COMPARISON OF CALIBRATION AND UNCERTAINTY ANALYSIS METHODS: CASE STUDY OF NZOIA RIVER SWAT MODEL

AHMED ADEL
Water Resources Research Institute, National Water Research Center
Egypt

DURGA LAL SHRESTHA, ANN VAN GRIENSVEN, DIMITRI SOLOMATINE
UNESCO-IHE Institute for Water Education
P.O. Box 3015, 2601 DA Delft, the Netherlands
and Water Resources Section, Delft University of Technology

This paper presents comparison of different optimization and uncertainty analysis methods in distributed hydrological modeling. A distributed hydrological model using the Soil and Water Assessment Tool (SWAT) has been built to simulate daily flow in Nzoia River in Kenya. Model parameters of SWAT were calibrated using four different optimization algorithms: parameter solution (PARASOL), adaptive clustering covering (ACCO), genetic algorithm (GA) and multi start (M-Simplex). All generated sets of parameters during the optimization process were analyzed and used to estimate parameter uncertainty. The obtained estimates of model uncertainty were also compared with those obtained with GLUE (generalized likelihood uncertainty estimation) and UNEEC (Uncertainty estimation based on local error and clustering) method. The comparison result shows reasonable level of consistency between different methods. The differences of the uncertainty bounds of the prediction of the stream flow estimated by various methods can be explained by the fact that UNEEC is not a Monte Carlo based method, and the other methods are, and by the differences in sampling strategies of the Monte Carlo based methods.

INTRODUCTION

Increasingly, decision-makers in the field of water resource depend on computer models which are becoming more and more complex. The question may arise to what extent they can trust information from models. These computer models are just a simplified, conceptual representation of a part of the hydrologic cycle used to understand, manage and forecast hydrologic processes. These hydrologic models can be distinguished into knowledge-driven and data-driven modeling. The first type, also called physically-based model uses governing partial differential equations such as Saint-Venant equations for surface flow, Richards’ equation for movement of water in soils, or Boussinesq equation for ground water. Soil and water assessment tool (SWAT) (Arnold and Fohrer [1]) is an example of this type. On the other hand, data-driven modeling uses statistics and machine
learning to treat the hydrological systems as a black box and try to find relationships (possibly, parsimonious ones) between historical inputs and outputs with limited knowledge of the physical behavior of the system. Artificial neural networks (ANN); M5 Model trees are examples of this type of modeling (for the latter, see Solomatine and Dulal [13]). When building a physically based hydrological model, a modeller, after selecting model, collecting required data and estimating parameters, should study the sensitivity of model outputs to changes in parameters, in order to identify the most effective parameters of the model. Then parameters calibration can be carried out for those sensitive parameters and an accepted range of model outputs could be determined using uncertainty analysis (see, e.g., Solomatine et al. [14]; Hall and Solomatine [6]).

Sensitivity analysis, a first step towards model calibration, received a lot of attention, for instance, a comprehensive review of more than a dozen sensitivity analysis methods has been reported by Hamby [7]). Van Griensven et al. [17] presented a combination of Latin hypercube simulation (LH) and one factor at a time (OAT) sampling method (Morris [9]), it was referred to LH-OAT sensitivity analysis method and implemented in SWAT.

Hydrological models conceptualize and aggregate the complex, spatially distributed, and highly interrelated water, energy, and vegetation processes in a watershed through relatively simple mathematical equations with model parameters that often do not represent directly measurable entities and must therefore be estimated using input and output measurements (Vrugt et al. [18]). This process of estimating parameter is called model calibration and it has been received increased scientific interest in the last two decades. For example, Duan et al. [4] presented a global optimization procedure called shuffled complex evolution algorithm (SCE-UA); applications of this algorithm can be found elsewhere (e.g., Duan et al. [4]; Sorooshian et al. [15]). Madsen et al. [8] compared three different automatic calibration methods, and they found that none of those calibration methods are superior with respect to all the performance measures considered. Solomatine et al. [14] introduced adaptive cluster covering (ACCO) algorithm and they reported that ACCO is efficient algorithm for calibration of hydrological model parameters. Genetic algorithm (GA) is another popular global optimization method which is a particular class of evolutionary algorithms. GA was extensively used in the calibration of hydrological model parameters (see, e.g., Franchini [5]).

Indeed, it is almost impossible to obtain a unique optimal value for model's parameters using auto-calibration methods but one has to find a probability distribution of parameters that represents the knowledge of parameter values and the reliability of modelling results must be assessed by estimating their uncertainty. Van Griensven and Meixner [16] developed and applied a new multi-objective optimisation and uncertainty estimation method called parameter solutions (PARASOL); which uses two methods to select “good” results out of simulations generated through calibration process based on an objective threshold; these selected simulations provide the uncertainty bounds on the model outputs.
Generalized likelihood uncertainty estimation (GLUE) is a widely used uncertainty analysis method. It was originally introduced by Beven and Binley [2] and it became widespread in many engineering applications. GLUE is based on the premise that there are many different model structures and many different parameter sets within the chosen model structure that may be acceptable in reproducing the observed behavior of the system. This concept is called “equifinality” (Beven and Freer [3]) and is opposed to the principle of “optimality” stating that it is possible to select an optimal (best) model.

Most of known uncertainty estimation methods study only parameters uncertainty, which deprives from quantify the effect of other sources of uncertainty (e.g., model uncertainty and input uncertainty); hence a new method so called UNEEC (Uncertainty estimation based on local errors and clustering) for quantifying total uncertainty of rainfall runoff models was proposed by Shrestha and Solomatine [11, 12].

In this paper, results of applying different calibration and uncertainty analysis algorithms on case study of Nzoia River in Kenya are presented. We used distributed hydrological model SWAT to simulate daily river flows, and four different calibration algorithms namely PARASOL, ACCO, GA, M-Simplex to calibrate the parameters of this model. All generated sets of parameters during the optimization process were analyzed and used to estimate parameter uncertainty. The obtained estimates of model uncertainty were also compared with those obtained with GLUE and UNEEC methods.

APPLICATION

Soil and water assessment tool (SWAT)

Soil and water assessment tool (SWAT) is used to simulate daily river flow of the considered case study. SWAT is able to simulate very large basins or a variety of management strategies without excessive investment of time or money. Also it is a continuous time modelling tool that enables its users to study long-term impacts. Furthermore, the software is easy to use and it is well documented with very details and clear. Importantly, it is a free software and public domain which makes it easy to use it even in the limited budget projects. SWAT separates hydrological process into land phase and routing phase. Land phase of the hydrologic cycle controls the amount of water, sediment, nutrient and pesticide loadings to the main channel in each subbasin. While routing phase defines the movement of water, sediments, etc. through the channel network of the watershed to the outlet. More details about SWAT governing equations are found elsewhere (see, e.g., Neitsch et al. [10]).

Nzoia River Case Study Watershed Description

Nzoia catchment is in the western region of Kenya where most of this region is highlands on side of the eastern Rift valley. It lies between latitudes 1°30’N and 0°05’S and longitudes 34° and 35° 45’E. The Nzoia River originates from Cherangani Hills at a mean elevation of 2300 m above sea level and drains into Lake Victoria at an altitude of
1000 m. It runs approximately South-West with a catchment area of about 12,900 km² with population more than 3 million. Although Kenya lies in the equator; annual rainfall over most of the country is surprisingly low and rather varied from year to year. The mean annual discharge of the catchment is 1800 x 10⁶ m³. Mean annual rainfall varies from a maximum of 1100 to 2700 mm and a minimum of 600 to 1100 mm. The catchment experiences four seasons in a year as a result of the inter-tropical convergence zone. There are two rainy seasons and two dry seasons, namely, short rains (October to December) and the long rains (March to May). The dry seasons occur in the months of January to February and June to September.

Figure 1. River networks of the Nzoia catchment. On the right shows the physical location of the catchment.

RESULTS AND DISCUSSION

Sensitivity and calibration analysis
Sensitivity analysis was performed before calibration of the model parameters to reduce the number of parameters to be calibrated. LH-OAT method was used to analyze the sensitivity of the model parameters; this method is already implemented in SWAT.
Because we were interested in river flow only, without considering sediment or water quality, sensitivity analysis was carried out using 27 parameters that may have a potential to influence river flow. The prior ranges of the parameters are based on the hydrological knowledge and the recommendations of the SWAT manual. Ten most sensitive parameters were selected for calibration and uncertainty analysis (see Table 1).

Table 1. Comparison of parameter calibration of the SWAT model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Ranges</th>
<th>PARASOL</th>
<th>ACCO</th>
<th>GA</th>
<th>M-Simplex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cn2</td>
<td>35-98</td>
<td>54.205</td>
<td>58.085</td>
<td>63.925</td>
<td>66.557</td>
</tr>
<tr>
<td>Alpha_Bf</td>
<td>0-1</td>
<td>0.916</td>
<td>0.997</td>
<td>0.497</td>
<td>0.648</td>
</tr>
<tr>
<td>Rchrg_Dp</td>
<td>0-1</td>
<td>0.312</td>
<td>0.270</td>
<td>0.233</td>
<td>0.211</td>
</tr>
<tr>
<td>Ch_K2</td>
<td>0-150</td>
<td>12.379</td>
<td>139.779</td>
<td>142.232</td>
<td>145.704</td>
</tr>
<tr>
<td>Surlag</td>
<td>0-10</td>
<td>0.004</td>
<td>1.058</td>
<td>4.714</td>
<td>1.317</td>
</tr>
<tr>
<td>Gwqmn</td>
<td>0-5000</td>
<td>0.058</td>
<td>57.076</td>
<td>332.347</td>
<td>713.928</td>
</tr>
<tr>
<td>Canmx</td>
<td>0-10</td>
<td>9.549</td>
<td>3.536</td>
<td>0.273</td>
<td>6.227</td>
</tr>
<tr>
<td>ESCO</td>
<td>0-1</td>
<td>0.945</td>
<td>0.998</td>
<td>0.637</td>
<td>0.541</td>
</tr>
<tr>
<td>Sol_Awc¹</td>
<td>-0.25-0.25</td>
<td>-0.156</td>
<td>0.236</td>
<td>-0.087</td>
<td>0.033</td>
</tr>
<tr>
<td>GW_REVAP</td>
<td>0.02-0.2</td>
<td>0.060</td>
<td>0.049</td>
<td>0.087</td>
<td>0.062</td>
</tr>
</tbody>
</table>

¹ This parameter is varied relatively with respect to the original value.

Table 1 shows the ranges of the parameter values used to calibrate the SWAT model by all optimization algorithms. Please note that only parameter Sol_AWC has relative change in the value of the parameter during the optimization. Calibrated model parameter values with all optimization algorithms are also reported in Table 1. The results show the considerable difference in the value of the model parameters across the different algorithms, although the model performance is similar (see Figure 2). Figure 2 presents efficiency and performance of the optimization algorithms. It was observed that all optimization algorithms resulted in similar model performances (in terms of Nash-Sutcliff efficiency). Among them, ACCO, GA and M-Simplex algorithms require approximately equal value of function evolutions (about 1000), while PARASOL requires more than 4 times more of these. Figure 3 depicts the simulated hydrographs in a part of calibration data. It is observed that none of the optimization algorithms are able to capture the peak flows.
Figure 2. Comparison of calibration results

Uncertainty analysis
Parameters set and simulations generated during the optimization process are valuable, so they are analyzed and used to estimate parameter uncertainty. This requires the definition of likelihood measure. We used Nash-Sutcliffe efficiency (Beven and Binley, [2]) as the basis for calculating the likelihood measure. All the simulations gathered by ACCO, GA, M-Simplex during the optimization process is assigned likelihood values.

In case of GLUE, we used threshold value of 0.5 (Nash-Sutcliffe efficiency) to reject non-behavioral models. Note that the volume error is about 1.7% (overestimation) with respect to the model calibrated with PARASOL (corresponding Nash-Sutcliffe efficiency of 0.69). However, we have not checked the volume error at the rejection threshold of 0.5. Total number of 1000 behavioral model simulations with a likelihood value greater than 0.5 are retained for the uncertainty analysis. Please note that in case of ACCO, GA,
M-Simplex algorithms, since initial random sampling are only a fraction (less than 10%) of total sampling and they sample over the entire parameter space with a focus of solutions near the optimum, all simulations are retained for the uncertainty analysis. All the retained simulations of runoff are likelihood weighted and ranked to form a likelihood weighted cumulative distribution function of runoff from which chosen quantiles can be selected to represent the parameter uncertainty.

Figure 4. Comparison of 90% prediction interval estimated with (a) PARSOL (b) GLUE, and (c) UNEEC methods for a part of data (7 March, 1977 to 28 January, 1978). Dots indicate the observed flows.

In UNEEC method, fuzzy clustering was used to partition the input space into 5 clusters. We used M5 model tree as a regression function to model the uncertainty. Please note that UNEEC estimates “total” model uncertainty based on model error, while other methods estimate parameter uncertainty and hence the comparison with UNEEC is performed purely for illustration.

Figure 4 shows 90% prediction intervals estimated with PARASOL (top figure), GLUE (middle figure) and UNEEC (bottom figure) for a part of input data. It is observed that the uncertainty bounds estimated by PARASOL are very narrow as compared to the other two. Interestingly enough most of the observed data fall outside the uncertainty bounds. It is important to keep in mind that the PARASOL only addresses the parameter
uncertainty as determined by the data availability. Other sources of uncertainty such as input uncertainty, structural problems or system variability that may not fully be captured by the observed period are not considered.

Figure 5 shows two statistics namely PICP (prediction interval coverage probability) and MPI (mean prediction interval). The former statistic measures the probability that the observed values lie within the estimated prediction intervals. The latter statistic computes the average width of the prediction intervals and gives an indication of the model uncertainty (Shrestha and Solomatine, [11, 12]). The best uncertainty analysis method is that one which covers more observed data inside the prediction interval with narrow bound. The results show the significant differences among the methods in estimating these two statistics. As mentioned previously, PARASOL gives very low PICP and MPI. On the other hand, ACCO produces PICP very close to the desired degree of confidence (i.e., 90%) with very high value of MPI.

![Figure 5 Comparison of uncertainty measure statistics between different uncertainty analysis algorithms](image)

**CONCLUSIONS**

This paper presents the experiments with different calibration and uncertainty estimation in distributed hydrological models. Soil and water assessment tool (SWAT) is used to simulate daily flow on a case study of the Nzoia River in Kenya. Sensitivity analysis was useful for reduction of the number of model parameters to be used for the calibration process. Four optimization algorithms were considered. All parameters and simulations generated during the optimization process were analyzed and used to estimate parameter uncertainty. The estimated model uncertainty are also compared with those obtained with GLUE (generalized likelihood uncertainty estimation) and UNEEC (uncertainty estimation based on local error and clustering) method.

The comparison between optimization algorithms was based on their ability to match the modelled discharge to the observed data and computational cost. Nash-Sutcliffe efficiency was used to measure models fitness and the number of the model runs
(function evaluations) was considered as the indicator of the computational cost. The comparison results show that ACCO algorithm was the fastest method; it was able to achieve reasonably good objective function value (however not the best one compare to the other methods) in about 1000 times of model runs. On the other hand PARASOL achieved marginally better objective function value but required more than 4 times more function evaluations.

Two uncertainty statistics were used to compare the uncertainty estimation algorithms; the first one is the amount of observed flow that falls inside of 90% prediction intervals (i.e. PICP) and the other one is the mean width of those prediction intervals (i.e. MPI). As these prediction intervals are wider, the decision making will be safe, but with higher uncertainty. On the other hand, when prediction intervals are narrower, decision making will be more risky and unreliable. Decision maker should be satisfied if around 90% of observed data fall within the prediction intervals with minimum width of the prediction interval. The comparison between methods shows that some of the methods (e.g. ACCO, and UNEEC) were apple to achieve reasonable amount of observed data within the prediction interval. While PARASOL gives very low value of coverage probability which suggests that this method may not be used for the uncertainty analysis. Pure random sampling like GLUE may not be suitable for uncertainty analysis of distributed model due to heavy computational demand. The comparison result shows that there are considerable differences of the uncertainty bounds estimated by various methods, and at the same time there is certain consistency in their predictions. It is worth stressing that these differences are inevitable – they are explained by the approach to uncertainty analysis, differences in sampling strategies, and the types of the handled uncertainty.

Acknowledgment
The work described in this manuscript was partly supported by the European Community’s Sixth Framework Program through the grant to the budget of the Integrated Project FLOODsite, Contract GOCE-CT-2004-505420.

REFERENCES


