DEPARTMENT OF
ECOLOGY
State of Washington

Lake Whatcom Models Review

Independent review performed by Tetra Tech

for the Washington State Department of Ecology

April 2008
Publication no. 09-10-013

Publication and Contact Information

This report is available on the Department of Ecology's website at
www.ecy.wa.gov/biblio/0910013.html

For more information contact:

Water Quality Program
Bellingham Field Office
Washington Department of Ecology
1440 – 10th St., Suite 102
Bellingham, WA 98225

Phone: (360) 715-5200

Washington State Department of Ecology www.ecy.wa.gov/

Lake Whatcom Models Review

---

by
Dr. Jonathan Butcher
Tetra Tech

Funded by the U.S. Environmental Protection Agency

Water Quality Program
Washington State Department of Ecology
Olympia, Washington 98504-7600

This page is purposely left blank

TETRA TECH

Cape Fear Building, Suite 105
3200 Chapel Hill-Nelson Highway, PO Box 14409
Tel 919.485.8278 Fax 919.485.8280 www.tetratech.com

10306 Eaton Place, Suite 340, Fairfax, VA 22030
Tel 703.385.6000 Fax 703.385.6007 www.tetratech.com

To:	Steve Hood, Washington Ecology Paul Pickett, Washington Ecology Dave Ragsdale, USEPA Region 10
From:	Dr. Jonathan Butcher
CC:	Jayne Carlin, USEPA Region 10 Bob Cusimano, Washington Ecology Jessica Koenig, Tetra Tech Andrew Parker, Tetra Tech
Subject:	**Lake Whatcom Models Review**
Date:	April 17, 2008

The Washington Department of Ecology (WDOE) is developing a nutrient TMDL for Lake Whatcom, near Bellingham, WA. WDOE requested an independent review of the models and modeling documents prepared for this project. Attached you will find a summary of Tetra Tech's review of the HSPF watershed model and CE-QUAL-W2 model developed for Lake Whatcom. The review included evaluation of model inputs and outputs, model documentation and model code (for W2). Based on the reviews, this memo provides answers to the following questions:

1. Does the model appropriately simulate the physical, chemical, and biological processes relative to the decision needs?
2. Does the model appropriately incorporate relevant and available data?
3. Does the model reasonably predict observed water quality conditions given the available information?
4. Were standard modeling procedures and protocols followed?
5. What are the model's shortcomings? What are the implications of these shortcomings on conclusions drawn from the model?

Contents

Tables

Figures

1. HSPF Watershed Model

The TMDL package includes an HSPF watershed model for hydrology and pollutant transport. The model input files supplied run successfully and produce results that appear to replicate those shown in the report. Documentation of the model is contained in the report *Final Report for Lake Whatcom Watershed TMDL Model Project* (The Cadmus Group and CDM, July 18, 2007)

The HSPF model was created from an existing HFAM (Hydrocomp Forecast and Analysis Modeling) project developed to analyze water supply in the Bellingham area, which contains the Lake Whatcom watershed as a subset. The HFAM is a proprietary tool of Hydrocomp, Inc.; however, the hydrologic simulation portion is based on HSPF, and is thus readily transferable to HSPF. CDM, the subcontractor for the watershed model, apparently completed this transference on an expedited schedule, and then pursued some additional refinement of the hydrology. The HFAM was reported to be calibrated for Bellingham, which should provide a strong starting point for the HSPF application. However, no calibration report for the HFAM was provided, and the brief discussion in the Cadmus/CDM does not provide a detailed discussion of how individual model parameters were derived in relation to watershed physical characteristics.

1.1. METEOROLOGICAL FORCING

One of the most important constraints on the accuracy of a watershed model is the accuracy of the meteorological data that drive it. For the Lake Whatcom application, the key inputs are precipitation and potential evapotranspiration (PET). (As the application does not consider snow due to the elevation range present in the watershed, other meteorological inputs, such as air temperature, wind, and solar radiation, do not have direct impacts on the hydrology.) Any inaccuracies in these forcing functions will directly translate into inaccuracies in the simulation of hydrology.

Precipitation input is based on three hourly precipitation gauges maintained by the City of Bellingham, located around the perimeter of the lake. A key issue in any watershed model is the degree to which point measurements of rainfall are correlated with the spatial input of precipitation across a watershed area. While the three gauges provide measurements around the lake, all are located near the elevation of the lake itself. Orographic effects are expected to result in increased precipitation at higher elevations. This issue is addressed through use of lapse rates, applied as a constant multiplier for each HSPF subbasin. These lapse rates are described as being summarized from the elevation band multipliers developed for the HFAM model; however, the backup justification for the development of the lapse rates (which appear to reflect both elevation and aspect) is not provided in the Cadmus/CDM report. In addition, the “precipitation factors” were varied somewhat during calibration

The City of Bellingham precipitation gauges do not report to the National Climatic Data Center (NCDC), and do not appear to be available online. As a result, I have not been able to directly verify the accuracy of the precipitation data included in the model. Comparison between gauges can, however, be used to check for internal consistency and flag major potential errors. A plot of monthly summed precipitation depth at the Brannian and Smith Creek gages reveals one apparent problem (Figure 1)

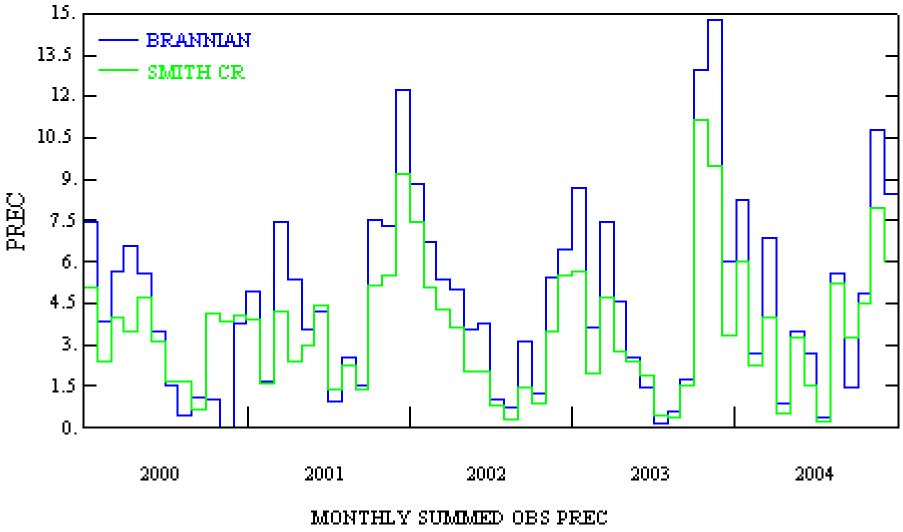

Figure 1. Comparison of Monthly Precipitation Depth (in.) Reported for Brannian and Smith Creek Gauges

During most months, total precipitation at these two gauges shows a similar pattern, with generally higher precipitation at the Brannian gauge (southern end of the lake) as opposed to the Smith Creek gauge (northern end of the lake). In November of 2000, however, the model input shows zero precipitation at the Brannian gauge, whereas about 4 inches was recorded at Smith Creek. Further, the October precipitation total at Brannian is much lower than that at Smith Creek. This strongly suggests the presence of uncorrected missing data. A separate report that analyzed the water budget of Lake Whatcom for 2000-2001[1] states

> *Precipitation data from the Water District #10, Division 30 gauge were substituted for the Brannian Creek location for the months of October and November 2000 and March and April 2001, because the Brannian Creek gauge was inoperable and the two gauges were found to be comparable during other months. Precipitation data from the Water District #10 District Office gauge were substituted for the Smith Creek location for the months of October and November 2000, because the Smith Creek gauge was inoperable.*

The precipitation data contained in the WDM file for the model do contain data for Brannian Creek in March-April 2001 and Smith Creek for October-November 2000 – but not for Brannian Creek in November 2000. This suggests that some patching of the precipitation record might have occurred, but that it is not complete, with at least the November 2000 Brannian Creek missing data being unpatched. A more detailed QA review of the precipitation data thus appears warranted (but would require access to the original data).

Attributing missing precipitation data as zeros (as occurred at least for November 2000 at Brannian Creek) will have an impact on model performance, and might well have skewed the hydrologic calibration effort.

PET is the other major meteorological control on hydrology. The report states that "hourly calculated potential evapotranspiration data from the Smith Creek station is used for all subwatersheds in the model." The report does not document the method used to calculate PET, which appears to have been

---

[1] Matthews, R.A., M. Hilles, J. Vandersypen, R.J. Mitchell, and G.B. Matthews. 2006. Lake Whatcom Monitoring Project, 1999/2000, Final Report. Available online at http://ceratium.ietc.wwu.edu/IWS2/lakestudies/lakewhatcom/online_html/2001/final_01su8.html.

taken from the HFAM input. The selection of approach can have a significant effect on the model. It appears as though PET was calculated based on solar radiation, wind, and air temperature at the Smith Creek station using Jensen-Hayes, Penman, or some other method.

A plot of the daily total PET used by the model looks generally reasonable, and the annual totals are in line with the national maps of annual lake evaporation[2]. However, examination of the data for 2004 does reveal a problem at the end of the simulation (Figure 2). Specifically, after 30 September 2004, all values appear to be set to either 0.01 in/hr or zero.

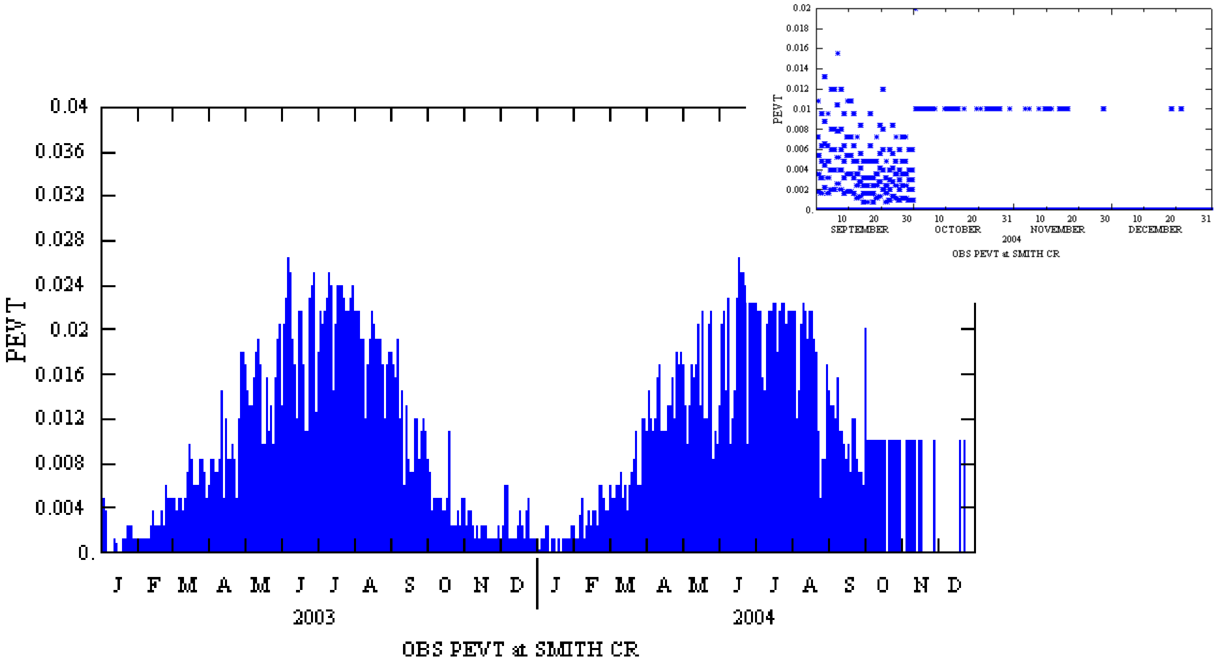

Figure 2. Hourly PET Values (in) used by Lake Whatcom Watershed Model for 2003-2004

Note: Inset shows values assigned for September-December, 2004.

1.2. HSPF MODEL SETUP

The model uses seven pervious land use categories from the 2001 NLCD, plus a single impervious cover, assigned as a fixed percentage of developed land based on NLCD imperviousness estimates. This is not technically correct, as HSPF should be set up to represent directly connected impervious area (DCIA), not total impervious area. Likely compensating for this is the fact that NLCD satellite imagery is not able to identify small fractions of impervious area in rural landscapes, particularly under tree canopy. The assumption that the same impervious fraction applies to all developed land in the watershed is also suspect. However, given the small amount of developed land, these assumptions are unlikely to result in significant errors.

A variety of hydrologic soil groups and slopes are present in the watershed. The model setup does not directly make use of these data, but rather assigns parameters during calibration to individual watersheds. This is an acceptable practice where individual subwatersheds are small enough and structured so as to contain relatively homogenous soils and slopes. This might be the case, but is not demonstrated in the report. Potential problems could arise in translation from gaged and calibrated watersheds (where the parameters presumably represent a weighted average across the contained soils and slopes) to ungaged

---

[2] Kohler, M.A., T.J. Nordenson, and D.R. Baker. 1959. Evaporation Maps for the United States. U.S. Weather Bureau Technical Paper 37.

watersheds. The report notes that the assignment to ungaged watersheds was made "based on a review of land use, soil hydrologic group, and geographic proximity," but details are not provided and the adequacy of this assignment cannot be evaluated. It might have been preferable to subdivide the land use categories by overlaying land use with soil hydrologic group and slope range.

One problem is created because water/wetlands is defined as an upland pervious land use. The "water" pervious land segments are then assigned parameters like those for other pervious land segments, with unsaturated soil moisture storage, infiltration, lower zone ET, etc. This is not correct. Wetlands can be represented as upland land segments in HSPF (for instance, using the high water table option), but the parameters will be quite different from those for dry soils. Fortunately, water/wetlands constitute only 1 percent of the land area in the watershed, so the overall impact will be small. In the Mirror Lake subwatershed, however, 10 percent of the area is in water/wetlands, and hydrologic predictions for this subwatershed will be significantly biased. Because the water area is included within the upland area, the model correctly does not assign either precipitation or ET to the stream reach segments.

1.3. HYDROLOGIC PARAMETERS

Parameters for the HSPF model generally fall within the recommended ranges given in BASINS Technical Note 6[3], but do not always vary between watersheds in ways that make intuitive sense. The model is set up as a cross-tabulation overlay of subwatersheds and land use classes and does not subdivide land use classes by soil hydrologic group or slope. While HSPF models are frequently set up in this way, many of the model parameters can reasonably be anticipated to vary with soil hydrologic group and slope, and classification by land use only limits the capability for refinement of the model. Key hydrologic parameters are set as follows:

LZSN, lower zone nominal soil moisture storage, is set at the sub-basin level, and varies from 2.0 to 11.7 inches in the Lake Whatcom watershed model, whereas the HFAM model assumed a constant value of 5.0. The calibrated values cover most of the "possible" range of 2.0 to 15.0 inches cited in Technical Note 6, and both the high and low values fall outside the "typical" range of 3.0 to 8.0 inches. LZSN is an index to soil storage and not a directly measurable physical property itself, but is related to both precipitation patterns and soil characteristics (e.g., available water capacity over the rooting depth). For the Lake Whatcom watershed, the precipitation component of LZSN should be relatively consistent across the watershed, while there might be some moderate amount of variation with soil type (most of the soils within the watershed are in Hydrologic Groups B and C). The summary of calibrated values provided in Figure 5.1 of the ARM model user's manual[4] suggest LZSN should typically be around 8.0 for lower elevations and 14.0 for higher elevations in western Washington. Thus, the low calibrated LZSN values of 2-3 in. for Austin Creek, Olsen Creek, and Smith Creek appear suspect. Further, estimates for LZSN and INFILT typically have an inverse correlation relative to total flow volume, raising the possibility that a false calibration has been obtained that replicates total volume, but not the components of the hydrograph. The modelers do not report a water balance analysis, which is general good practice for watershed models. In particular, a hydrograph separation program could be applied to the gage data and the resultant estimates of quickflow and baseflow compared to model predictions of surface runoff and groundwater discharge.

INFILT, the index of infiltration capacity, is primarily a function of soil characteristics and is typically varied according to soil hydrologic soil group. In the Lake Whatcom watershed model, INFILT is first varied in accordance with land use, then scaled into three groups by subbasin (with higher values for basins 72, 80, 85, and 90; lower values for basins 20 and 25). Variation among natural land use classes is atypical, and the variation among subbasins does not follow soil hydrologic group (each grouping of

---

[3] USEPA. 2000. BASINS Technical Note 6, Estimating Hydrology and Hydraulic Parameters for HSPF. EPA-823-R00-012. Office of Water, U.S. Environmental Protection Agency, Washington, DC.

[4] USEPA. 1978. User's Manual for Agricultural Runoff Management (ARM) Model. EPA 400/3-78-080. Environmental Research Laboratory, Athens, GA.

subbasins contains a mix of B and C soils). The values that are assigned (0.02 to 0.1) seem somewhat low for the mix of soils cited, as the typical range is 0.05 – 0.1 for C soils and 0.1 – 0.4 for B soils. (The original HFAM model had a constant value of 0.08, appropriate to C soils.) The low value of INFILT has apparently been compensated by increasing the value of INTFW (interflow inflow parameter), as both INFILT and INTFW divert water from direct runoff to delayed runoff.

AGWRC, the groundwater recession coefficient, is set at discrete values of 0.97, 0.98, or 0.99, and is varied by subbasin. AGWRC is a complex function of climate, topography, soils, and land use, and some variability is expected (but not implemented for Lake Whatcom) by land use, with higher values for forest. In most applications, AGWRC would be set through analysis of hydrograph recession curves, but no such discussion is presented here.

DEEPFR, the fraction of infiltrating water lost to deep aquifers, is set to 0.14 in all subbasins, referencing a citation that total deep aquifer recharge in the watershed is approximately 2 inches per year. However, because no upland water balance is presented, the details of this assignment are not clear.

LZETP, lower zone ET factor, is set to a seasonal pattern but is not varied by land use. A value of 0.9 is assigned for 1 May to 1 November, with a value of 0.1 for the remainder of the year. Use of a seasonal pattern is typical, although a more gradual curve, rather than a step-function, is typical to represent changes in density of vegetation and stage of plant growth. Further, LZETP is expected to vary with land use, with higher values for forest than for grassland, and more winter ET for evergreens than for deciduous forest.

CEPSC, initial interception (in.), is varied by land use and subbasin, but not by season. This parameter is primarily a function of canopy density, and so should vary by land use, but probably not by subbasin. Further, seasonal variability is expected, particularly for the deciduous forest land cover.

In sum, while there do not appear to be any gross errors in the specification of hydrologic parameters, there are questions regarding specific values. Calibration of HSPF is challenging because there are multiple free parameters that tend to be correlated with one another, making it difficult to achieve a unique calibration. The problem is that a general fit can often be obtained by compensating errors unless multiple tests and constraints are imposed to ensure that all portions of the water balance are adequately represented. Incorrect assignment of parameter values, particularly with regard to distributing flow between surface and groundwater, might also impact representation of water quality.

1.4. HYDROLOGIC CALIBRATION APPROACH

The HSPF model is set up to run for calendar years 2000 through 2004. The entire period is used for model calibration. There is not a separate validation period to test performance of the calibrated model, although this could likely have been done for 1995-1999 observations.

Calibration of hydrology focuses on six gages. Three of these have continuous USGS gaging for the calibration period (Olsen, Mill, and Euclid Creek). The other three have 15-min gaging from the Western Washington University (WWU). While the UWW data are available online, they have frequent gaps and are not as useful for water balance evaluation. In addition, there is a USGS gage on Brannian Creek (12201960) with data for the whole calibration period, but was apparently not used in the calibration. This gage provides a useful independent check on the calibration.

The calibration approach focuses on relative error, relative absolute error, root mean square error, and the correlation coefficient, along with the Nash-Sutcliffe (NS) and McMahon (MH) coefficients of model fit efficiency. All six measures are performed on untransformed flow series, although the errors in such series will typically exhibit scale dependence (heteroscedasticity) because observations are constrained to be greater than or equal to zero. Without transformation, the first four measures focus on overall fit to the total time series and will be strongly influenced by the high-flow events that constitute a large part of the total annual flow volume. The NS statistic compares the predictive ability of the model to the predictive ability of the average and is also strongly influenced by high outliers. The MH statistic works with square root transformed data, reducing, but not eliminating, the influence of high outliers.

Notably lacking from the calibration approach is any attempt to evaluate the water balance. The choice of calibration statistics will require the model to honor total annual flow volume, but not necessarily to correctly represent the different portions of the hydrograph. Typical procedure for calibration of hydrology in HSFP would involve additional measures of fit to the high and low flow ranges, seasonal flow volumes, and storm volumes to ensure that a realistic water balance is attained[5].

As would be anticipated, the calibrated model provides a reasonable fit to mean daily flow, as well as to seasonal flows on the basis of November-April and April-October (this seasonal division is not really adequate, as it does not isolate the low-flow period), although there is often a divergence between observed and predicted maximum daily flow. The NS coefficients appear acceptable (75-89%) for four of the six reported calibration gages, but are rather low at Euclid Creek (44%) and Olsen Creek (42%).

1.5. HYDROLOGIC CALIBRATION RESULTS

To further investigate the calibration, we applied some of the standard HSPF hydrologic calibration measures to the four stations with USGS gaging concurrent to the calibration run (including Brannian Creek, not examined in the original calibration report).

Statistical summaries of the status of hydrologic calibration are provided in Table 1 through Table 4. Percent errors that exceed the recommended boundaries for HSPF applications are highlighted.

---

[5] Lumb, A.M, R.B. McCammon, and J.L. Kittle, Jr. 1994. User's Manual for an Expert System (HSPEXP) for Calibration of the Hydrological Simulation Program – Fortran. Water-Resources Investigations Report 94-4168. U.S. Geological Survey, Reston, VA.

Table 1. Statistical Summary of Hydrologic Calibration, Euclid Creek (USGS 12202400)

Simulated Flow		**Observed Flow Gage**	
REACH OUTFLOW FROM DSN 85, LW_Existing 3.25-Year Analysis Period: 10/1/2001 - 12/31/2004 Flow volumes are (inches/year) for upstream drainage area Original files		**USGS 12202400 EUCLID CR AT EUCLID AVE AT BELLINGHAM, WA** Drainage Area (sq-mi): 0.54	
Total Simulated In-stream Flow:	**14.28**	Total Observed In-stream Flow:	**13.02**
Total of simulated highest 10% flows:	**7.10**	Total of Observed highest 10% flows:	**7.35**
Total of Simulated lowest 50% flows:	**0.656**	Total of Observed Lowest 50% flows:	**0.412**
Simulated Summer Flow Volume (months 7-9):	**0.355**	Observed Summer Flow Volume (7-9):	**0.133**
Simulated Fall Flow Volume (months 10-12):	**7.02**	Observed Fall Flow Volume (10-12):	**7.51**
Simulated Winter Flow Volume (months 1-3):	**5.82**	Observed Winter Flow Volume (1-3):	**4.38**
Simulated Spring Flow Volume (months 4-6):	**1.09**	Observed Spring Flow Volume (4-6):	**1.00**
Total Simulated Storm Volume:	**3.17**	Total Observed Storm Volume:	**3.62**
Simulated Summer Storm Volume (7-9):	**0.1177**	Observed Summer Storm Volume (7-9):	**0.0464**

Errors (Simulated-Observed)	*Error Statistics*	*Recommended Criteria*
Error in total volume:	9.73	10
Error in 50% lowest flows:	**59.28**	**10**
Error in 10% highest flows:	-3.42	15
Seasonal volume error - Summer:	**167.73**	**30**
Seasonal volume error - Fall:	-6.45	30
Seasonal volume error - Winter:	**32.86**	**30**
Seasonal volume error - Spring:	9.00	30
Error in storm volumes:	-12.47	20
Error in summer storm volumes:	**153.88**	**50**
Nash-Sutcliffe Coefficient of Efficiency, E:	0.434	Model accuracy increases as E or E' approaches 1.0
Baseline adjusted coefficient (Garrick), E':	0.526	

Table 2. Statistical Summary of Hydrologic Calibration, Mill Creek (USGS 12202420)

Simulated Flow		**Observed Flow Gage**	
REACH OUTFLOW FROM DSN 90, LW_Existing 2.55-Year Analysis Period: 6/1/2002 - 12/31/2004 Flow volumes are (inches/year) for upstream drainage area Original files		**USGS 12202420 MILL CREEK NEAR BELLINGHAM, WA** Drainage Area (sq-mi): 0.79	
Total Simulated In-stream Flow:	**12.54**	Total Observed In-stream Flow:	**14.04**
Total of simulated highest 10% flows:	**5.82**	Total of Observed highest 10% flows:	**6.73**
Total of Simulated lowest 50% flows:	**0.622**	Total of Observed Lowest 50% flows:	**0.385**
Simulated Summer Flow Volume (months 7-9):	**0.467**	Observed Summer Flow Volume (7-9):	**0.471**
Simulated Fall Flow Volume (months 10-12):	**6.27**	Observed Fall Flow Volume (10-12):	**7.03**
Simulated Winter Flow Volume (months 1-3):	**4.65**	Observed Winter Flow Volume (1-3):	**5.26**
Simulated Spring Flow Volume (months 4-6):	**1.16**	Observed Spring Flow Volume (4-6):	**1.28**
Total Simulated Storm Volume:	**2.25**	Total Observed Storm Volume:	**2.53**
Simulated Summer Storm Volume (7-9):	**0.1687**	Observed Summer Storm Volume (7-9):	**0.1773**

Errors (Simulated-Observed)	*Error Statistics*	*Recommended Criteria*		
Error in total volume:	**-10.65**	**10**		
Error in 50% lowest flows:	**61.42**	**10**		
Error in 10% highest flows:	-13.51	15		
Seasonal volume error - Summer:	-0.80	30		
Seasonal volume error - Fall:	-10.79	30		
Seasonal volume error - Winter:	-11.67	30		
Seasonal volume error - Spring:	-9.38	30		
Error in storm volumes:	-11.31	20		
Error in summer storm volumes:	-4.84	50		
Nash-Sutcliffe Coefficient of Efficiency, E:	0.783	Model accuracy increases as E or E' approaches 1.0		
Baseline adjusted coefficient (Garrick), E':	0.626			

Table 3. Statistical Summary of Hydrologic Calibration, Olsen Creek (USGS 12202300)

Simulated Flow		**Observed Flow Gage**	
REACH OUTFLOW FROM DSN 20, LW_Existing 3.25-Year Analysis Period: 10/1/2001 - 12/31/2004 Flow volumes are (inches/year) for upstream drainage area Original Files		**USGS 12202300 OLSEN CREEK NEAR BELLINGHAM, WA** Drainage Area (sq-mi): 3.78	
Total Simulated In-stream Flow:	**43.32**	Total Observed In-stream Flow:	**43.55**
Total of simulated highest 10% flows:	**23.84**	Total of Observed highest 10% flows:	**25.40**
Total of Simulated lowest 50% flows:	**1.605**	Total of Observed Lowest 50% flows:	**2.620**
Simulated Summer Flow Volume (months 7-9):	**1.489**	Observed Summer Flow Volume (7-9):	**0.975**
Simulated Fall Flow Volume (months 10-12):	**23.10**	Observed Fall Flow Volume (10-12):	**25.40**
Simulated Winter Flow Volume (months 1-3):	**15.30**	Observed Winter Flow Volume (1-3):	**13.05**
Simulated Spring Flow Volume (months 4-6):	**3.43**	Observed Spring Flow Volume (4-6):	**4.12**
Total Simulated Storm Volume:	**13.34**	Total Observed Storm Volume:	**12.79**
Simulated Summer Storm Volume (7-9):	**0.5119**	Observed Summer Storm Volume (7-9):	**0.1992**

Errors (Simulated-Observed)	*Error Statistics*	*Recommended Criteria*		
Error in total volume:	-0.53	10		
Error in 50% lowest flows:	**-38.72**	**10**		
Error in 10% highest flows:	-6.13	15		
Seasonal volume error - Summer:	**52.71**	**30**		
Seasonal volume error - Fall:	-9.05	30		
Seasonal volume error - Winter:	17.23	30		
Seasonal volume error - Spring:	-16.81	30		
Error in storm volumes:	4.30	20		
Error in summer storm volumes:	**157.04**	**50**		
Nash-Sutcliffe Coefficient of Efficiency, E:	0.456	Model accuracy increases as E or E' approaches 1.0		
Baseline adjusted coefficient (Garrick), E':	0.443			

Table 4. Statistical Summary of Hydrologic Calibration, Brannian Creek (USGS 12201960)

Simulated Flow		**Observed Flow Gage**	
REACH OUTFLOW FROM DSN 45, LW_Existing 3.25-Year Analysis Period: 10/1/2001 - 12/31/2004 Flow volumes are (inches/year) for upstream drainage area Original Files		**USGS 12201960 BRANNIAN CREEK AT S BAY DR NR WICKERSHAM, WA** Drainage Area (sq-mi): 3.36	
Total Simulated In-stream Flow:	**41.84**	Total Observed In-stream Flow:	**42.29**
Total of simulated highest 10% flows:	**15.51**	Total of Observed highest 10% flows:	**19.63**
Total of Simulated lowest 50% flows:	**4.175**	Total of Observed Lowest 50% flows:	**2.710**
Simulated Summer Flow Volume (months 7-9):	**1.492**	Observed Summer Flow Volume (7-9):	**0.676**
Simulated Fall Flow Volume (months 10-12):	**19.80**	Observed Fall Flow Volume (10-12):	**20.39**
Simulated Winter Flow Volume (months 1-3):	**15.08**	Observed Winter Flow Volume (1-3):	**15.70**
Simulated Spring Flow Volume (months 4-6):	**5.47**	Observed Spring Flow Volume (4-6):	**5.52**
Total Simulated Storm Volume:	**4.12**	Total Observed Storm Volume:	**7.69**
Simulated Summer Storm Volume (7-9):	**0.1200**	Observed Summer Storm Volume (7-9):	**0.1222**

Errors (Simulated-Observed)	*Error Statistics*	*Recommended Criteria*		
Error in total volume:	-1.06	10		
Error in 50% lowest flows:	**54.03**	**10**		
Error in 10% highest flows:	**-21.02**	**15**		
Seasonal volume error - Summer:	**120.77**	**30**		
Seasonal volume error - Fall:	-2.90	30		
Seasonal volume error - Winter:	-3.96	30		
Seasonal volume error - Spring:	-0.93	30		
Error in storm volumes:	**-46.45**	**20**		
Error in summer storm volumes:	-1.86	50		
Nash-Sutcliffe Coefficient of Efficiency, E:	0.594	Model accuracy increases		
Baseline adjusted coefficient (Garrick), E':	0.624	as E or E' approaches 1.0		

Further details of the model calibration fit for these stations are provided in Figure 3 through Figure 30 and Table 5 through Table 8. The model does well at reproducing total flow volume. However, there are also several consistent problems with the calibration fit:

- All four models appear to underpredict the largest daily peak flows.
- Three out of four models (Euclid, Mill, and Olsen) show elevated levels of uncertainty in simulation of the winter months.
- Three out of four models (Euclid, Mill, and Brannian) show a distinct overprediction of flows during the driest periods. (Note, however, that the observed late summer flows in these creeks are near zero.)

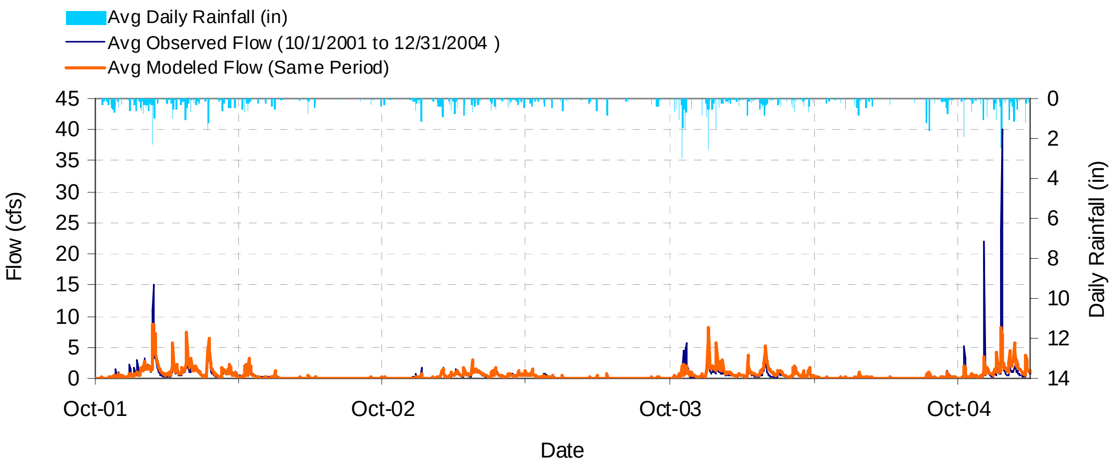

Figure 3. Mean Daily Flow: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

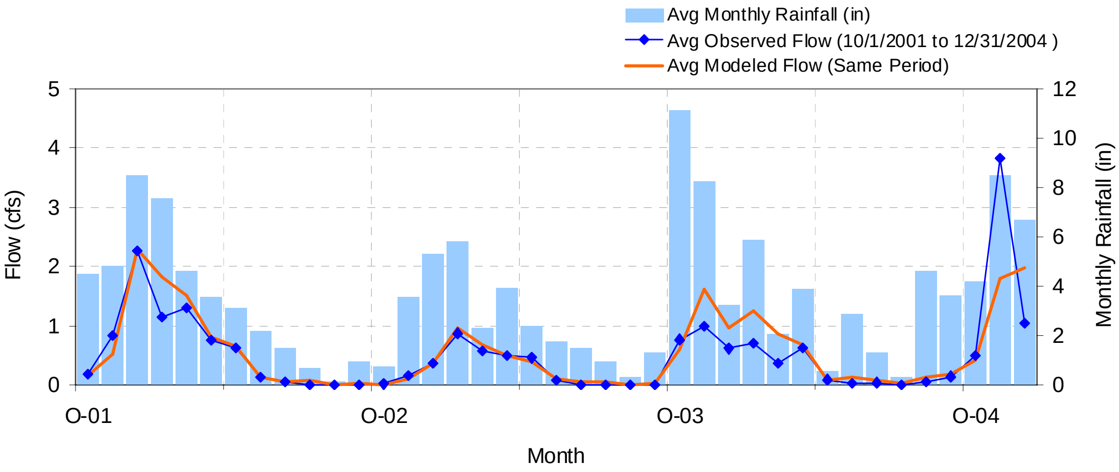

Figure 4. Mean Monthly Flow: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

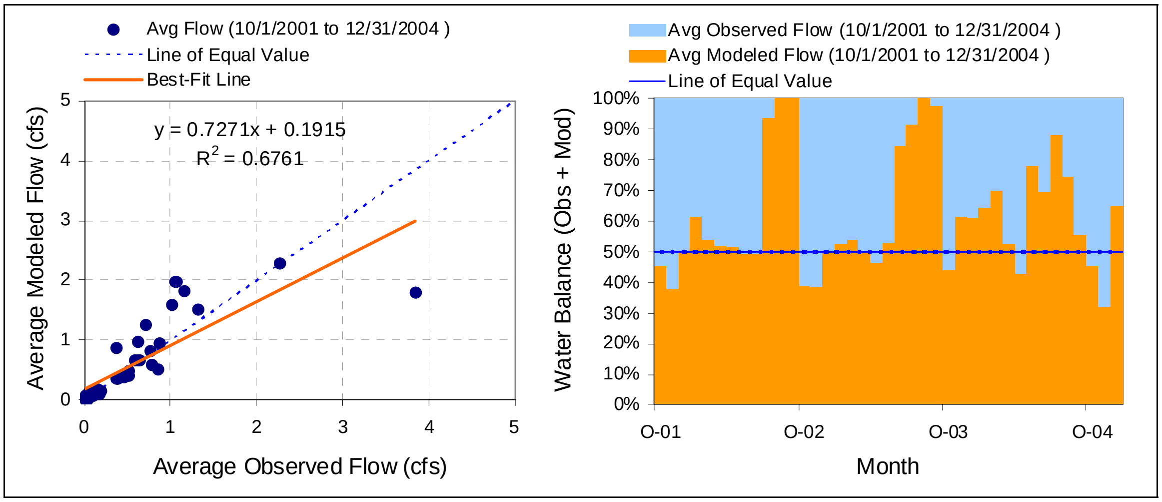

Figure 5. Monthly Flow Regression and Temporal Variation: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

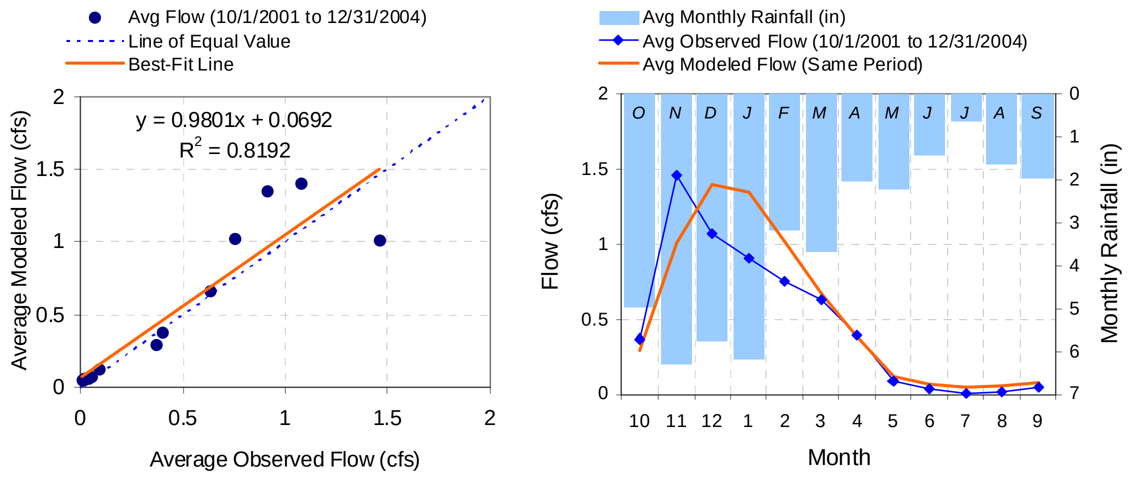

Figure 6. Seasonal Regression and Temporal Aggregate: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

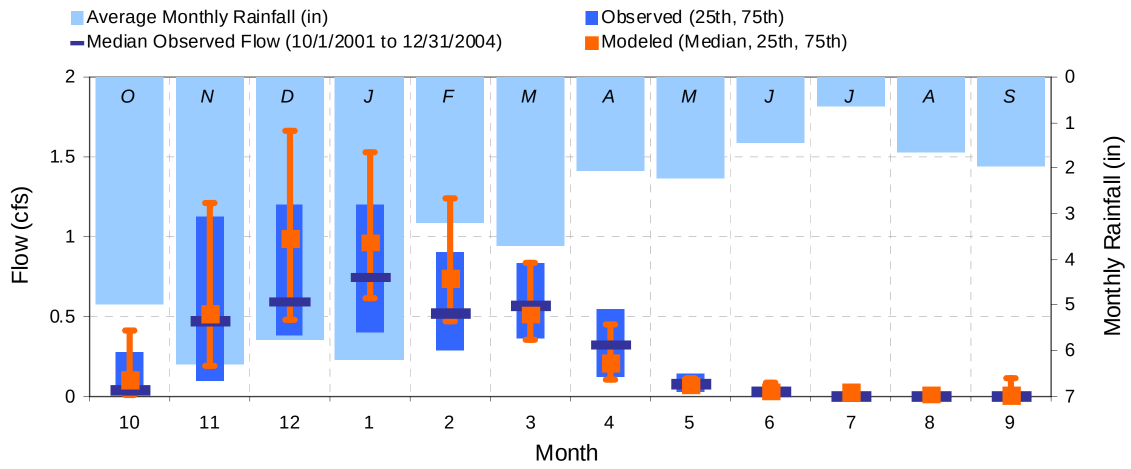

Figure 7. Seasonal Medians and Ranges: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

Table 5. Seasonal Summary: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

MONTH	OBSERVED FLOW (CFS)				MODELED FLOW (CFS)			
	MEAN	MEDIAN	25TH	75TH	MEAN	MEDIAN	25TH	75TH
Oct	0.37	0.04	0.02	0.27	0.30	0.10	0.01	0.41
Nov	1.46	0.48	0.09	1.13	1.01	0.52	0.19	1.21
Dec	1.08	0.60	0.38	1.20	1.40	0.99	0.48	1.66
Jan	0.91	0.75	0.40	1.20	1.35	0.97	0.62	1.53
Feb	0.75	0.52	0.29	0.90	1.02	0.74	0.47	1.24
Mar	0.63	0.57	0.36	0.84	0.67	0.51	0.36	0.84
Apr	0.40	0.33	0.12	0.55	0.38	0.21	0.10	0.45
May	0.09	0.08	0.03	0.14	0.12	0.07	0.04	0.12
Jun	0.03	0.03	0.01	0.05	0.06	0.03	0.02	0.08
Jul	0.00	0.00	0.00	0.00	0.05	0.03	0.02	0.05
Aug	0.02	0.00	0.00	0.00	0.06	0.01	0.01	0.02
Sep	0.05	0.00	0.00	0.03	0.08	0.01	0.01	0.11

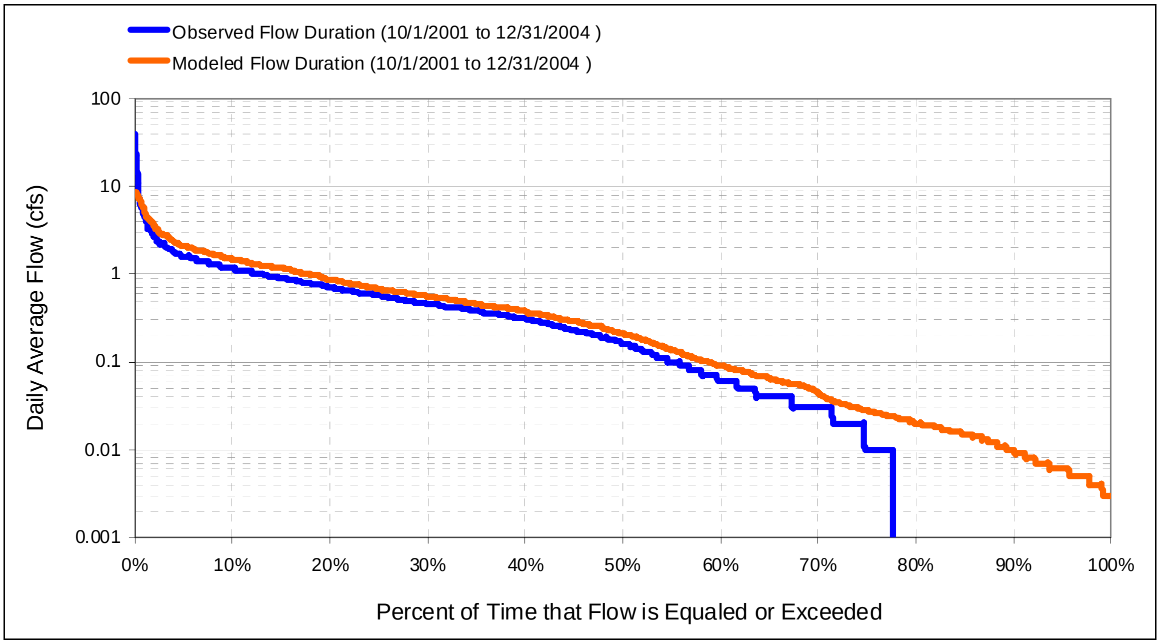

Figure 8. Flow Exceedence: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

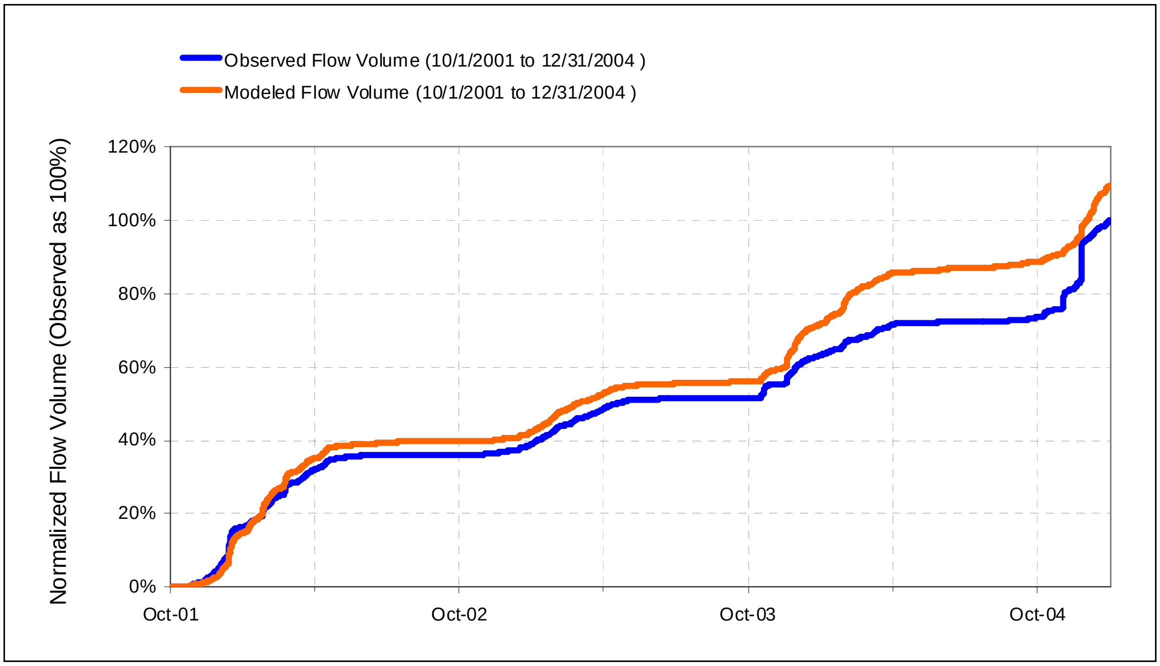

Figure 9. Flow Accumulation: USGS 12202400 Euclid Cr at Euclid Ave at Bellingham, WA

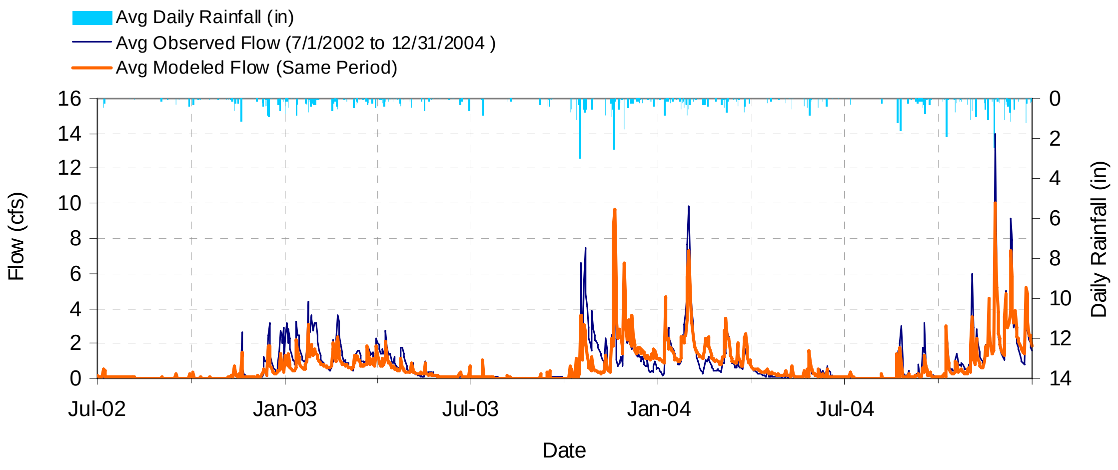

Figure 10. Mean Daily Flow: USGS 12202420 Mill Creek near Bellingham, WA

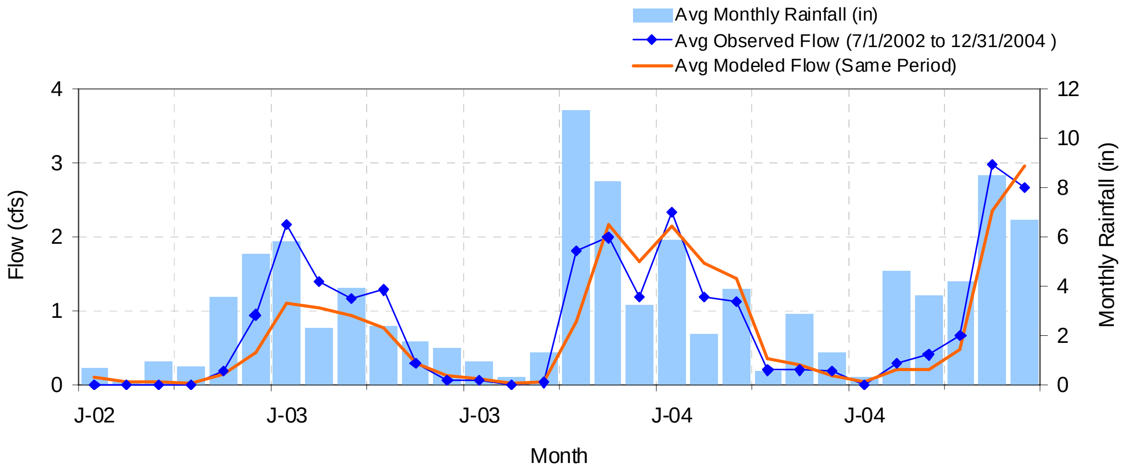

Figure 11. Mean Monthly Flow: USGS 12202420 Mill Creek near Bellingham, WA

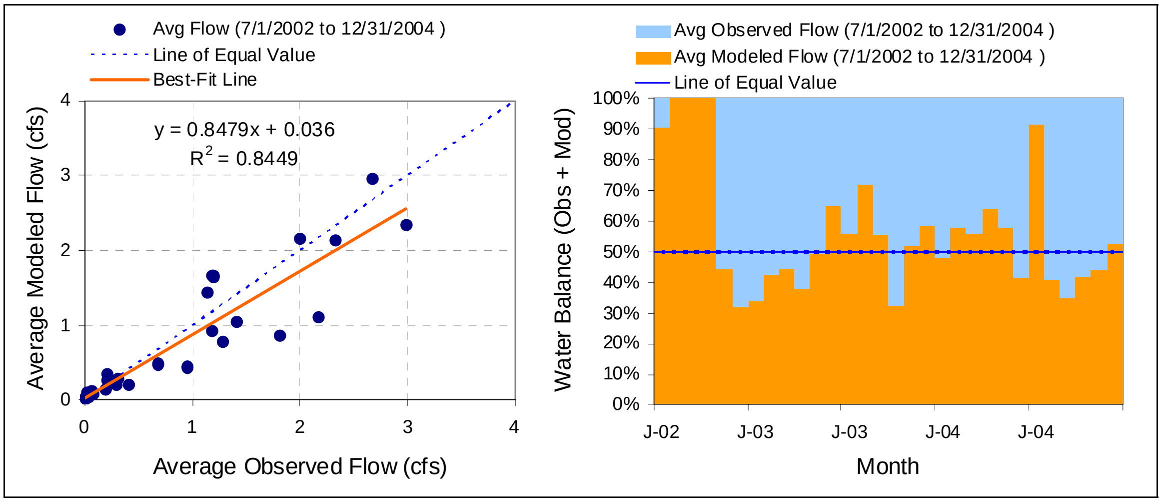

Figure 12. Monthly Flow Regression and Temporal Variation: USGS 12202420 Mill Creek near Bellingham, WA

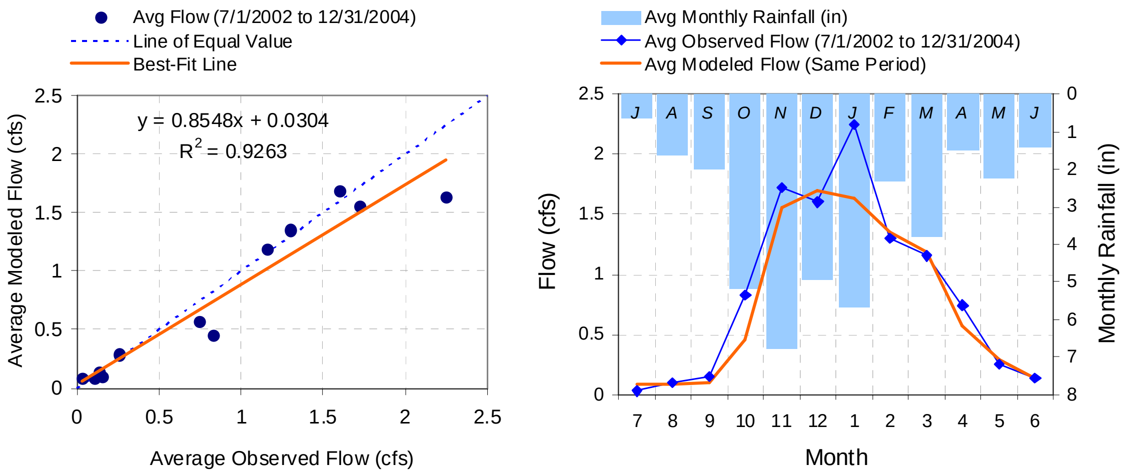

Figure 13. Seasonal Regression and Temporal Aggregate: USGS 12202420 Mill Creek near Bellingham, WA

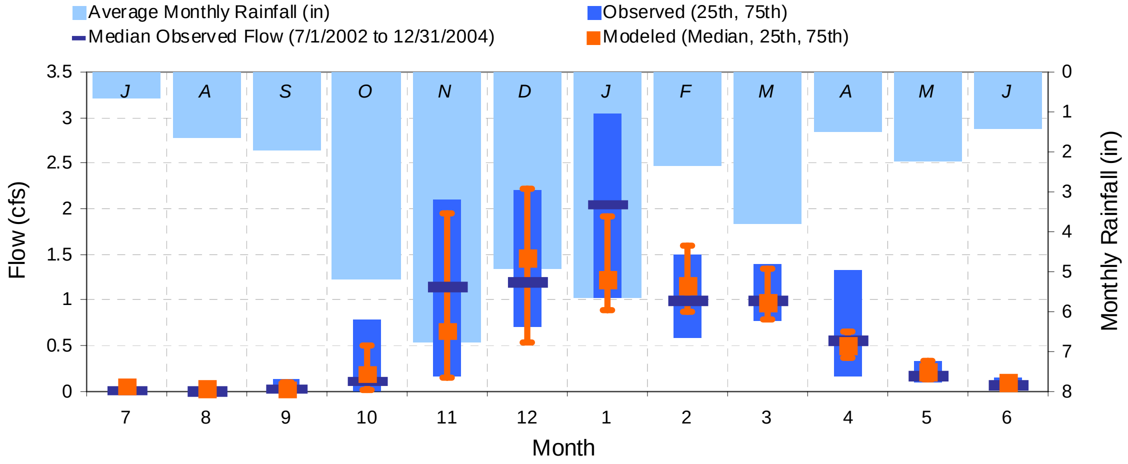

Figure 14. Seasonal Medians and Ranges: USGS 12202420 Mill Creek near Bellingham, WA

Table 6. Seasonal Summary: USGS 12202420 Mill Creek near Bellingham, WA

MONTH	OBSERVED FLOW (CFS)				MODELED FLOW (CFS)			
	MEAN	MEDIAN	25TH	75TH	MEAN	MEDIAN	25TH	75TH
Jun	0.12	0.06	0.04	0.13	0.14	0.09	0.07	0.13
Jul	0.03	0.01	0.00	0.05	0.08	0.06	0.04	0.08
Aug	0.10	0.00	0.00	0.01	0.09	0.03	0.02	0.04
Sep	0.15	0.03	0.00	0.13	0.10	0.02	0.01	0.09
Oct	0.83	0.11	0.00	0.78	0.46	0.19	0.01	0.51
Nov	1.72	1.15	0.17	2.10	1.55	0.66	0.15	1.94
Dec	1.60	1.20	0.71	2.20	1.69	1.46	0.53	2.22
Jan	2.25	2.05	1.03	3.05	1.63	1.22	0.89	1.92
Feb	1.30	1.00	0.59	1.50	1.35	1.16	0.88	1.60
Mar	1.16	1.00	0.78	1.40	1.18	0.97	0.78	1.35
Apr	0.74	0.56	0.16	1.33	0.57	0.50	0.37	0.65
May	0.25	0.17	0.10	0.34	0.29	0.22	0.14	0.33

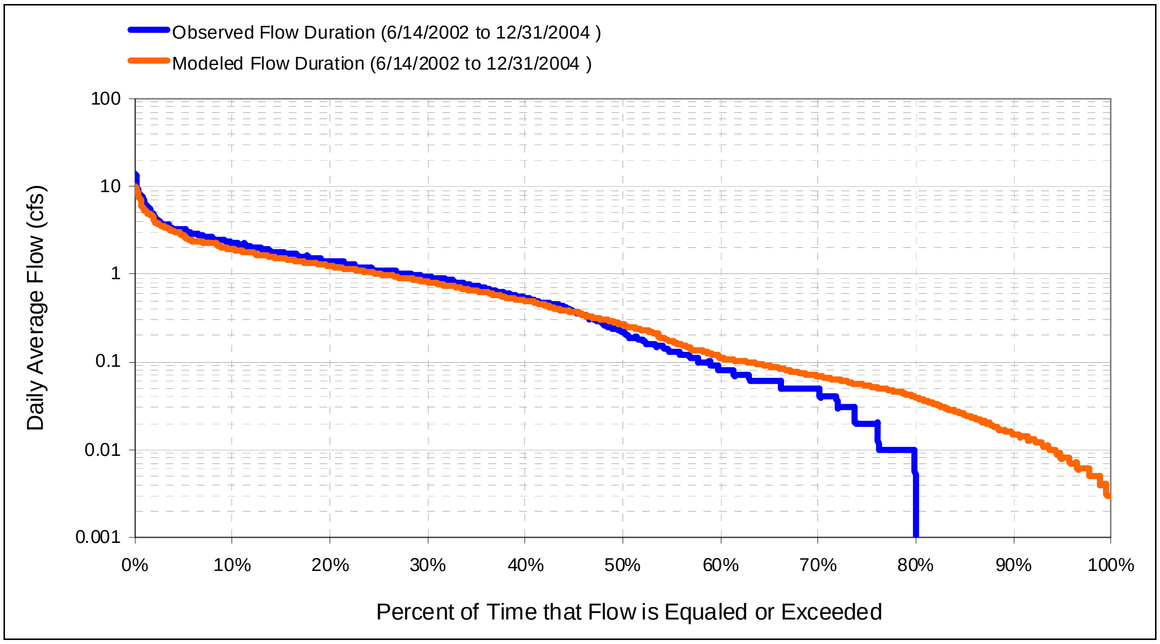

Figure 15. Flow Exceedence: USGS 12202420 Mill Creek near Bellingham, WA

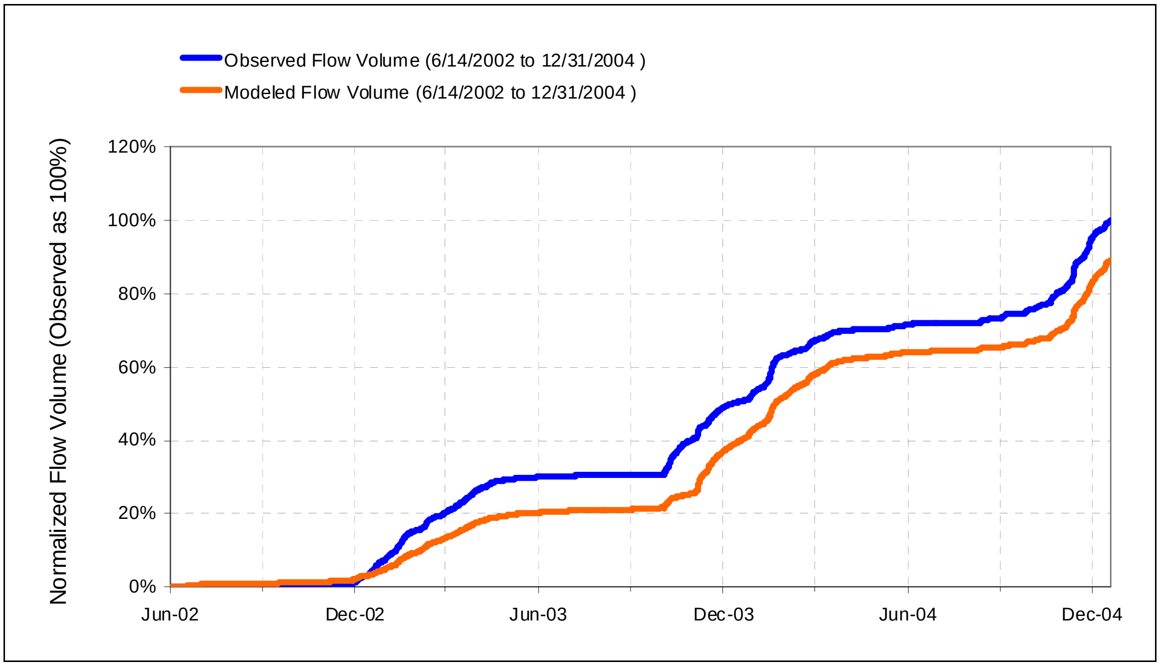

Figure 16. Flow Accumulation: USGS 12202420 Mill Creek near Bellingham, WA

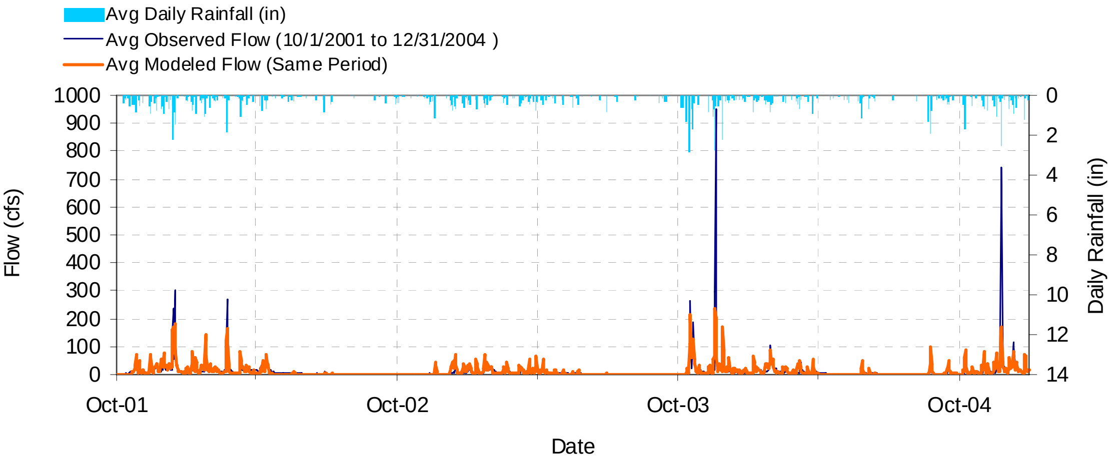

Figure 17. Mean Daily Flow: USGS 12202300 Olsen Creek near Bellingham, WA

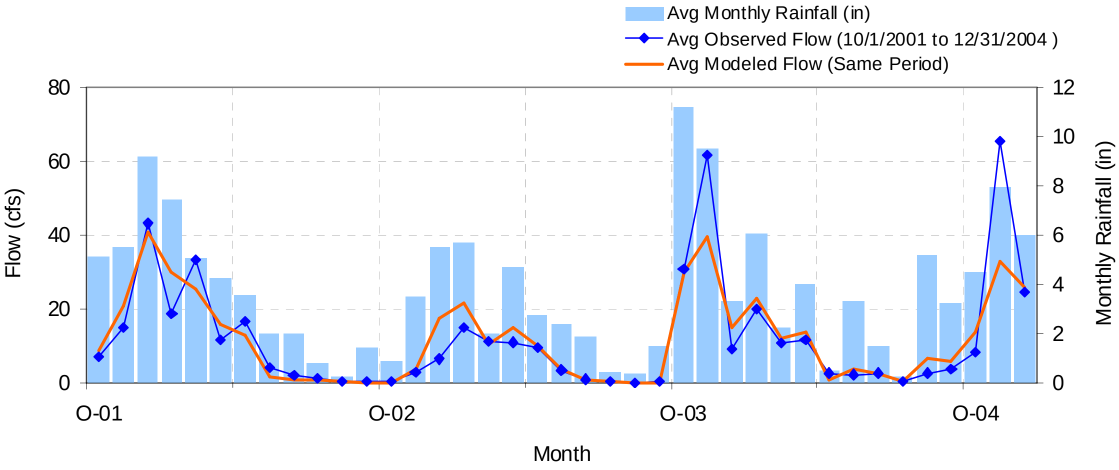

Figure 18. Mean Monthly Flow: USGS 12202300 Olsen Creek near Bellingham, WA

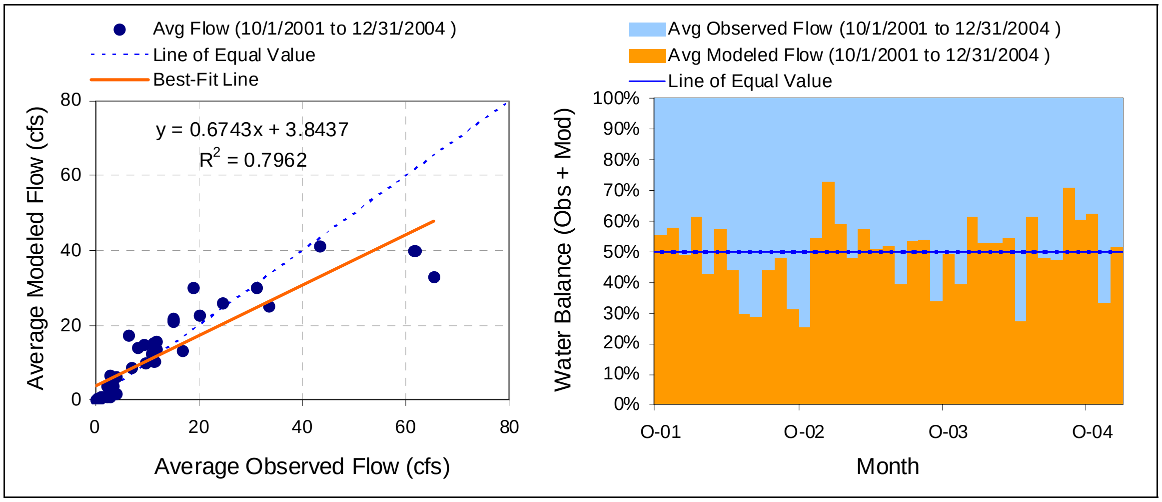

Figure 19. Monthly Flow Regression and Temporal Variation: USGS 12202300 Olsen Creek near Bellingham, WA

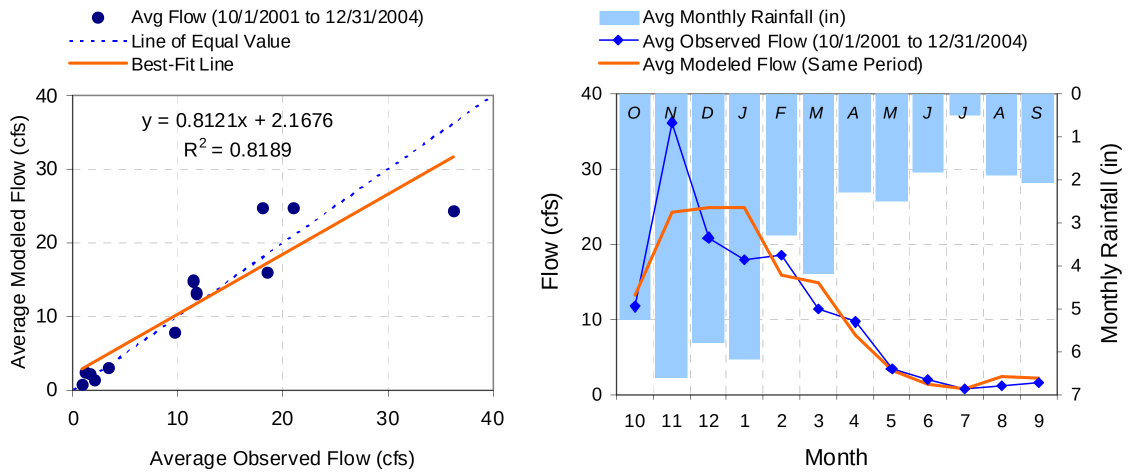

Figure 20. Seasonal Regression and Temporal Aggregate: USGS 12202300 Olsen Creek near Bellingham, WA

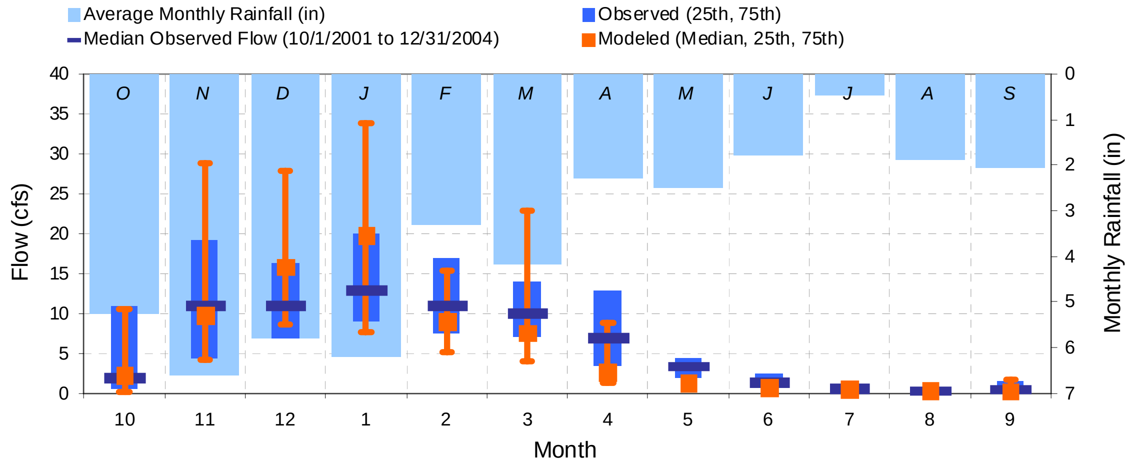

Figure 21. Seasonal Medians and Ranges: USGS 12202300 Olsen Creek near Bellingham, WA

Table 7. Seasonal Summary: USGS 12202300 Olsen Creek near Bellingham, WA

MONTH	OBSERVED FLOW (CFS)				MODELED FLOW (CFS)			
	MEAN	MEDIAN	25TH	75TH	MEAN	MEDIAN	25TH	75TH
Oct	11.71	1.90	0.55	11.00	13.21	2.23	0.19	10.57
Nov	36.26	11.00	4.40	19.25	24.34	9.78	4.24	28.78
Dec	20.92	11.00	6.88	16.25	24.83	15.83	8.56	27.84
Jan	18.01	13.00	9.00	20.00	24.89	19.75	7.63	33.92
Feb	18.51	11.00	7.50	17.00	15.98	8.96	5.13	15.35
Mar	11.45	10.00	7.00	14.00	14.88	7.57	4.01	22.93
Apr	9.69	6.95	3.48	12.75	7.99	2.65	1.23	8.86
May	3.35	3.40	1.80	4.30	3.12	1.24	0.75	1.96
Jun	2.00	1.40	1.00	2.45	1.38	0.65	0.49	1.05
Jul	0.79	0.59	0.43	0.81	0.71	0.58	0.42	0.72
Aug	1.13	0.32	0.23	0.44	2.48	0.30	0.25	0.36
Sep	1.60	0.48	0.32	1.40	2.17	0.22	0.18	1.59

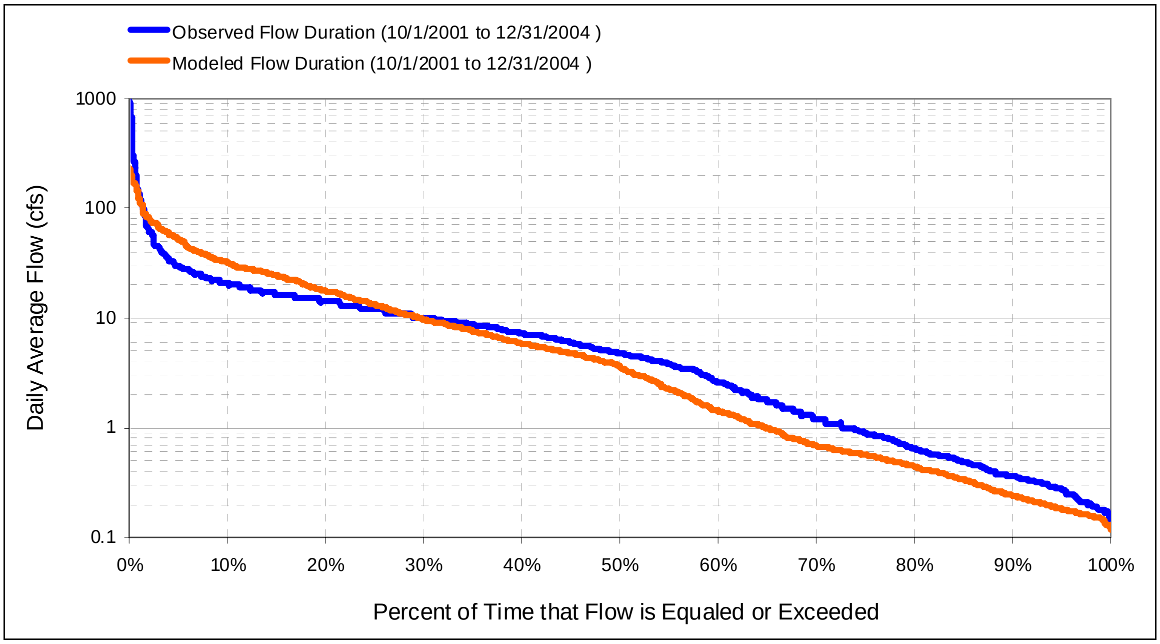

Figure 22. Flow Exceedence: USGS 12202300 Olsen Creek near Bellingham, WA

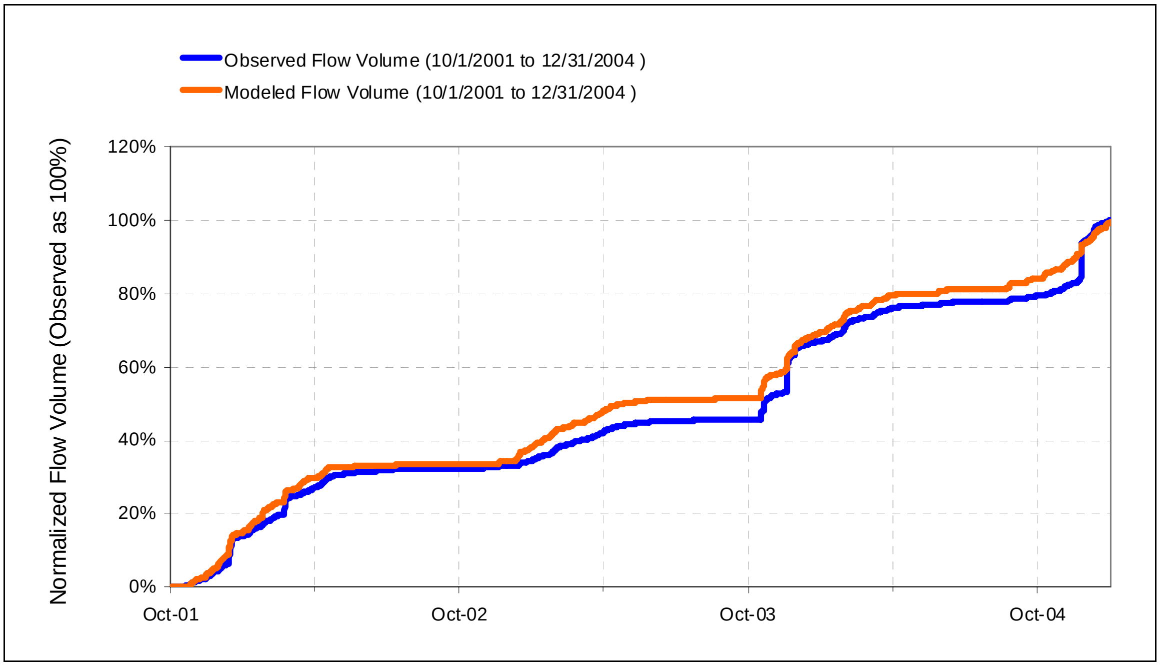

Figure 23. Flow Accumulation: USGS 12202300 Olsen Creek near Bellingham, WA

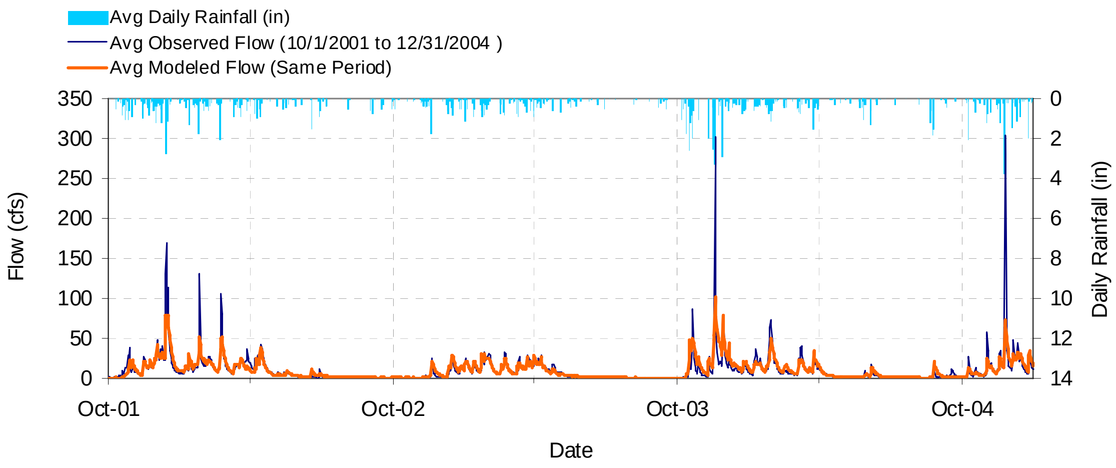

Figure 24. Mean Daily Flow: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

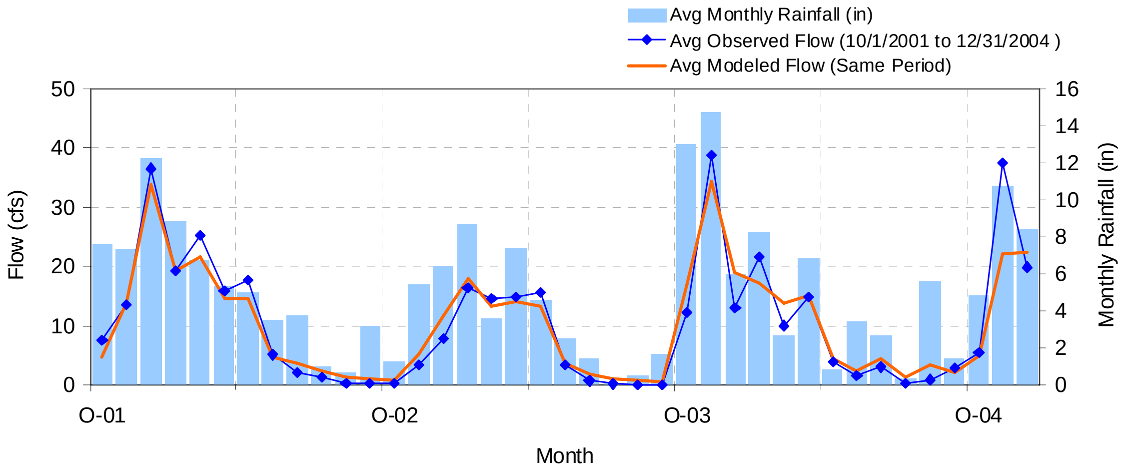

Figure 25. Mean Monthly Flow: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

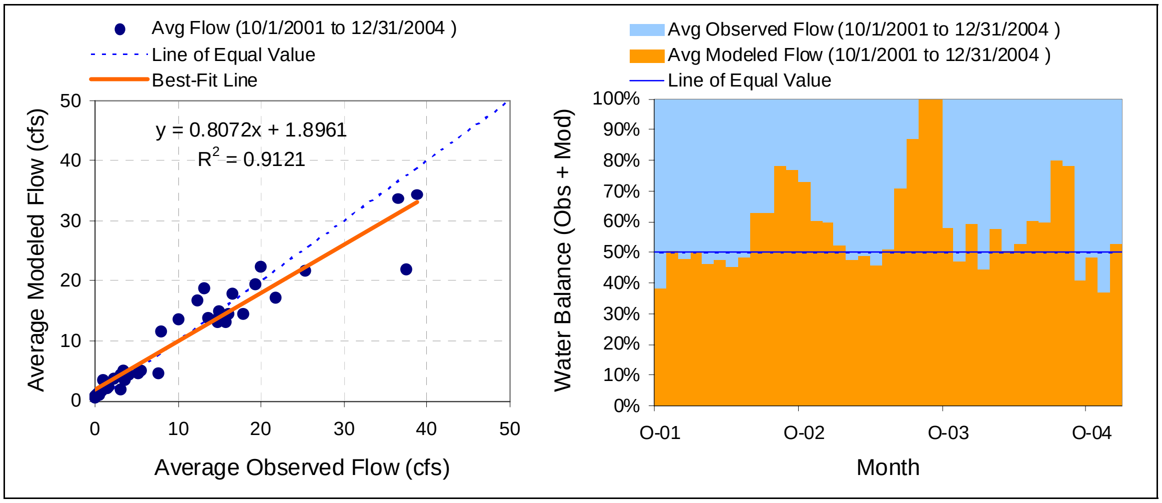

Figure 26. Monthly Flow Regression and Temporal Variation: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

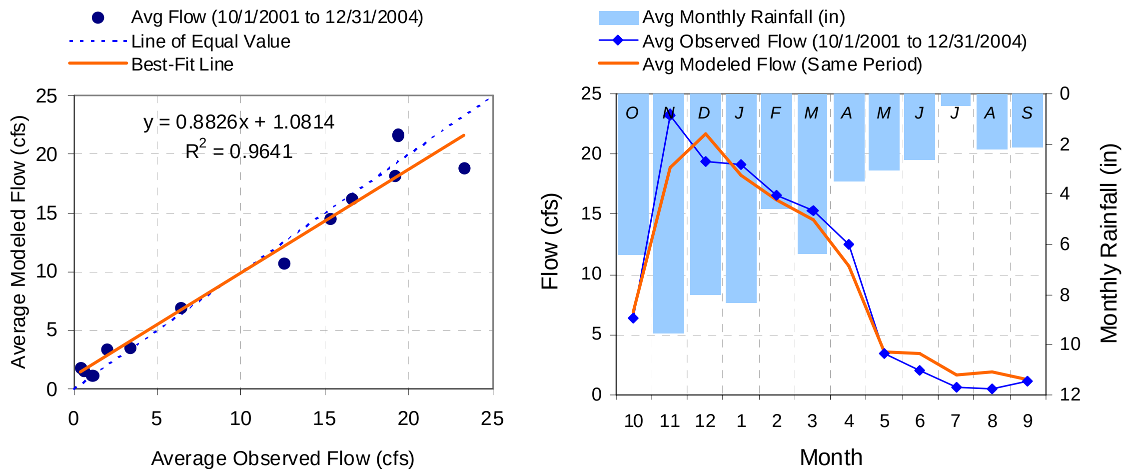

Figure 27. Seasonal Regression and Temporal Aggregate: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

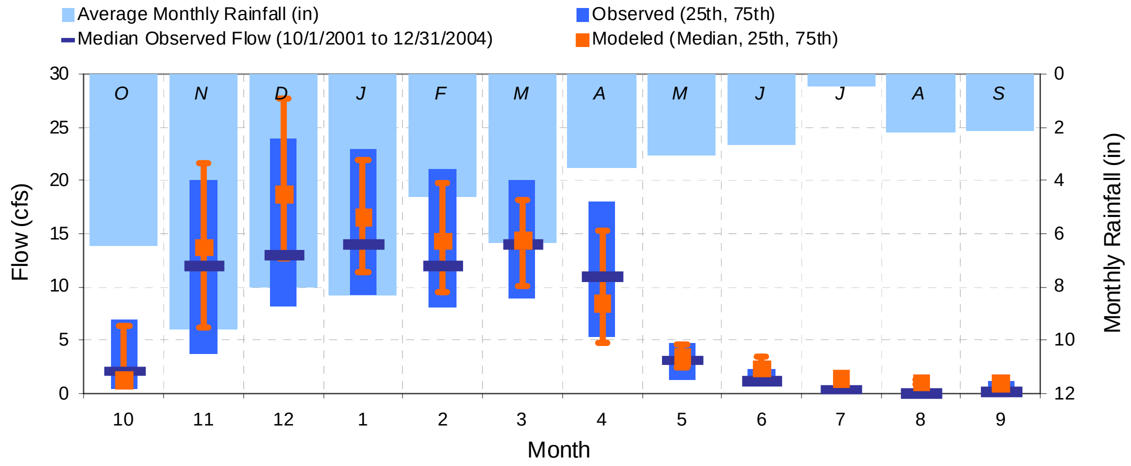

Figure 28. Seasonal Medians and Ranges: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

Table 8. Seasonal Summary: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

MONTH	OBSERVED FLOW (CFS)				MODELED FLOW (CFS)			
	MEAN	MEDIAN	25TH	75TH	MEAN	MEDIAN	25TH	75TH
Oct	6.39	2.15	0.35	6.85	6.91	1.28	0.90	6.28
Nov	23.32	12.00	3.78	20.00	18.92	13.74	6.12	21.68
Dec	19.36	13.00	8.28	24.00	21.69	18.77	12.65	27.76
Jan	19.17	14.00	9.20	23.00	18.23	16.59	11.43	21.86
Feb	16.60	12.00	8.10	21.00	16.22	14.34	9.54	19.80
Mar	15.26	14.00	8.90	20.00	14.58	14.41	10.05	18.22
Apr	12.54	11.00	5.25	18.00	10.77	8.44	4.69	15.24
May	3.38	3.10	1.20	4.80	3.61	3.40	2.44	4.52
Jun	2.00	1.15	0.78	2.28	3.37	2.35	1.83	3.50
Jul	0.62	0.38	0.18	0.65	1.61	1.38	1.16	1.74
Aug	0.44	0.00	0.00	0.38	1.88	1.05	0.88	1.22
Sep	1.10	0.18	0.00	1.18	1.26	0.99	0.70	1.44

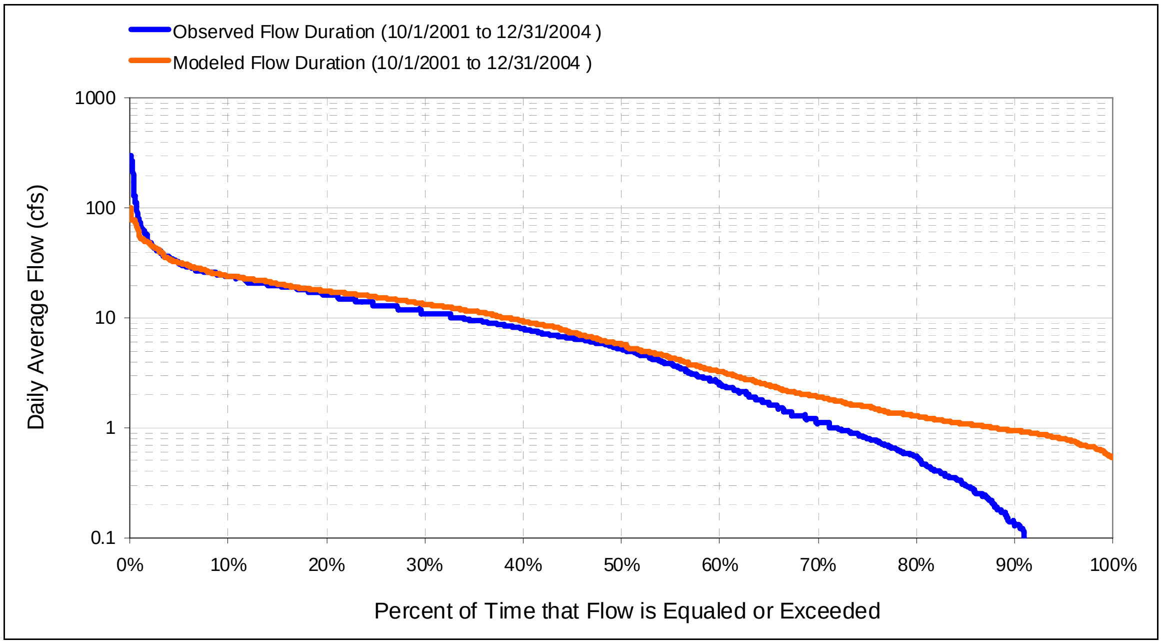

Figure 29. Flow Exceedence: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

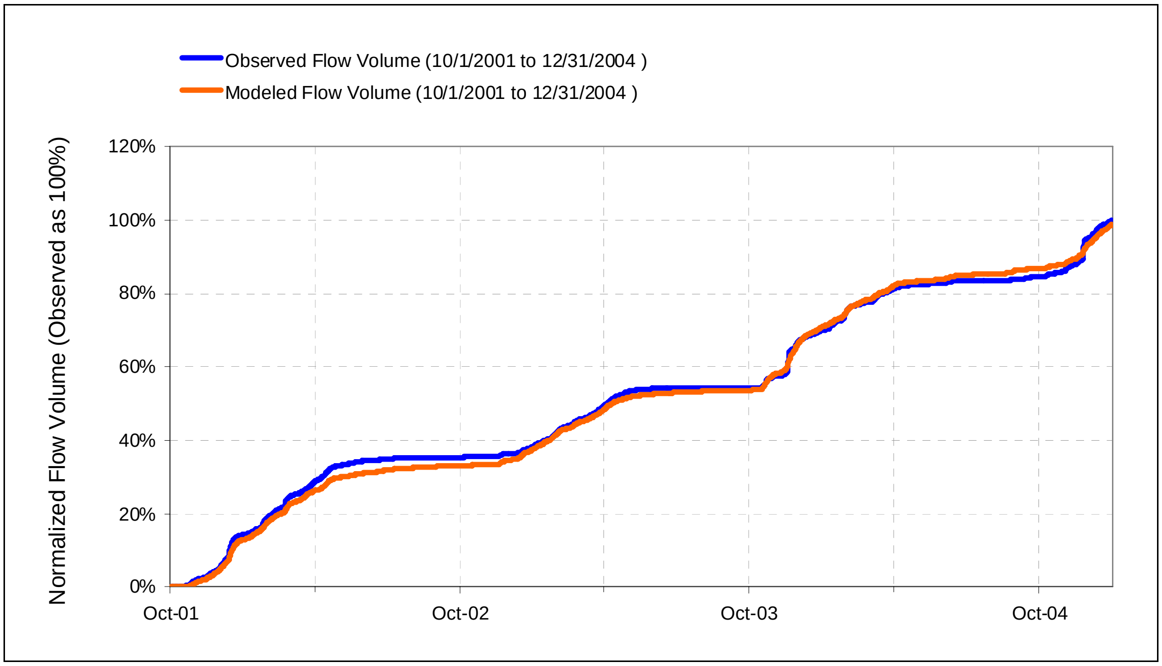

Figure 30. Flow Accumulation: USGS 12201960 Brannian Creek at S Bay Dr nr Wickersham, WA

Do the apparent discrepancies in the hydrologic calibration matter? That very much depends on the intended purposes of the model. The model application does a good job of reproducing total flow volume. Due to the choice of calibration metrics, the model does appear to under-predict storm peaks, while over-predicting summer low flows. In terms of driving the lake model hydraulics, total volume is most important, and the over-prediction of low flows will have little effect on lake model predictions. Under-prediction of stormflow peaks will have some impacts on the representation of mixing events in the lake. More importantly, storm peaks will constitute the majority of the transport of sediment and sediment-associated pollutants (such as phosphorus) into the lake, so under-estimation of these peaks could translate into a low bias on estimates of pollutant delivery. Perhaps of greater concern, inaccuracies in the representation of tributary water balance could bias the attribution of loading sources and associated load allocations for individual tributaries.

1.6. WATER QUALITY MODEL

In contrast to hydrology, the water quality simulation of the Lake Whatcom watershed is a new application, evidently constructed on a shortened schedule. Because the primary concern for the lake is phosphorus loading, the model focuses on simulation of total phosphorus (TP), simulated as a buildup/washoff process. There are two potential problems with this approach, particularly as applies to a largely rural watershed. First, TP consists of a variety of inorganic and organic chemical forms, which might move through the watershed in different ways and exhibit different bioavailability. A simple buildup/washoff representation (which is most appropriate for impervious surface loading) will have inherent limitations. Second, much of the movement of phosphorus is likely to occur in association with the movement of particulate matter (either as orthophosphate sorbed to sediment or as organic detritus). The model is not calibrated for sediment – although doing so could have improved the TP simulation.

There were evidently some problems in representing TP load in the Middle Fork Diversion, estimated to contribute 11 percent of the TP load to Lake Whatcom. This diversion is outside the scope of the watershed model, but needs to be represented for the lake model. The watershed modeling report (p. 2-5) states that the WDOE regression model of TP loads in the diversion was rejected in favor of a constant concentration assumption, which would introduce additional uncertainty into the simulation of TP load. However, the lake model apparently does use the regression model to establish inputs from the diversion (personal communication from Paul Pickett, WDOE, to Jonathan Butcher, Tetra Tech, 15 April 2008).

Results of the TP simulation are compared graphically to observed data through time-series and box and whisker plots. No quantitative evaluation of the fit is provided. While the graphical comparisons generally appear reasonable, such comparisons can also be misleading. The stated procedure of modifying parameters “to improve the goodness of fit with available observed water quality concentrations” cannot be rigorously carried out without use of some quantitative measures of the fit. Further, it is the responsibility of the modelers to provide an estimate of uncertainty in model predictions, which can, at least in part, be developed through quantitative goodness-of-fit measures, particularly when applied to an independent model validation time period. (As noted above, the Lake Whatcom watershed model does not employ a validation test.)

It is noted that watershed models are frequently presented without quantitative calibration results, and data limitations can significantly limit the utility of such an analysis. The need for such measures ties back to the intended uses of the model. The Quality Assurance Project Plan (QAPP) produced for the watershed model lists the primary objectives as determining monthly average and maximum storm event loads at the mouths of the tributaries along the lake. It further states, “applications of absolute criteria for model acceptance or rejection based on rigorous comparisons with the observed data are not appropriate for this effort” due to the high level of uncertainty that was anticipated associated with the observed data. While “absolute criteria for model acceptance” might not be appropriate in this case, some measures need to be used (and reported) to guide the calibration process. For example, average (or median) error and average absolute error could be used to track the precision and bias, respectively, of the model during calibration. Even if numeric acceptance criteria are not used, it would be useful to implement a goal of maximizing precision and minimizing bias in the calibration.

In addition to quantitative goodness of fit measures on concentration, the water quality calibration could consider other evaluations, such as comparison of observed versus predicted load-duration curves and comparison of model-predicted annual loads to statistically derived annual loads. The report does provide a comparison of model results to phosphorus loading predictions from the WDOE regression model – but again the comparison is made only graphically, with no discussion of the evident divergence between the two results during certain periods.

In contrast to the hydrology model, the TP simulation appears to predict a range of storm event concentrations that is greater than is observed. The reasons for and implications of this are not adequately addressed in the report, although the phenomenon is noted.

There are apparently no data with which to evaluate the simulation of loading rates from individual land use classes in the watershed other than the relative differences between individual watersheds. (No subwatershed has more than 5.8% developed land, and no station has more than 3.1% agricultural land.) The differences among subwatersheds are likely due in large part to differences in precipitation and soils, and do not provide a strong foundation for attributing loads to individual land use types. The modelers took the approach of setting buildup/washoff coefficients based on national-scale event mean concentration recommendations (from NURP, etc.), but these national numbers might not reflect conditions in Washington. Additional discussion of this issue should have been provided, preferably with consideration of more local data.

The modelers correctly note that the “watershed model should be considered a work in progress”, but conclude “the model is adequately set up to evaluate potential scenarios within the watershed.” The latter statement is key: Despite the shortcomings in the model calibration noted above, the model application does appear adequate for examination of the relative difference between existing, natural, and full buildout conditions.

2. HSPF to CE-QUAL-W2 Translator

The 30 November 2007 “Amendment to Lake Whatcom TMDL Final Modeling Report” documents the creation of a translator from the HSPF model output to CE-QUAL-W2 input – although applications of the W2 model using the translator do not seem to have been published at this time.

There are several important things to note regarding the translator. The HSPF model produces hourly output of flow and total phosphorus. In contrast, the CE-QUAL-W2 model is set up to use daily input, specified as step functions (un-interpolated) in card TRIB INT. W2 also requires input for multiple state variables, including five species of phosphorus (inorganic PO_4 plus labile and refractory, particulate and dissolved organic phosphorus) as well as other constituents such as temperature, nitrogen, algae, and so on. In addition, the W2 model uses as input the flows and loads from 22 Water Resource Inventory Area (WRIA) subwatersheds, which do not match up exactly with the HSPF subwatersheds. To address these complexities, the translator bases the input files on those derived for the earlier calibrated version of the W2 model (by WRIA), for which the tributary input data were based on regression against observed data. The translator then replaces the data for flow and total phosphorus in these files with the results of the HSPF simulation. To do this, a weighting scheme is employed to distribute the HSPF subwatersheds to the WRIA subwatersheds. Total phosphorus from the HPSF model is apportioned to the five W2 input variables based on the percentages determined in the original tributary input files. These factors have the following implications:

- Although the lake model runs on a 5-minute time step, tributary inputs from the HSPF model have been averaged to daily values. This is adequate for general lake responses to phosphorus loading, but not appropriate to support the model in its current configuration to evaluate responses to sub-daily transients.

- For scenarios, the approach assumes that constituents other than phosphorus are independent of scenario land use (maintaining existing concentrations). This becomes progressively more inappropriate as scenario conditions differ more from calibration conditions. However, this will only cause a problem with the intended uses of the model to the extent that algal response and resulting DO concentrations are sensitive to loads other than phosphorus. The HSPF model is not currently calibrated for other constituents, so this might be the best option available.

- Similarly, the approach assumes that the fractionation of total phosphorus into inorganic and dissolved and particulate, labile and refractory organic components is fixed at the calibration level. These fractions might also change in response to scenarios, but again the HSPF model is not set up to provide these inputs.

The translator is implemented through a complex VBA code. It appears that the developers put a lot of initial effort into making the code flexible, but curtailed this effort at some point due to time constraints. As a result, some portions of the code are general and flexible, while other portions are hard-coded. For instance, when reading in the HSPF output, the code will only read years starting with “200”, and would need to be modified to simulate 1999 and earlier.

3. CE-QUAL-W2 Model

A working copy of the CE-QUAL-W2 model (W2) was also provided for this review, and runs successfully. The W2 model of Lake Whatcom was completed prior to the development of the watershed model, with the major lake model calibration report being delivered in July 2005[6] (updated by memorandum in February 2007 with revised statistics reported in November 2007 and February 2008). In the original model, instead of using watershed model output, “flows for un-gauged watersheds were estimated by calculating the area proportional flow based on a nearby gauged creek.” Pollutant loading time series for monitored tributaries were “developed from data and using regressions fitted to the data” (using regressions developed by WDOE), while “for watersheds that lacked data, the constituent file of a similar watershed was used.” Watershed model output was first used for the November 2007 run.

This sequencing is the opposite of what would ideally occur. It would have been preferable to first develop the watershed model as a means of integrating the observed data and extrapolating to unmonitored watersheds, and then use the watershed model output to drive the lake model.

A file entitled New Model Statistics_2-20-08.doc was supplied, containing model statistics for the November 2007, January 2008, and February 2008 runs. This file does not include any explanatory text, but internal notes in the model input/output (confirmed by personal communication from Paul Pickett, WDOE) show that it represents “Calibration with HSPF tributary flow and TP.” No graphs were provided for these later runs. Comparison of the February 2008 results (using HSPF boundary conditions) to the February 2007 results (using the regression model boundary conditions) shows that prediction of chlorophyll *a* generally improved (better on 14 of 15 statistics, representing three measures at 5 stations), while the DO fit was of similar quality (better on 7, worse on 7, tied on 1 statistic). Unfortunately, substitution of the HSPF flows and loads does not appear to improve the phosphorus simulation: the fit for ortho-phosphorus is worse on 15 of 15 measures, while the fit for total phosphorus is better on 6, tied on 4, and worse on 5 measures. The magnitude of the change, however, appears to be insignificant.

3.1. MODEL PARAMETERS

Parameters used in the calibrated W2 model are generally in line with “typical” values recommended by the model developers. The major differences are in the specifications for the simulation of algae, for which three algal groups (chlorophytes, chrysophytes, and cyanobacteria) are represented. (Note that the

---

[6] Berger, C.J. and S.A. Wells. 2005. Lake Whatcom Water Quality Model. Technical Report EWR-03-05. Prepared for Washington Department of Ecology by Portland State University, Portland, OR.

parameters for algal group 3 are incorrectly described as “algal type 2” in the parameter table.) The temperature parameters controlling optimal growth are all set lower than the typical values, but this is likely appropriate for the latitude and location of Lake Whatcom and is compensated for by the specification of a lower fractional growth rate than default at the lower boundary for the optimal growth range. Algal growth, mortality, and excretion rates are all increased above typical values, which will tend to increase the rate of cycling between inorganic and organic nutrient pools; however, the parameters for stoichiometric equivalent between organic matter and phosphorus are set well below typical values, which might have the opposite effect. Similarly, the light saturation intensity for maximum algal growth is set well below typical values, but the extinction coefficient for inorganic suspended solids is set at 10 times the typical value – again having somewhat of a compensating effect, while changing the shape of the vertical distribution curve for algal density.

Unfortunately, the report provides no discussion of the rationale and process used to develop the calibration. No sensitivity analysis or evaluation of the correlation between parameter estimates is provided (most likely due to long model run times). The generally good quality of fit of the model does, however, suggest that it provides a reasonable representation of the behavior of Lake Whatcom.

3.2. CODE MODIFICATION

The CE-QUAL-W2 model was updated from version 3.2 to version 3.5 between August 2006 and August 2007. Some of these changes were made specifically to enhance the Lake Whatcom model.

The Release Notes for version 3.5 are dated August 27, 2007, and this is apparently the version used in the November 2007 and February 2008 model runs supplied for this review. The model calibration report was, however, last revised in February 8, 2007. Tables of error statistics for water quality parameters were separately updated for the November 2007-February 2008 model runs, but these runs also include substitution of HSPF output for the earlier regression-based tributary boundary files. It is thus not possible to directly discern the impact of model code changes versus changes in boundary conditions for these runs. Changes to the model code after February 2007 included error corrections for sediment heating (when the model added and subtracted layers), CBOD settling, and the logic for algal negative settling. The Whatcom model does not simulate CBOD and does not use negative algal settling, so the impacts of these changes after February 2007 should have had little effect on the Lake Whatcom calibration.

The primary code change that affects the Lake Whatcom model is the addition of variable stoichiometry of sediments (this is covered in the version 3.5 User’s Manual, but not in the Release Notes). The code now calculates the concentration of phosphorus within organic matter in the sediment, accounting for settling inputs of RPOM, LPOM, and algae, as well as epiphyton death (subroutine SEDIMENTP). The subroutine includes a first-order release term, SEDDP. SEDDP is calculated in subroutine KINETIC_RATES (line 9906) as a function of the general sediment decay rate and the current concentration of phosphorus in the sediment. The kinetic term is multiplied by a zero-one variable (DO3) that evaluates to zero unless the oxygen concentration is greater than 1 E-10 mg/L, thus preventing sediment decay under fully anoxic conditions, as was done in previous versions for the decay and oxygen demand of organic matter. Within subroutine PHOSPHORUS (line 10270+) the phosphorus resulting from the decay of sediment organic matter is added to the total inorganic phosphorus flux term, PO4SS (equation B-33 in the manual). PO4SS is equivalenced as one of the vectors in the source/sink matrix, CSSK, and thus included in the general tridiagonal solution for water quality. First order release of nitrogen and carbon from the sediment is handled in an analogous fashion.

The version 3.5 manual gives the following description of phosphorus release:

> *Sediment contribution of phosphorus to overlying waters can be simulated in three ways. In the first, the sediment compartment accumulates particulate organic matter and algae,*

> *which then decay. This is modeled as a 1st-order process. However, sediment phosphorus release depends upon sediment age, chemistry, overlying phosphorus concentrations, and other factors not included in the sediment compartment. In the second, sediments can be assigned a release rate for phosphorus that is independent of sediment concentrations. Sediments are modeled as a "black box" using a zero-order rate. Phosphorus release is only allowed to occur if the overlying water dissolved oxygen concentration is less than a minimum value [O2LIM]. The third method is a combination of the first two where organic materials accumulate and decay in the sediments along with a background decay rate independent of organic matter accumulation in the sediments.*

The statement that release "is only allowed to occur if the overlying water dissolved oxygen concentration is less than a minimum value" is somewhat misleading here. First, the "oxygen-limit" variable (KDO) is not actually a limit; rather it is a half-saturation constant that simulates a gradual change in rates from oxic to anoxic conditions, in this case by multiplying the rate times $\{1 – O_2/(KDO + O_2\}$, such that rates increase with decreasing oxygen concentration. Second, the KDO limitation is applied by the code only to the zero-order release. This means that the new first-order release rate for phosphorus is not dependent on oxygen conditions (except that it shuts off during fully anoxic conditions). Both nitrogen and phosphorus releases are simulated as a function of the user-specified sediment decay rate, the constituent concentration in the sediment, and a temperature rate multiplier. This creates a potential conceptual problem in the formulation, as inorganic phosphorus released from the sediment is prone to co-precipitation with ferrous hydroxide complexes under oxic conditions, while ammonium releases are not. Because there are no user-specified variables to differentiate the rates of inorganic phosphorus and inorganic nitrogen generation by organic matter decay in the sediment, this could introduce an imbalance between sediment source predictions of net phosphorus and nitrogen releases.

Both first-order and zero-order sediment releases of phosphorus can be specified simultaneously in the model. Indeed, this is done in the Lake Whatcom model, although the zero-order release rate (0.0001) is one tenth of the default. One conceptual problem here is that the zero-order release, implemented by the subroutine PHOSPHORUS, is not linked back to the sediment phosphorus accounting, so that the zero-order releases do not deplete the sediment phosphorus stores, potentially resulting in mass-balance errors.

The Lake Whatcom application was set up using an iterative approach to establish initial conditions, in which output from a model run was stored and then used as initial conditions for a subsequent run. This is accomplished using the longitudinal profile (LPR) option for initial conditions, flagged by setting the variable T2I to -2. The code contains a hardwired modification that causes writing of two output LPR files at the end of the simulation (corresponding to the two constituent waterbodies specified for the Whatcom model). The code to accomplish this (starting at line 7450) appears to be correct, with the exception of epiphyton. For epiphyton, the program expects to read sets of data for each branch, for each epiphyton group within the branch, and for each segment within a group-branch set. The output code writes a set of output for an epiphyton group within each branch and then starts again with the next epiphyton group. This does not create a problem for this application because epiphyton are not implemented for Lake Whatcom.

It is also worth noting that the code for writing the LPR output is not affected by user termination ("STOP_PUSHED"); thus the LPR file will be written regardless of whether the program terminates normally or is terminated prematurely by the user, which could lead to some confusion.

In general, the new code appears to perform as intended, and adds flexibility to the model. There are some relatively minor conceptual enhancements that could be pursued, but these appear unlikely to cause any significant problems with the Lake Whatcom simulation.

3.3. MODEL CALIBRATION

The Lake Whatcom model was initially implemented for the period from February 14, 2002 – December 31, 2003 and takes about 20 hours to run. The model start in the middle of February was based on the availability of boundary condition and monitoring data. The start date was later moved back to January 1, 2002, providing two full years of simulation. The HSPF model has now been run for 2000-2005, but the lake model application period does not seem to have been extended.

The model has thus been calibrated and run over two growing seasons, and no separate validation tests have been undertaken. Calibration to a relatively short period incurs some risk, as there is not sufficient leverage to properly resolve all parameter values, and there is no guarantee that results will be applicable to other years with other inflow, outflow, and weather conditions. This occurs in part because the best-fit parameter estimates in a complex model like CE-QUAL-W2 typically exhibit strong cross-correlation so that there might be many sets of parameter values that fit equally well to a single year, but might have different implications for prediction of years with different boundary forcings. In addition, a 1.9 year model run might not provide enough time to damp out the effects of initial condition specifications, such as phosphorus stores in the sediment. The modelers have attempted to address this through iteration of the model. Initial conditions for calibration were set from observed data, but the TMDL application was set up using an iterative approach to model periods longer than two years by rerunning the calibration years using initial conditions from the end of the previous run – but this makes the implicit assumption that lake conditions at the end of 2001 (prior to the start of the 2002 simulation) are similar to those at the end of 2003. No discussion is provided to justify whether 2002-2003 is indeed a representative time period for evaluating the behavior of this lake.

As to validation, current guidance no longer explicitly demands testing with a separate independent validation period, but rather suggests a more general process of model evaluation "to determine whether a model and its analytical results are of a quality sufficient to serve as the basis for a decision."[7] The important point is that some type of analysis needs to be provided regarding the suitability of the model for its intended purposes. A traditional validation test is valuable, when feasible, because (1) it helps determine whether the calibration is robust (or, alternatively, represents over-fitting to limited data), and (2) it provides a direct measure of the predictive ability of the model. If lake data are available outside of the 2002-2003 range, it is recommended that W2 be extended and run to the full period of the HSPF modeling to investigate these issues.

Calibration of the W2 model is evaluated both statistically and graphically in the report. The statistics include mean error, mean absolute error, and root mean square error. The selection of these statistics is not explicitly justified in terms of the specific application objectives of the model, and there does not appear to have been a modeling QAPP. One problem with the selected statistics is that they are scale dependent and do not necessarily reflect the level of uncertainty in the model. Normalized measures of relative mean error and relative mean absolute error should also be provided. As an example, the mean error and mean absolute errors reported for ortho-phosphorus appear quite small (0.002 mg/L and 0.003 mg/L); however, the average ortho-phosphorus concentration seems to be only about 0.004 mg/L – so the relative mean error and mean absolute error are about 50% and 75% respectively (for the 2008 version). For chlorophyll *a* the average observed concentration is 2.36 µg/L, so a mean error of 0.02 translates into a relative mean error of only 0.8%, while the relative mean absolute error of 1.07 is 46%. (Note: Table 8 in the February 2008 results mistakenly labels the chlorophyll *a* results as mg/L, instead of µg/L.) On the other hand, the magnitude of the errors for ortho-phosphorus might be near the quantitation limit of the data (which is not reported), in which case they would not be a cause for concern.

---

[7] Pascual, P., N. Stiber, and E. Sunderland. 2003. Draft Guidance on the Development, Evaluation, and Application of Regulatory Environmental Models. Council for Regulatory Environmental Modeling, Office of Science Policy, U.S. Environmental Protection Agency, Washington, DC.

In fitting a model, it is generally desirable to force the mean error (a measure of bias) toward zero while simultaneously minimizing the mean absolute error and/or root mean square error (measures of precision). For the reported constituents, the average relative mean error appears to be a small percentage of the mean, with the exception of ortho-phosphorus (which might be constrained by the precision of the observed data). Therefore, the model does not seem to contain much system-wide bias. There also does not appear to be a consistent spatial bias, as the signs of mean error in most cases differ between stations LW1, LW2, and Intake (all in the north basin of the lake).

For the hydrology calibration, the model appears to provide a close fit to observed water levels – although this is somewhat misleading as "water levels were calibrated by adding a distributed flow file for branch 1 to compensate for the error in inflow/outflow measurements…" In other words, the modeled water level is forced to an optimal fit to observations by assuming that any discrepancy is due to errors in inflow and outflow measurements. Further, this water balance adjustment cannot properly be called a "hydrodynamic calibration", as no evidence on flow velocities or mixing dynamics is presented. Instead, the quality of the hydrodynamic calibration must be evaluated implicitly based on model ability to represent thermal and pollutant profiles.

The documents provided do not contain an evaluation of water level error statistics using the HSPF model as input to the lake model. It is thus not possible to tell whether the distributed flow correction is appropriate to the linked model application, which replaces gaged inflows with modeled flows. (Any error associated with lake outflows would be unaffected.) It would also be worth giving some thought as to how these "corrections" might be affected by different TMDL scenarios.

The fit for water temperature seems to place the location of the thermocline well and generally captures the onset and breakdown of stratification. The statistics presented for the temperature fit, however, including the R^2 value of 99.6%, are in some ways misleading, as they are based on many individual points (more than 30 in many cases) in vertical profiles, which typically include multiple samples from a typically isothermal hypolimnion. It would have been useful to also report the model's ability to predict temperature at a specific fixed depth (e.g., 1 m).

DO predictions also seem reasonable. They capture the onset of hypoxia well and have mean error and mean absolute error well less than 1 mg/L – one measure of a good-quality fit. As with temperature, the statistics are in a sense inflated by matching to many points in a vertical profile, including multiple hypolimnetic points with approximately steady DO. During some periods (e.g., late 2003), the DO profiles appear to show a spurious peak at around 15 m depth. This likely represents an overprediction of algal production at this depth, perhaps due to under-estimation of light extinction.

Prediction of ortho-phosphorus is much less precise, as noted above, with relative errors on the order of 50 percent. This is not surprising, as the ambient concentrations are quite low, likely affected by method precision and probably controlled by algal uptake. Of greater concern, the predicted vertical ortho-phosphorus profile under stratified conditions sometimes shows a pattern of elevated concentrations in the hypolimnion that is not matched in observed data.

For chlorophyll *a*, the overall model does not exhibit bias (mean error of 0.8%) and the general fit is good, with a mean absolute error of 1.07 µg/L or 46%, and R^2 of 75.3% (for the February 2008 run) – but there are large discrepancies in many individual points (not surprising for chlorophyll *a*). As with the other water quality measures, the comparison is based on multiple individual measurements over an extended vertical profile and the inclusion of multiple low-concentration points from the hypolimnion might inflate the statistics. (Note that a scatter plot was provided for the February 2007 run, but not for the February 2008 run.) The fine-scale vertical distribution of algae is always difficult to predict, and it might be informative to also compare average observed and predicted concentrations over the depth of the epilimnion.

Interestingly, the original (2005) model had a slightly better R^2, but tended to underpredict observed chlorophyll *a* concentrations above 7 µg/L. The 2007 model predicts these high concentrations much more accurately, but at the cost of over-predicting many observations in the 1-4 µg/L range.

4. Summary

The documents provided do not include an overall statement of project purposes and decision needs, introducing some limitations on evaluation of suitability of the modeling products for their intended uses.

The CE-QUAL-W2 lake model application appears to have been carefully developed and provides a credible fit to observed data. Its major shortcoming is calibration to a single time period of less than 2 years without any validation test. Extension of the model to additional years is advisable, if data are available to support this.

In contrast to the lake model, the HSPF watershed model was apparently developed with a limited schedule and budget. While the model was able to make use of an existing HFAM model, there are some shortcomings in the hydrologic calibration, although these appear to be of limited significance to the role of the watershed model in supporting the lake modeling. The water quality simulation is limited to TP. The calibration for TP is evaluated only graphically, without quantitative statistics, and is rightly noted as “a work in progress.” Better results could likely be attained by simulating inorganic and organic phosphorus separately (after calibrating for sediment). In addition, the watershed model is not used to predict nitrogen loading.

Despite these shortcomings, the watershed model does appear adequate to fulfill its stated use of examining the relative differences between existing, natural, and full buildout conditions.

The scope for this review requested evaluation of five general questions, presented below.

1. Does the model appropriately simulate the physical, chemical, and biological processes?

CE-QUAL-W2 is a well-established and tested 2-D simulation model for lakes and reservoirs that appears to be an appropriate choice for Lake Whatcom. The model was enhanced in the course of this project to incorporate variable sediment stoichiometry. All models incorporate simplifying assumptions, some of which have been noted above for the Lake Whatcom application. However, it does appear that the model does simulate all of the most significant physical, chemical, and biological processes that affect nutrient, algal, and DO response in the lake.

The HSPF watershed model provides only a simplified estimate of phosphorus load generation and transport processes in the watershed. The approach simulates TP loading as a buildup/washoff process on all upland types, whereas the movement of inorganic phosphorus from pervious lands is likely tied to sediment loading while organic phosphorus flux is likely tied to organic matter loading, neither of which is calibrated. However, the application appears to be adequate for examination of management scenarios. Most notably, substitution of HSPF output for WDOE regression interpolations of tributary data resulted in little change in the fit of the lake model, suggesting that the HSPF model provides an adequate representation of boundary conditions for the lake simulation.

2. Does the model appropriately incorporate relevant and available data?

Both the watershed and lake models do appear to “appropriately incorporate relevant and available data.” It is not possible, however, to draw firm conclusions here, because neither of the model calibration reports (both of which are rather brief) provides a thorough discussion of the extent of available data.

3. Does the model reasonably predict observed water quality conditions given the available information?

The model appears to do a credible job of predicting observed water-quality conditions in Lake Whatcom. The quality of model fit appears good for this type of application. The main drawback is application to a single application period of less than 2 years, which does not allow evaluation of the performance of the model on other time periods.

4. Were standard modeling procedures and protocols followed?

Pascual *et al.* (*op. cit.*) list the following best practices for model evaluation: peer review of models, QA project planning including data quality assessment, model corroboration, and sensitivity and uncertainty analysis.

Neither the lake model nor the watershed model includes full corroboration tests. In particular, neither contains a validation application. While formal model validation runs are not an absolute requisite for regulatory application of environmental simulation models, such tests are valuable when feasible. At a minimum, additional discussion should be provided as to why the models can be expected to perform adequately outside the data period used for calibration.

Model uncertainty is addressed to the extent that comparisons between simulated results and observations are made. In the case of the watershed model, no quantitative evaluation is provided of model uncertainty for TP prediction. Meaningful sensitivity analyses have not been undertaken. No sensitivity analyses are presented for the lake model, while the watershed model report contains only a brief discussion of model sensitivity to parameters controlling the water balance and pollutant load generation that mostly restates the obvious. A more formal sensitivity analysis would examine how the response of the model can be apportioned to changes in the model's input and would ideally include propagation of uncertainty from the watershed model to the key decision points in the lake model. Such efforts can be difficult and time-consuming with complex environmental models, but are important for understanding the reliability of model scenario results.

Development of the lake model appears to have proceeded without a modeling QAPP. A general QAPP was prepared for the watershed model. However, in neither case is the selection of calibration metrics explicitly justified in terms of the intended decision uses of the models. Despite these shortcomings, both the lake and watershed model appear to have been developed in a reasonable manner by experienced modelers.

5. What are the model's shortcomings? What are the implications of these shortcomings on conclusions drawn from the model?

The paired models provide two key functions: they provide causal interpretation of current conditions in Lake Whatcom and form a basis for evaluation of natural baseline and future management scenarios. The models seem to be on a firm footing with the first function, as they provide a credible representation to observed in-lake data for 2002-2003. Application to scenarios is less certain, due to the lack of validation. In particular, there is likely considerable uncertainty regarding the full build out scenario because the land use for calibration conditions contained little developed land, based developed land parameters on national literature values, and is thus poorly constrained to guarantee accurate representation of loading from developed land in the Bellingham area.

Another shortcoming relative to evaluation of scenarios is that the watershed model is used to predict only flow and TP loading, while other water quality constituents are predicted using WDOE regression equations. As noted above, this might introduce errors for scenario analysis. Whether this is a significant problem or not depends on whether the model results relevant to decision purposes are sensitive to the concentrations of other parameters, such as inorganic nitrogen – a question that might have been answered through a sensitivity analysis.

Simulation models are only approximations of real systems, and all models have shortcomings. As the statistician George E.P. Box noted, “All models are wrong; some are useful.”[8] Despite various shortcomings noted above, the Whatcom models appear credible and useful, although improvements will always be possible. The key question here is whether the models are sufficiently well developed to satisfy their intended decision purposes relevant to developing a TMDL and associated load allocations and wasteload allocations for Lake Whatcom. The models do appear generally suitable for this purpose insofar as lake responses of interest are strongly controlled by flow and TP load. To the extent that responses of interest also depend on other forcing functions (e.g., nitrogen loading, speciation of phosphorus load), the watershed component is incomplete. Notably, concerns about the ability of the model to accurately predict responses to full development are ultimately of limited significance to the TMDL, assuming that current loading rates need to be either maintained or reduced.

---

[8] Box, G. E. P. and N. R. Draper. 1987. Empirical Model-Building and Response Surfaces, p. 424. Wiley.