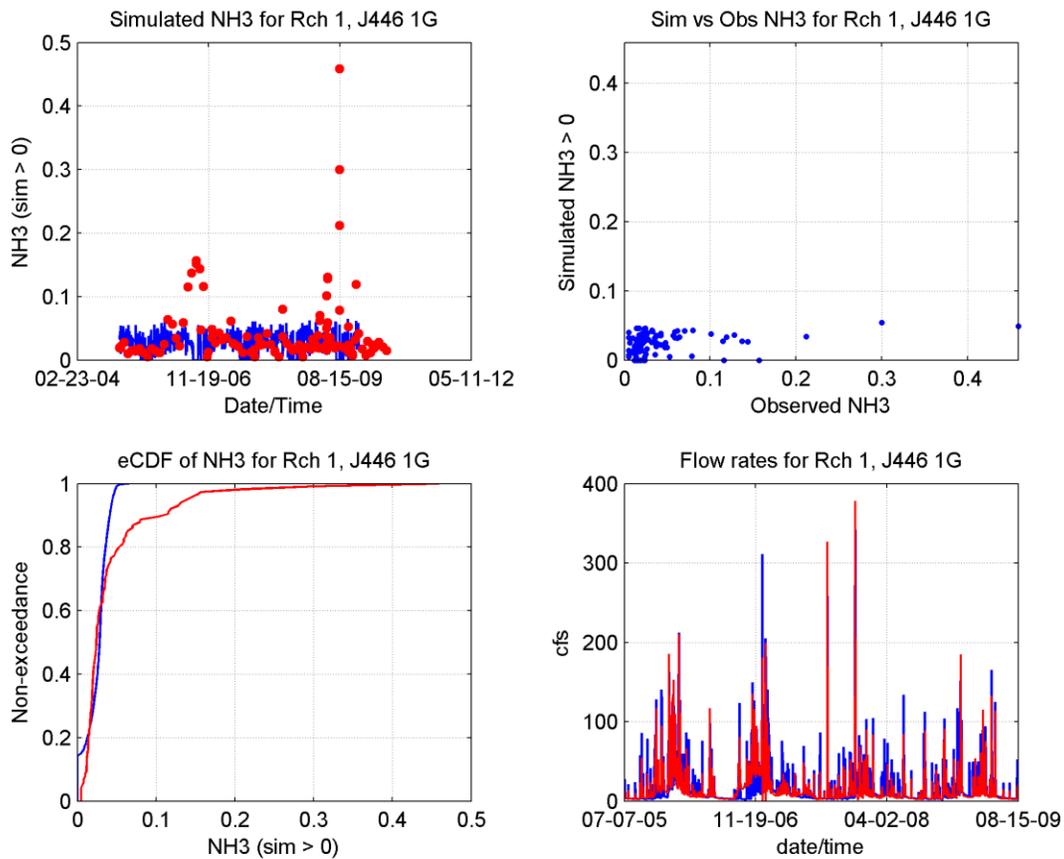


Figure 34 Time series, scatter plot, and cumulative distribution plots of calibrated for ammonia-n concentrations for the mainstem (1G). Red is observed and blue is simulated.

Nitrate-N

Nitrate-N is modeled similarly to ammonia, based on surface accumulation, washoff with surface runoff, and definition of monthly-varying interflow and groundwater concentrations. Calibration of nitrate and ammonia was largely done by adjusting the interflow and groundwater concentrations (and ammonia surface loading factors) by land use, until the errors were minimized at the eight calibration points. While the agreement was fairly good for nitrate (r -square > 0.50) at four of the eight calibration points (3G, 4GO, 6G, and 7G), the mainstem r -square value was 0.26 (Table 19). However, there is enough model error to fail the U-test (p -value approximately 0.01). This model error can be seen in the scatter plot and cumulative distribution as shown in Figure 35 below.

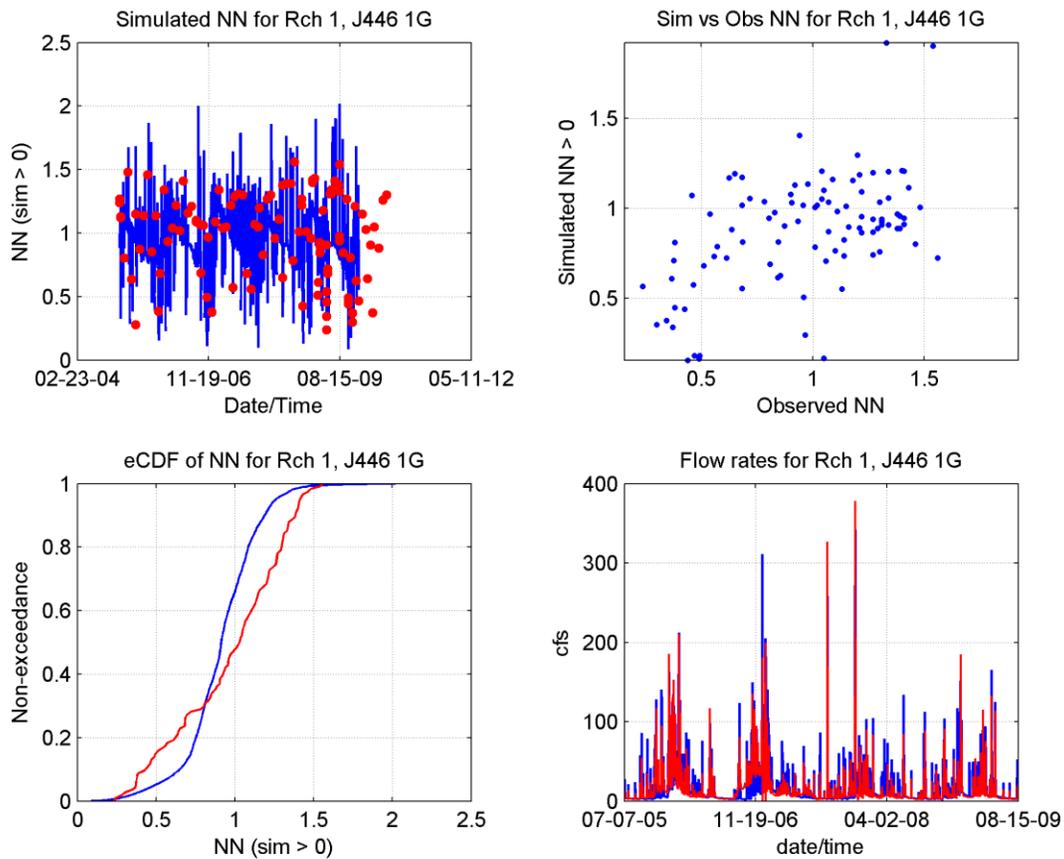


Figure 35 Time series, scatter plot, and cumulative distribution plots of calibrated for nitrate concentrations for the mainstem (1G). Red is observed and blue is simulated.

Orthophosphate-P

Orthophosphate-P is modeled similarly to ammonia with one major difference. Because of its propensity to be bound to solids, it is associated with sediment instead of surface runoff and is also associated with sediment in the stream, and loadings are calculated based on surface accumulation, wash-off in association with sediment that is transported to the stream, and definition of monthly-varying interflow and groundwater concentrations. Similar to ammonia-n, two drawn out events are observed with concentrations twice other peaks reported during a five year period. Correlations are poor, but have similar distributions (Figure 36). Without further understanding the cause of those elevated concentrations, annual loadings would be substantially higher than supported by the multiple years of observed data. Orthophosphate concentrations were calibrated by adjusting the land use-specific interflow and groundwater concentrations and the surface parameters (potency factors) seasonally to achieve a fit ignoring those two events. The statistical measures indicate the model is poorly calibrated with respect to instantaneous

values (Table 20), but passes the U-test of fitness with a p-value of 0.41. Note that storms produce spikes of PO4, which is primarily from the surface-generated particulate P.

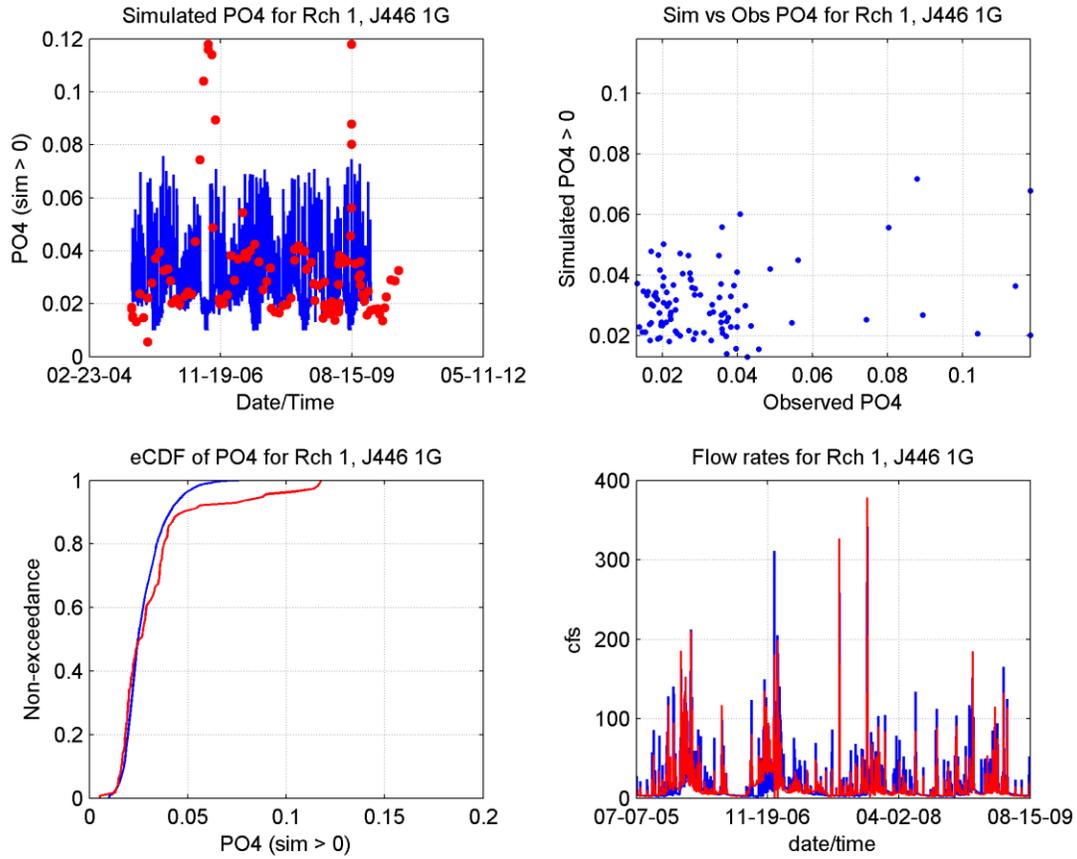


Figure 36 Time series, scatter plot, and cumulative distribution plots of calibrated for orthophosphorus concentrations for the mainstem (1G). Red is observed and blue is simulated.

Table 18 Summary statistics for calibration of ammonia-N.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Ammonia-N (mg/L)	RMSE	0.06	0.06	0.02	0.28	0.04	0.28	0.11	0.10
	ME	-0.02	0.03	-0.01	-0.16	-0.02	-0.07	-0.05	-0.05
	RPD	-44%	130%	-50%	-90%	-94%	-62%	-70%	-67%
	r-square	0.08	0.02	0.18	0.07	0.14	0.00	0.03	0.24

Table 19 Summary statistics for calibration of nitrates.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Nitrate (mg/L)	RMSE	0.34	0.78	0.63	0.94	0.60	0.53	0.40	0.33
	ME	-0.09	-0.51	0.50	0.76	0.55	-0.04	0.22	0.12
	RPD	-10%	-37%	152%	502%	500%	-4%	24%	16%
	r-square	0.26	0.27	0.54	0.53	0.27	0.17	0.51	0.64

Table 20 Summary statistics for orthophosphate.

Parameter	Statistic	Mainstem (1G)	Billy Creek (2G)	Totem Lake Trib. (3G)	Wetland outlet (4GO)	Wetland inlet (4GI)	West Branch (5G)	East Branch (6G)	North Branch (7G)
Ortho-phosphate (mg/L)	RMSE	0.02	0.04	0.03	0.03	0.02	0.06	0.03	0.02
	ME	0.00	-0.03	0.03	0.02	0.02	-0.01	0.00	0.00
	RPD	-8%	-42%	148%	121%	134%	-18%	7%	9%
	r-square	0.05	0.32	0.00	0.34	0.05	0.13	0.02	0.18

6.0. DISCUSSION

There are two approaches in calibrating a model when multiple locations of observed data are available: 1) build multiple sub models (seven to eight would be required for this study) and calibrate each one with likely unique parameterization, or 2) minimize model error among all observation locations and weight parameters with best fits focusing on Juanita Creek mainstem. Since land use in the study area is fairly uniform with the exception of Totem Lake drainage subbasin dominated by commercial land use, there was not enough variability to support multiple calibrated models generating multiple unique parameter datasets. Thus, model calibration was conducted minimizing error among the calibration points, but when necessary, giving more weight to the mainstem.

The rigor of the statistical testing of the calibrated model was one of three levels; with the most comprehensive testing include using a suite of statistics characterizing various aspects of the hydrologic regime and how well the model replicates it. Based on the suite of statistics used, the model is well calibrated at all locations except during the lowest flow conditions in Totem Lake tributary. This deficiency is not believed to substantially affect outcomes from the various stormwater management scenarios evaluated for this study.

The next level of rigor includes four statistics (root-mean-square-error, mean error, relative percent difference, and r-square) applied to simulated and observed water quality parameters requiring competency predicting instantaneous conditions throughout the period of record. These parameters include water temperature, dissolved oxygen, dissolved copper, and fecal coliforms. Because copper is sediment associated, the same level of statistical rigor was applied to suspended solids as well. Overall the model is well calibrated for water temperature, and dissolved oxygen throughout the study area. However, as complexity of a parameter increases either mechanistically and/or with uncertainty, the model accuracy diminishes. Total copper was generally simulated with acceptable accuracy among all the calibration points except near the outlet of the large wetland (4G0). As previously mentioned this is likely due to plant uptake in the wetland not represented in the model. Dissolved copper simulations were generally less accurate but more variable among calibration points. Emphasis was given to best characterizing the mainstem of Juanita Creek. However, processes simulating dissolved copper were statistically best modeled at the confluence of the Totem Lake tributary (3G). Further evaluations of copper at the various calibration points using visual inspections of time series, scatter plots, and cumulative distributions reveal that other than slight shifts in timings, or the occurrence of an “anomalous” event, simulated copper concentrations are well characterized at all but one calibration point (4G0).

Simulated fecal coliform concentrations generally follow observed concentrations except for the highest concentrations when the model is under simulating those conditions. It is believed that there are multiple cross-connections, illicit connections, and failing onsite septic systems within the Juanita Creek basin. These types of conditions can generate instantaneous loads not predictable with any consistency without substantially more modeling effort and likely a different modeling scheme or framework. Thus while the

model accuracy is relatively poor, it is believed to be usable for evaluating mitigation strategies relative to each other in the study.

Nutrients were generally less accurately modeled, but were tested for equivalency using the Mann-Whitney *U*-test—the least level of statistical rigor. All but nitrates passed the test focusing the mainstem of Juanita Creek. This model deficiency is likely a result of the before mentioned wide spread septic sewer loads as revealed in the monitoring of fecal coliform concentrations. So while the model fails the test, it still generally reflects conditions in the basin and can be used for such applications with this understanding.

Overall the model is weakest in the Totem Lake tributary system (3G, 4GO, and 4GI), but this is understandable given the greater level of uncertainty characterizing the hydrodynamics and nutrient cycle for Totem Lake (proper), a large wetland downstream, and the flat highly vegetative stream reach connecting the two. Notwithstanding these uncertainties, the model does characterize the various elements throughout the study area, but with some inconsistency. We conclude that the model is sufficiently well calibrated throughout the basin for assessing stormwater management scenario impacts on flows. We also conclude that the model is sufficiently well calibrated only in the mainstem for assessing stormwater management scenario impacts on water quality.

7.0. REFERENCES

- Bicknell, B.R., J.C. Imhoff, J.L. Kittle Jr., T.H. Jobs, and A.S. Donigian, Jr. 2005. Hydrological Simulation Program - Fortran (HSPF). User's Manual for Release 12. U.S. EPA National Exposure Research Laboratory, Athens, GA, in cooperation with U.S. Geological Survey, Water Resources Division, Reston, VA.
- DeGasperi, C.L., H.B. Berge, K;R. Whiting, J.J. Burkey, J.L. Cassin, and R.R. Fuerstenberg. 2009. Linking Hydrologic Alteration to Biological Impairment in Urbanizing Streams of the Puget Lowland, Washington, USA. *Journal of the American Water Resources Association* 45(2):512-533.
- Horner, Richard R., Joseph J. Skupien, Eric H. Livingston, and H. Earl Shaver. 1994. Fundamentals of Urban Runoff Management: Technical and Institutional Issues. Prepared by the Terrene Institute, Washington, DC, in cooperation with the U.S. Environmental Protection Agency.
- King County. (in progress). Development of a Stormwater Retrofit Plan for Water Resources Inventory Area (WRIA) 9, and Estimation of Costs for Retrofitting all Developed Lands of Puget Sound. EPA funded.