SWAT-CUP
Calibration and Uncertainty Programs for SWAT

K.C. Abbaspour[1], M. Vejdani[2], S. Haghighat[2]
[1]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstr 133, P.O. Box 611, 8600 Duebendorf, Switzerland
[2] Neprash Gostar Software Development, Tehran, Iran
Email: abbaspour@eawag.ch

Keywords: Hydrologic modelling, calibration, uncertainty analysis, SWAT, SWAT-CUP

EXTENDED ABSTRACT

Distributed watershed models are increasingly being used to support decisions about alternative management strategies in the areas of landuse change, climate change, water allocation, and pollution control. For this reason it is important that these models pass through a careful calibration and uncertainty analysis. Furthermore, as calibration model parameters are always conditional in nature the meaning of a calibrated model, its domain of use, and its uncertainty should be clear to both the analyst and the decision maker. Large-scale distributed models are particularly difficult to calibrate and to interpret the calibration because of large model uncertainty, input uncertainty, and parameter non-uniqueness. To perform calibration and uncertainty analysis, in recent years many procedures have become available. As only one technique cannot be applied to all situations and different projects can benefit from different procedures, we have linked, for the time being, three programs to the hydrologic simulator Soil and Water Assessment Tools (SWAT) (Arnold et al., 1998) under the same platform, SWAT-CUP (SWAT Calibration Uncertainty Procedures). These procedures include: Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992), Parameter Solution (ParaSol) (van Griensven and Meixner, 2006), and Sequential Uncertainty FItting (SUFI-2) (Abbaspour, et al., 2007). In this paper we describe SWAT-CUP and the three procedures and provide an application example using SUFI-2.

Inverse modelling (IM) has often been used to denote a calibration procedure which uses measured data to optimize an objective function for the purpose of finding the best parameters. In recent years IM has become a very popular method for calibration. IM is concerned with the problem of making inferences about physical systems from measured output variables of the model (e.g., river discharge, sediment concentration). This is attractive because direct measurement of parameters describing the physical system is time consuming, costly, tedious, and often has limited applicability. In large-scale distributed applications most parameters are almost impossible to measure as they are lumped and; hence, do not carry the same physical meaning as they did in their small-scale applications. For example, soil parameters such as hydraulic conductivity, bulk density, water storage capacity are but fitting parameters in the large scale. Because nearly all measurements are subject to some uncertainty and the models are only approximations, the inferences are usually statistical in nature. Furthermore, because one can only measure a limited number of (noisy) data and physical systems are usually modelled by continuum equations, no hydrological inverse problem is really uniquely solvable. In other words, if there is a single model that fits the measurements there will be many of them and a large number of parameter combinations can lead to acceptable modelling results. Our goal in inverse modelling is then to characterize the set of models, mainly through assigning distributions (uncertainties) to the parameters, which fit the data and satisfy our presumptions as well as other prior information.

To make the parameter inferences quantitative, one must consider 1) the error in the measured data (driving variables such as rainfall and temperature), 2) the error in the measured variables used in model calibration (e.g., river discharges and sediment concentrations, nutrient loads, etc.), and 3) the error in the conceptual model (i.e., inclusion of all the physics in the model that contributes significantly to the data). The latter uncertainty could especially be large in large-scale watershed models.

1. INTRODUCTION

Definition and quantification of calibration uncertainty in distributed hydrological modeling has become the subject of much research in recent years. Two points must be considered in model calibration, 1) parameter *non-uniqueness*, which means there are (infinitely) many good solutions because of different parameter combinations and, 2) parameter *conditionality*, which means any calibrated model is only conditionally calibrated and cannot be applied in an "absolute sense" to all cases.

Parameter non-uniqueness arises from the fact that IM is inherently non-unique. This is because of the large number of local minima associated with any given objective function.

To draw an analogy, the space of the goal function, *g*, could be likened to a block of "Swiss cheese" with many holes (i.e., a multimodal objective function). Each hole represents a local minimum, with the size of the hole in any direction representing the range of uncertainty at that location.

Figure 1 shows the "Swiss cheese" effect in the response surface of an objective function. In this Figure, for a better visualization, the objective function is inverted so that the minima are represented as peaks.

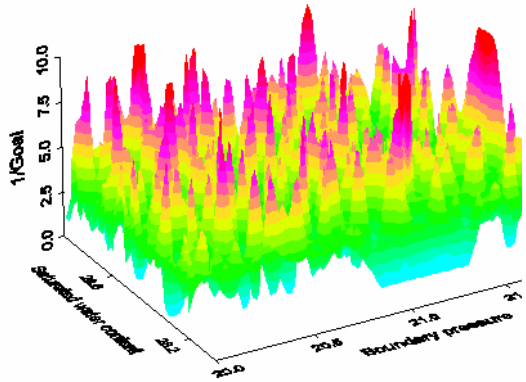

Figure 1. Response surface for an inverted goal function. Local minima are represented as peaks

This Figure shows that for any given objective function, there exists many parameter sets for which the objective values are not significantly different from each other, i.e., there are many potential solutions based on quite different parameter sets. As there are many potential solutions, each parameter optimization routine finds one such minimum for a given objective function. Hence, the search for that one absolute global minimum in hydrologic problems, where the parameters are generally lumped, is not very meaningful. Therefore, as the problem of parameter optimization is not unique, it is important that we define how a model is calibrated and what the magnitude of the prediction uncertainty is.

It has been shown in previous works that "calibrated model parameters" are always "conditional" in nature (Abbaspour et al., 1999). In other words, there is no uniquely defined parameter set, and calibrated parameters are always conditioned on the measured data, simulation model, calibration routine, objective function, etc. Hence, while one calibrated model can produce good discharge results and poor sediment results, another calibrated model based on the same data can produce better sediment and poorer discharge results. Furthermore, it was shown in Abbaspour et al. (2007) that while a watershed model calibrated based on discharge, sediment, nitrate, and phosphorus concentration at the watershed outlet produced good simulations of these variables, it failed to give correct loads from various landuses. After including the landuse loads in the calibration process, the model produced good simulations of the variables at the watershed outlet as well as acceptable loads from landuses. In a further application, the later model, however, still failed to produce correct discharges at other stations inside the watershed. Hence, there are different degrees of calibration as well as different calibrated models for the same project. The applicability of a calibrated watershed model, therefore, is also conditional and a certain "calibrated model" cannot be used for all purposes. When modelling the effect of landuse change, it is therefore essential to verify first if the model gives correct loads for various landuses or not.

As only one calibration/uncertainty analysis procedure may not apply to all situations, we linked several of them in the same platform to SWAT. The objective of this paper is to briefly introduce SWAT-CUP.

2. MATERIALS AND METHODS

2.1. SWAT-CUP

The objectives of SWAT-CUP (SWAT Calibration and Uncertainty Procedures) is to: 1) integrate various calibration/uncertainty analysis procedures for SWAT in one user interface, 2) make the calibrating

procedure easy to use for students and professional users, 3) make the learning of the programs easier for the beginners, 4) provide a faster way to do the time consuming calibration operations and standardize calibration steps, and 5) add extra functionalities to calibration operations such as creating graphs of calibrated results, data comparison, etc. The program is written in C# programming platform. SWAT-CUP has the interface shown in Figure 2 and currently supports programs SUFI2, GLUE, and ParaSol and runs for SWAT versions 2000, 2005. Upon choosing a procedure, the program guides the user step by step through the input files necessary for running each program.

Figure 2. The interface of SWAT-CUP containing program SUFI2, GLUE, and ParaSol for calibration and uncertainty analysis of SWAT

2.2. Model Parameterization

Parameterization of a watershed model is a difficult task as there are a large number of possibilities. For example consider a soil map. Should similar soils in different parts of the region be given the same parameters? There is no reason why this should be so, as the same soil may have different parameters in different places because it is in a different climatic region or under a different landuse or soil management. This may, therefore, result into thousands of parameters; hence, some kind of integration is necessary. The interface linking SWAT to various calibration programs allows parameter aggregation on the basis of hydrologic group, soil, landuse, and subbasin specifications formulated as:

x__<parname>.<ext>__<hydrogrp>__<soltext>__<landuse>__<subbsn>

where x__ is a code to indicate the type of change to be applied to the parameter. For example, v__ means the default parameter is replaced by a given value, a__ means a given quantity is added to the default value, and r__ means the existing parameter value is multiplied by (1 + a given value); <parname> is the SWAT parameter name; <ext> is the SWAT file extension code for the file containing the parameter; <hydrogrp> is the soil hydrological group ('A','B','C' or 'D'); <soltext> is the soil texture; <landuse> is the landuse category; and <subbsn> is the subbasin number, crop index, or fertilizer index. Any combination of the above factors can be used to describe a parameter identifier; hence, providing the opportunity for a detailed parameterization of the system. Omitting the identifiers <hydrogrp>, <soltext>, <landuse>, and <subbsn> allows global assignment of parameters.

The program SWAT-CUP coupling various programs to SWAT has the general concept shown in Figure 3. The steps are: 1) calibration program writes model parameters in model.in, 2) swat_edit.exe edits the SWAT's input files inserting the new parameter values, 3) the SWAT simulator is run, and 4) swat_extract.exe program extracts the desired variables from SWAT's output files and write them to model.out. The procedure continues as required by the calibration program.

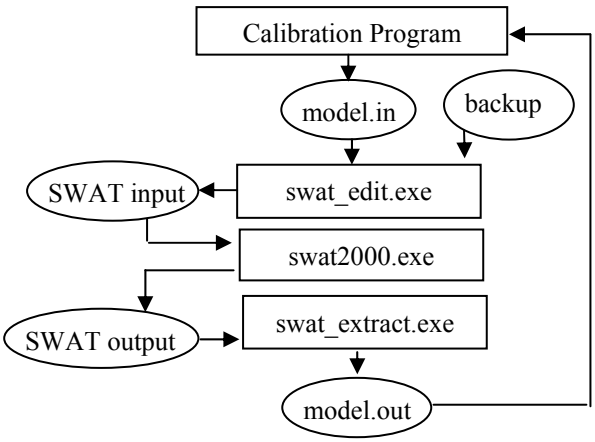

Figure 3. Interaction between a calibration program and SWAT in SWAT-CUP

2.3 Calibration/Uncertainty Analysis Procedures

GLUE. The Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992) was introduced partly to allow for the possible non-uniqueness (equifinality, ambiguity or non-

identifiability) of parameter sets during the estimation of model parameters in over-parameterized models. The procedure is simple and requires few assumptions when used in practical applications. GLUE assumes that, in the case of large over-parameterized models, there is no inverse solution and, hence, that the estimation of a unique set of parameters, which optimize goodness-of fit-criteria given the observations, is not possible. The technique is based on the estimation of the weights or probabilities associated with different parameter sets, based on the use of a subjective likelihood measure to derive a posterior probability function, which is subsequently used to derive the predictive probability of the output variables. In Romanowicz et al., (1994) a statistically motivated, more formal equivalent of GLUE is developed, where the likelihood function is explicitly derived based on the error between the observed outputs and those simulated by the model. This formal approach is equivalent to a Bayesian statistical estimation: it requires assumptions about the statistical structure of the errors. GLUE is usually applied by directly weighting the outputs of multiple model realizations to form a predictive distribution of a variable of interest. Prediction uncertainties are then related to variation in model outputs, without necessarily adding an additional explicit error component. There is thus an interesting question as to whether an appropriate choice of likelihood measure can result in similar results from the two approaches.

There are a number of possible measures of model performance that can be used in this kind of analysis. The only formal requirements for use in a GLUE analysis are that the likelihood measure should increase monotonously with increasing performance and be zero for models considered as unacceptable or non-behavioral. Application-oriented measures are easily used in this framework. Measures based on formal statistical assumptions, when applied to all model realizations (rather than simply in the region of an "optimal" model) should give results similar to a Bayesian approach when used within a GLUE framework (Romanowicz et al., 1994), but the assumptions made (additive Gaussian errors in the simplest cases) are not always easily justified in the case of nonlinear environmental models with poorly known boundary conditions. GLUE can currently support a likelihood measure expressed as the *Nash-Sutcliffe* coefficient.

ParaSol. The Parameter Solution (ParaSol) (van Griensven and Meixner, 2006) method aggregates objective functions (OF) into a global optimization criterion (GOC) and then minimizes these OF's or a GOC using the SCE-UA algorithm. The uncertainty analysis could then be performed with a choice between 2 statistical concepts.

The SCE algorithm is a global search algorithm for the minimization of a single function for up to 16 parameters (Duan et al., 1992). It combines the direct search method of the simplex procedure with the concept of a controlled random search, a systematic evolution of points in the direction of global improvement, competitive evolution and the concept of complex shuffling. In a first step (zero-loop), SCE-UA selects an initial 'population' by random sampling throughout the feasible parameters space for p parameters to be optimized (delineated by given parameter ranges). The population is divided into several "complexes" that consist of 2p+1 points. Each complex evolves independently using the simplex algorithm. The complexes are periodically shuffled to form new complexes in order to share information between the complexes. SCE-UA has been widely used in watershed model calibration and other areas of hydrology such as soil erosion, subsurface hydrology, remote sensing and land surface modeling. It was generally found to be robust, effective and efficient. The SCE-UA has also been applied with success on SWAT for the hydrologic and water quality parameters (van Griensven and Bauwens, 2003).

The type of objective functions used in ParaSol is limited to the sum of the squares of the residuals and the sum of the squares of the difference of the measured and simulated values after ranking (see van Griensven and Bauwens, 2003 for more information).

SUFI-2. The Sequential Uncertainty FItting ver. 2 (SUFI-2) program is similar in concept to GLUE, although it follows a different methodology to obtain posterior parameters from priors. In SUFI-2, parameter uncertainty accounts for all sources of uncertainties such as uncertainty in driving variables (e.g., rainfall), conceptual model, parameters, and measured data. The degree to which all uncertainties are accounted for is quantified by a measure referred to as the *p-factor*, which is the percentage of measured data bracketed by the 95% prediction uncertainty (95PPU). The 95PPU is calculated at the 2.5% and 97.5% levels of the cumulative distribution of an output variable obtained through Latin

hypercube sampling. As all forms of uncertainties are reflected in the measurements (e.g., discharge), the parameter uncertainties generating the 95PPU account for all uncertainties. Breaking down the total uncertainty into its various components is of some interest, but quite difficult to do, and as far as the authors are aware, no reliable procedure yet exists. Another measure quantifying the strength of a calibration/uncertainty analysis is the so called *d-factor*, which is the average thickness of the 95PPU band divided by the standard deviation of the measured data. SUFI-2, hence seeks to bracket most of the measured data with the smallest possible *d-factor*.

The concept behind the uncertainty analysis of the SUFI-2 algorithm is depicted graphically in Figure 4. This Figure illustrates that a single parameter value (shown by a point) leads to a single model response (Fig. 4a), while propagation of the uncertainty in a parameter (shown by a line) leads to the 95PPU illustrated by the shaded region in Figure 4b. As parameter uncertainty increases, the output uncertainty also increases (not necessarily linearly) (Fig. 4c). Hence, SUFI-2 starts by assuming a large parameter uncertainty (within a physically meaningful range), so that the measured data initially falls within the 95PPU, then decreases this uncertainty in steps while monitoring the *p-factor* and the *d-factor*. In each step, previous parameter ranges are updated by calculating the sensitivity matrix (equivalent to Jacobian), and equivalent of a Hessian matrix, followed by the calculation of covariance matrix, 95% confidence intervals of the parameters, and correlation matrix. Parameters are then updated in such a way that the new ranges are always smaller than the previous ranges, and are centered around the best simulation (for more detail see Abbaspour et al., 2007). The goodness of fit and the degree to which the calibrated model accounts for the uncertainties are assessed by the above two measures. An ideal situation would lead to a *p-factor* of about 100% and an *d-factor* near zero. When acceptable values of *d-factor* and *p-factor* are reached, then the parameter uncertainties are the desired parameter ranges. Further goodness of fit can be quantified by the R^2 and/or Nash-Sutcliff (NS) coefficient between the observations and the final best simulation.
If initially a set of parameter ranges cannot be found where the 95PPU brackets most of the data, for example, if the situation in Figure 4d occurs with the parameter uncertainties at physically meaningful limits, then the problem is not one of parameter calibration and the conceptual model must be re-examined. SUFI-2 can currently handle 6 different objective functions (two types of root mean square error, Chi square, Nasch-Sutcliffe, R^2, and bR^2, where b is the slope of the regression line between measured and simulated variable). For a comparison of the above methods in an application to Chaohe River Basin in China see Yang et al., (2007).

In future developments, SWAT-CUP will have more programs for calibration /uncertainty analysis as well as program to perform sensitivity analysis.

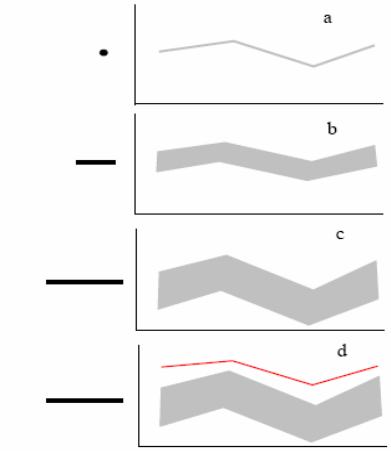

Figure 4. A conceptual illustration of the relationship between parameter uncertainty and prediction uncertainty

3. MODEL APPLICATION

SWAT-CUP was applied to Thur watershed in North-West Switzerland. The watershed was calibrated based on the biweekly measured discharge, sediment, nitrate, and total phosphorous loads at the watershed outlet at Andelfingen station (Figure 5). Water discharge was measured continuously. Concentrations of sediments (suspended solids), nitrate, and total phosphorous in the river water were determined in biweekly composite flow proportional samples. Corresponding biweekly loads were calculated as the product of biweekly average water discharge times concentration.

A constrained objective function was used to ensure correct loads were being simulated for different landuses. The objective function, g, and the constraints were formulated as follows:
Minimize:

$$g=\frac{1}{\sigma^2_{Q_m}}\sum_{i=1}^{130}(Q_m-Q_s)_i^2+\frac{1}{\sigma^2_{S_m}}\sum_{i=1}^{130}(S_m-S_s)_i^2+\frac{1}{\sigma^2_{N_m}}\sum_{i=1}^{130}(N_m-N_s)_i^2+\frac{1}{\sigma^2_{P_m}}\sum_{i=1}^{130}(P_m-P_s)_i^2$$

Subject to:

$0.1 \le S_{Forest} \le 0.3$ (t ha^{-1})

$1.5 \le S_{Agricultural} \le 6$

$2.2 \le N_{Forest} \le 16$ (kg N ha^{-1})

$19 \le N_{Agricultural} \le 47$

$15 \le N_{Pasture} \le 25$

$0.02 \le P_{Forest} \le 0.1$ (kg P ha^{-1})

$0.5 \le P_{Agricultural} \le 2.4$

$0.3 \le P_{Pasture} \le 1.2$

where Q is the average biweekly discharge (m^3 s^{-1}), S is the total biweekly sediment load in the river (t), N is the total biweekly nitrate (NO_3-N) load in the river (kg), P is the total biweekly total phosphorus load (kg), σ^2 is the variance, and m and s subscripts stand for measured and simulated, respectively. In the constraints, $S_{Landuse}$ is the average annual sediment load of the landuse in the watershed (t ha^{-1}), $N_{Landuse}$ is the average annual nitrate load of the landuse (kg N ha^{-1}), and $P_{Landuse}$ is the average annual total phosphorus load of the landuse (kg P ha^{-1}) all in the period of 1991-1995.

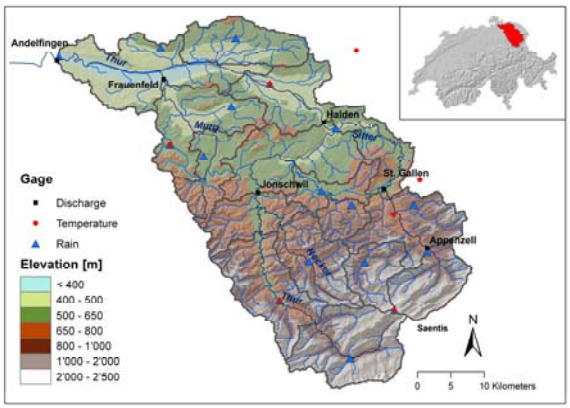

Figure 5. The Thur river basin with SWAT-delineated subbasins, digital elevation model, river network, and meteorological stations. Reproduced with the permission of swisstopo (BA067983)

3.1. Model Results

The Thur model was initially calibrated based on the discharge, sediment load, and nitrate and phosphorus concentrations at the watershed outlet at Andelfingen using the objective function above without constraints. This calibrated model produced excellent results of all variables for calibration and validation periods. However, when loads from various landuses were presented to local experts, they did not quite share our enthusiasm as they were off the range of their long-term measurements and experiences. In a second attempt at calibration we included the range of loads as constraints as shown above in the objective function and re-calibrated the model.

The results are shown in Figures 6-9 for, respectively, daily discharge, sediment, phosphorus, and nitrate loads. For daily discharge, 91% of the observed data is bracketed by the 95PPU (*p-factor*), which is an excellent result. The other measure of the goodness of calibration shown on the Figure is the *d-factor*, which quantifies the thickness of the 95PPU. The smaller this number, the smaller the uncertainties and the better is our calibration work. A value close to 1 is highly desirable for *d-factor* with a *p-factor* also close to 1. The result for biweekly sediment is shown in Figure 7. About 80% of the data were bracketed by the 95PPU and the *d-factor* had a value of 1.5. Most of the data not bracketed, were from the very small sediment loads, while all of the peaks were accounted for. The calibration and validation statistics show larger uncertainties than discharge. Results of the total phosphorus simulation in the river discharge are shown in Figure 8. Similar to sediment, the phosphorus simulation also has larger uncertainties. This is not surprising as much of the phosphorus moved with the sediment. Results of the nitrate simulation are given in Figure 9. Similar to the discharge, the nitrate simulation is also very good with small uncertainties, *d-factor* = 1, while bracketing 82% of the data for calibration and 84% for validation.

4. CONCLUSIONS

SWAT-CUP has a great practical appeal and it can be used for calibration and uncertainty analysis of watershed models.

5. REFERENCES

Abbaspour, K.C., Yang J., Maximov I., Siber R., Bogner K, Mieleitner J., Zobrist J., and Srinivasan R. (2007). Spatially-distributed modelling of hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. Journal of Hydrology, 333: 413-430.

Abbaspour, K. C., M. Sonnleitner, and R. Schulin. 1999. Uncertainty in Estimation of Soil Hydraulic Parameters by Inverse Modeling: Example Lysimeter Experiments. Soil Sci. Soc. of Am. J., 63: 501-509.

Arnold, J.G., Srinisvan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part I: model development. Journal of American Water Resources Association, 34 (1), 73-89.

Beven, K., Binley A. (1992). The Future of Distributed Models - Model Calibration and Uncertainty Prediction. Hydrological Processes, 6(3), 279-298.

Duan, Q., Gupta, V. K., and Sorooshian S. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models, Water. Resourc. Res., 28:1015-1031, 1992.

Romanowicz, R. J., Beven K., and Tawn J. 1994. Evaluation of Predictive Uncertainty in Nonlinear Hydrological Models Using a Bayesian Approach. In: Statistics for the Environment 2, Water Related Issues ,eds V. Barnett and K. F. Turkman, 297-315, Wiley, Chichester.

van Griensven, A. and Meixner T. 2006. Methods to quantify and identify the sources of uncertainty for river basin water quality models. Water Science and Technology, 53(1): 51-59.

van Griensven A. and W. Bauwens. 2003. Multi-objective auto-calibration for semi-distributed water quality models, Water. Resourc. Res. 39 (12): Art. No. 1348 DEC 16.

Yang, J., Abbaspour K.C., Reichert P., Yang H., and Xia J. 2007. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. Journal of Hydrology. In review.

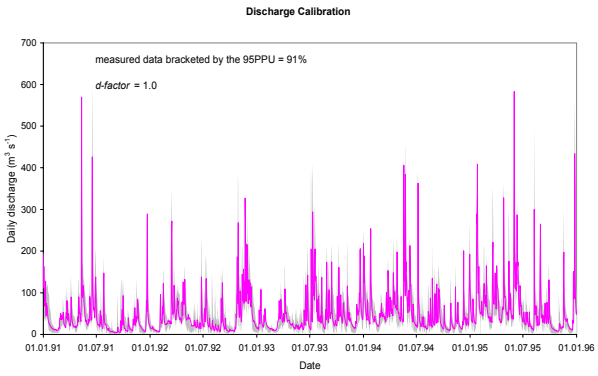

Figure 6. Simulated and observed daily discharges at the watershed outlet.

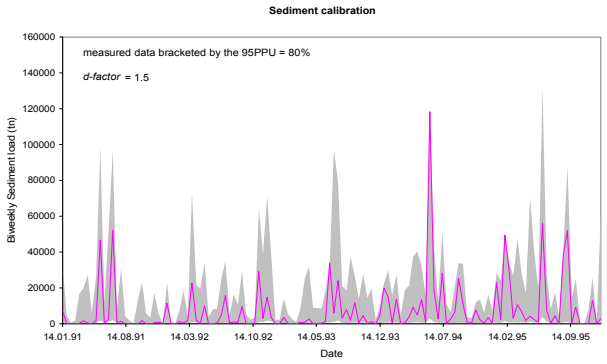

Figure 7. Simulated and observed biweekly sediment loads carried by the river at the watershed outlet

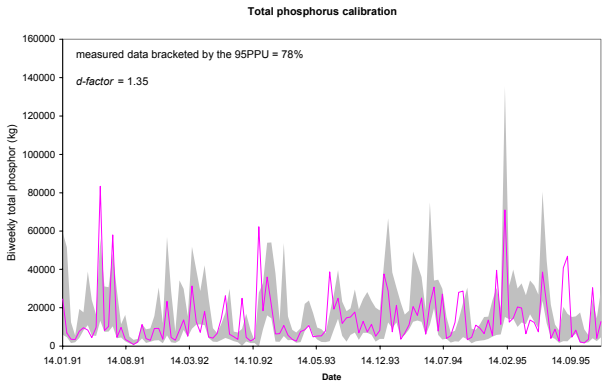

Figure 8. Simulated and observed biweekly total phosphorus loads carried by the river at the watershed outlet

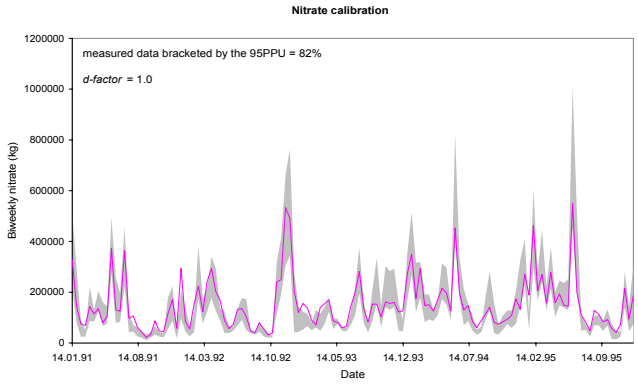

Figure 9. Simulated and observed biweekly nitrate loads carried by the river at the watershed outlet