

Step wise Multi-criteria Performance Evaluation of Rainfall-Runoff Models using WETSPRO

P.C. Shakti^{1a,b}, N.K. Shrestha² & P. Gurung³

¹ *Interuniversity Program in Water Resources Engineering (IUPWARE), Katholieke Universiteit Leuven and Vrije Universiteit Brussel, Belgium.*

^a *Graduate School of Life and Environmental Science, University of Tsukuba, Tsukuba, Japan.*

^b *National Research Institute for Earth Science and Disaster Prevention(NIED), Tsukuba, Japan.*

E-mail: shakti@hotmail.com

² *Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, Brussel Belgium.*

³ *Hydro Solution Pvt. Ltd, Swayambhu, Kathmandu, Nepal*

ABSTRACT

This paper illustrates a methodology to evaluate model's performance of rainfall runoff model using a tool called WETSPRO (Water Engineering Time Series PROcessing tool). Simulated results of physically based semi-distributed model - SWAT (Soil and Water Assessment Tool) for Kliene Nete watershed (581 km²), Belgium are considered in this study. Paper presents a series of sequential time series processing tasks to be performed to evaluate model's performance thoroughly. The problem of serial dependence and heteroscedasticity is addressed and model performance evaluation on different flow components (peak flows, low flows and volume) and flow volume is carried. Performance evaluation of both flow components on their extremes is also performed. Two most commonly used goodness-fit-statistics (Mean Square Error – MSE and Nash Sutcliff Efficiency – NSE) are used with number of complementary graphical plots for evaluation propose. Results indicated model's robust performance on peak flows although base flows are slightly underestimated especially for lower return periods. Cumulative flow volumes tend to be overestimated. Based upon the study, some recommendations are summarized to enhance model's ability to simulate the flows events.

Keywords: Rainfall runoff model, SWAT, WETSPRO, Kliene Nete, peak flows, low flows.

1. INTRODUCTION

Different types of rainfall-runoff models are being used by many research institutions, government organizations and other water related sectors for decision support systems. These models have to be calibrated and thoroughly validated. However, calibrating a rainfall runoff model is not an easy task due to large number of model parameters

involved especially in case of distributed rainfall runoff model (Willems, 2009). Consideration of several parameters while calibrating the rainfall runoff models tend to complicate the model calibration task and that would sometimes be highly time consuming (Dmitri et al., 2006). Rainfall runoff models have been calibrated using either a heuristic approach or using auto-calibrated algorithms. Heuristic approach might

be quite time consuming but it uses the experience and knowledge of modeler; while auto-calibrated routines are fast but might return the so called 'optimal' parameters based on the local optimum instead of global optimum. So, combined use of both approaches might result quick and near global optimal parameters values to match the field condition. Moreover, obtaining 'optimal' parameter values is also a case specific. Using a more general model structure would make the models over-parameterized and hence not identifiable. Nevertheless, calibrating rainfall runoff model sometime requires a step wise calibration scheme as described by Willems (2010) especially for a lumped conceptual model. Traditional approaches to validate the calibration of rainfall runoff models are based on some statistical indicators, for example the Nash Sutcliffe Efficiency – NSE (Nash and Sutcliffe, 1970). However it is difficult to characterize the different aspect of model performance of a particular rainfall runoff model with only one or two statistical indicator. The use of multi-objective set of preferably independent statistics is needed and consideration of supporting graphical criteria to evaluate how well the model has simulated to an events.

To support performance evaluation of a rainfall-runoff model, a Water Engineering Time Series PROcessing tool (WETSPRO) has been developed by Prof. Patrick Willems, Katholieke Universiteit Leuven, Belgium. The WETSPRO can be used to conduct flow filtering (quick and slow flow) using a numeric digital filter, to extract peak-over-threshold (POT) values and to construct different model evaluation plots (Willems, 2009). The WETSPRO follows a step wise multi-criterion performance evaluation methodology with the use of different supplement graphs. It makes the use of observed and simulated results separating them into different flow components. It also makes sure of using independent variables by extracting

nearly independent POTs. On the other hands, it also allows comparing cumulative volumes as well as model performance on more extreme events. More description about WETSPRO can be found on Willems (2009).

2. SCOPE OF THE STUDY

In present days, many rainfall-runoff models are extensively using for hydrological analysis worldwide as well as many statistical tools are developed to evaluate the model results. The performance evaluation of model results is mandatory for application in the field of water resources management. Consequently, main objective of this study focus on performance evaluation of the SWAT model results using statistical techniques, taking into account serial dependence of the model output flows and heteroscedasticity in the model residuals. Though the study is based on Belgium, it presents a methodology on how to evaluate model's performance on better way; making it applicable in any catchments in the world.

3. DATA AND METHODOLOGY:

Performance evaluation on the model results of a mid-sized Kliene Nete Watershed, 581km², Belgium (Figure-1) developed on Soil and Water Assessment Tool (SWAT) is illustrated in this study. SWAT is a physically-based continuous hydrological model that predicts the impact of land management practices on water, sediment and agricultural chemical yields in complex basins with varying soils, land use and management conditions (Arnold et al., 1998; Green et al., 2008; Neitsch et al., 2002, Srinivasan et al., 1998). The model simulation results of the Kliene Nete Watershed obtained (Shrestha et al., 2010) with resulted NSE values of 74 and 67% for the calibration (1/1/1994-12/31/1998) and validation

period (1/1/1999-12/31/2002) respectively is used for this study. Observed discharge data series at

Grobbendonk (outlet) with daily frequency is used for the evaluation process.

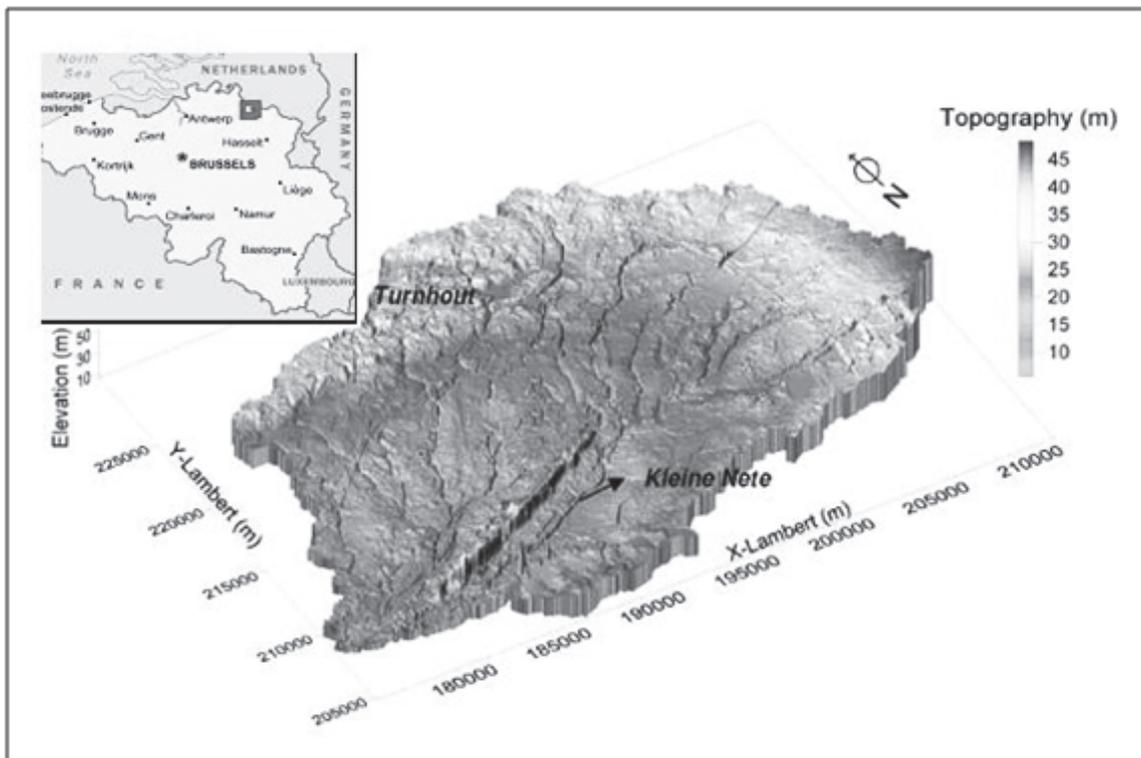


Figure-1: Topography of the Kleine Nete Watershed, location of the watershed is shown in the inset, (Dam et al., 2009).

There are various performance evaluation indicators that mentioned in many literatures. Among all the statistical tools, the Mean Square Error (MSE) and the NSE are two most frequently used statistical parameters for performance evaluation of simulated results. The MSE is a measure that considers the average random discrepancy between observed river flow discharge – $q_0(i)$ and simulated river flow discharge – $q_m(i)$, and is given as:

$$MSE = \frac{\sum_{i=1}^n (q_m(i) - q_0(i))^2}{n} \quad (1)$$

Where, $q_0(i)$ and $q_m(i)$ = the observed and simulated river

discharge, respectively and,

i = number of observations (1, n).

Of course, closer to zero MSE value is the ultimate target in case of model performance result, but due to much uncertainty in the model it is difficult to get desire value. The NSE is used to quantitatively describe the accuracy of model outputs and predicative accuracy and is given as:

$$NSE = \left(1 - \frac{\sum_{i=1}^n (q_m(i) - q_o(i))^2}{\sum_{i=1}^n (q_o(i) - \bar{q}_o)^2} \right) = \left(1 - \frac{MSE}{S_{q_o}^2} \right) \quad (2)$$

Where,

\bar{q}_o a $s_{q_o}^2$ and are the mean and variance of observed discharge series.

But, it has been observed that the model residuals typically increase with higher flow values (Willems, 2009). That means higher flow values receives more weight in evaluation of statistics (1) and (2) making them more sensitive to higher flows. Hence, lower value of MSE and higher NSE can be observed provided higher peaks are nicely matching even if the low flows (base flows) are not matching fairly. The situation is likely to be compounded because of the squared terms in the derivation of MSE and NSE. Hence, the flow values should be transformed by applying suitable transformation before they are used for evaluation of statistics such as (1) and (2). The Box and Cox transformation (Box and Cox, 1964) allows in such cases. The simplicity of Box and Cox transformation is that it has only one parameter ' λ ' and this transformation is used for this study. The Box and Cox transformation given by:

$$BC(q) = \begin{cases} (q^\lambda - 1) / \lambda, & \text{if } \lambda \neq 0 \\ \ln(q), & \text{if } \lambda = 0 \end{cases} \quad (3)$$

Where,

q = Observed or simulated discharge values and,
 λ = Parameter of Box-Cox transformation.

The Box-Cox transformation was applied to both observed and simulated river flow discharges in the determination of the MSE and hence NSE so as to have homoscedasticity (variances of the sampled elements in rainfall runoff modeling is not varying with varying ' q ') in model residuals. The parameter λ needs to be calibrated making the model residual homoscedasticity.

Also, when series of observations are used, model residuals often have a serial dependence – one flow event is dependent on previous event. This dependence will be higher for a series with small time step is considered (Willems, 2009). For our case, as the minimum time step of SWAT is one day, there will be no serial dependence for overland

flow (quick flow is divided into overland flow and interflow) because of small catchment with response time of around 8 hours. But for baseflow, the serial dependence is likely to affect due to longer recession time of base flow. This will cause problems on evaluating goodness-fit-statistics (1) and (2) as the events should be independent. This problem is addressed by extracting nearly independent POT values (*Figure-2*). The two subsequent events are considered as nearly independent if the following three conditions are fulfilled:

- the time length ' τ_s ' of the decreasing flank of the first event exceeds a time ' k_p ':

$$\tau_s > k_p \quad (4)$$

- the discharge drops down – in between the two events – to a fraction lower than ' f ' of the peak flow:

$$\frac{q_{\min}}{q_{\max}} < f \quad (5)$$

Or, close to the baseflow q_{base} :

$$\frac{q_{\min} - q_{\text{base}}}{q_{\max}} < f \quad (6)$$

- the discharge increment $q_{\max} - q_{\min}$ has a minimum height q_{lim} :

$$q_{\max} - q_{\min} > q_{\text{lim}} \quad (7)$$

Thus, according to the procedure of selecting nearly independent POTs has three parameters namely; k_p , f and q_{lim} . The parameter ' k_p ' can be taken equal to recession constant of quick flow or higher. Similarly, parameter ' f ', in equation (5) can be taken as the upper limit of base flow fraction in the peak flow. If equation (6) is considered, then it can be taken as 5% to 15%. And, the parameter ' q_{lim} ' can be taken as the upper limit of small noise peaks which needs to be avoided to be selected as POTs. It is obvious that sub-flow filtering prior to POT selection will make easy to select the parameters of POT selection for example ' k_p ' and ' f '. The sub-flow separation on WETSPRO is based on a generalization of the recursive digital

filter proposed by Chapman (1991) with a new filter parameter 'w' that represents the case-specific average fraction of the quick flow volumes over the total flow volumes. It is to be noted that the original Chapman filtering has a parameter 'k' which is the recession constant 'k' of the sub-flow to be separated. By this way the sub-flow filtering has two parameters for each sub-flow.

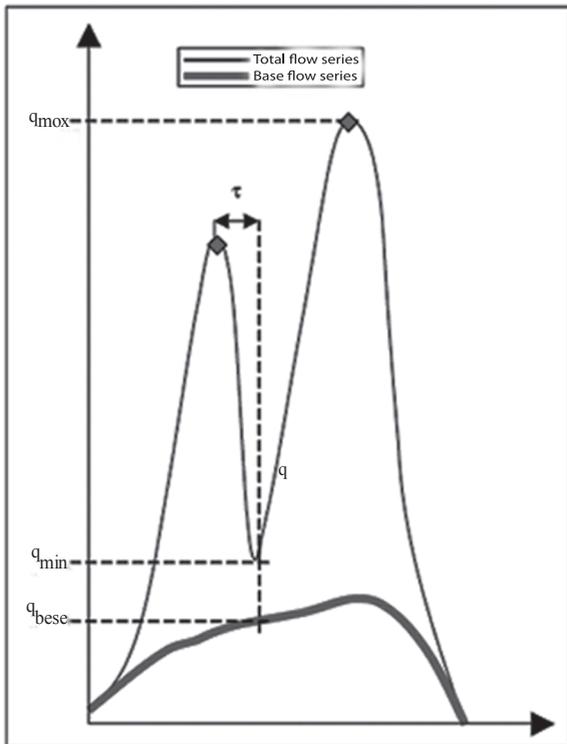


Figure-2: Parameters used in the criteria to select nearly independent POTs (Willems, 2009)

Following are some equations (depicted from Willems (2009)) after the parameter 'w' has been introduced.

$$f(t) = a_1 f(t-1) + a_2 (q(t) - a q(t-1)) \tag{8}$$

$$b(t) = q(t) - f(t) = a b(t-1) + a_3 (1-a) (f(t-1) + f(t)) \tag{9}$$

Where

$$a_1 = \frac{(2 + v) - v}{(2 + v - v a)} \tag{10}$$

$$a_2 = \frac{2}{(2 + v + v a)} \tag{11}$$

$$a_3 = 0.5 v \tag{12}$$

$$a = \exp\left(-\frac{1}{k}\right) \tag{13}$$

$$v = \left(\frac{(1-w)}{w}\right) \tag{14}$$

With,

$q(t)$ = Total flow in time 't' [M^3T^{-1}]

$b(t)$ = Slow flow component in time 't' [M^3T^{-1}]

$q(t)$ = Quick flow component in time 't' [M^3T^{-1}]

k = Recession constant [T]

w = Case specific fraction of quick flow volumes over the total volume [-]

4. RESULT AND DISCUSSION

4.1 CALIBRATION OF λ

Figure-3 is the plot of model residuals (difference between observed and simulated discharge) and simulated discharges. It is clear that the model residuals showed heteroscedasticity as more residual variance can be observed for higher discharges. To have homoscedasticity in model residuals, a value of 0.25 for λ has chosen with heuristic approach. After the transformation, the model residuals evenly distributed along 'zero-residual' line (thick black horizontal line).

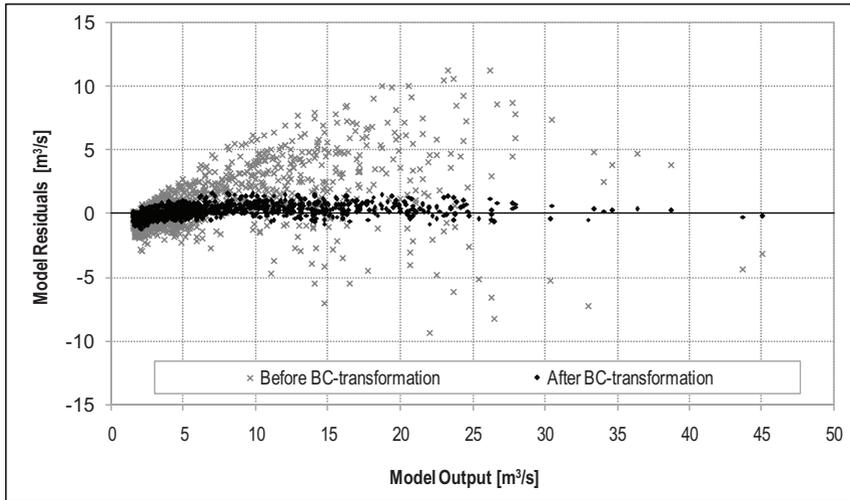


Figure-3: Plot of SWAT model residuals with simulated discharges

4.2 FLOW FILTERING

Filtering parameters ‘k’ and ‘w’ for baseflow separation have been chosen as 80 days and 0.43 respectively and same for interflow are 4 days and 0.4 respectively. Hence around 57% (1 - 0.43) of the total flow is contributed by the baseflow which is quite obvious for relatively flat catchment having high filtering because of having predominant sandy soil. Estimation of those

parameters has been done by visual inspection of the filter result plots especially the recession limb of the hydrograph. *Figure-4* is used to estimate the baseflow recession constant by making the slant dotted line as far as possible, parallel to the trend of recession of long dry period. *Figure-5* is the plot of interflow filter result along with filtered base flow for initial 1500 to 2000 days. It is to be noted that the days are counted as 0 (1/1/1994) to 3287 (12/31/2002).

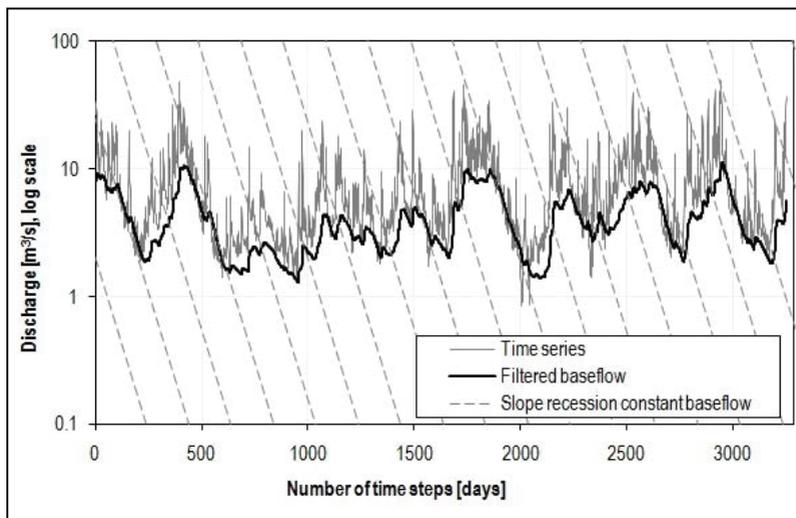


Figure-4: Base flow filter result

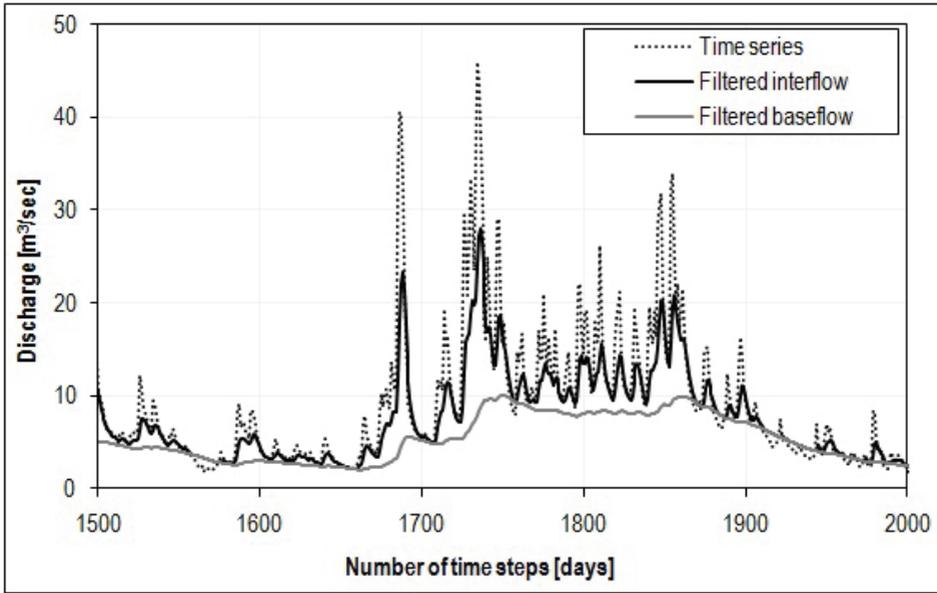


Figure-5: Interflow filter result

4.3 POT EXTRACTION

The parameters k_p , f and q_{lim} for extracting POTs for quick flow period are chosen as 8 days, 0.6 and 3 m^3/s . While the same for slow flow period are chosen as 80 days, 1 and 0.5 m^3/s . Since selection of POT is not really based upon any strict speculation, this

is also a trial and error process and the parameters can be well adjusted by the user depending on the POT plots. But the methods and the parameter's sensitivity should be well understood. *Figure-6* shows the results for the initial 1000 to 2500 days. Altogether 118 POTs for quick flow and 34 POTs for slow flows are extracted.

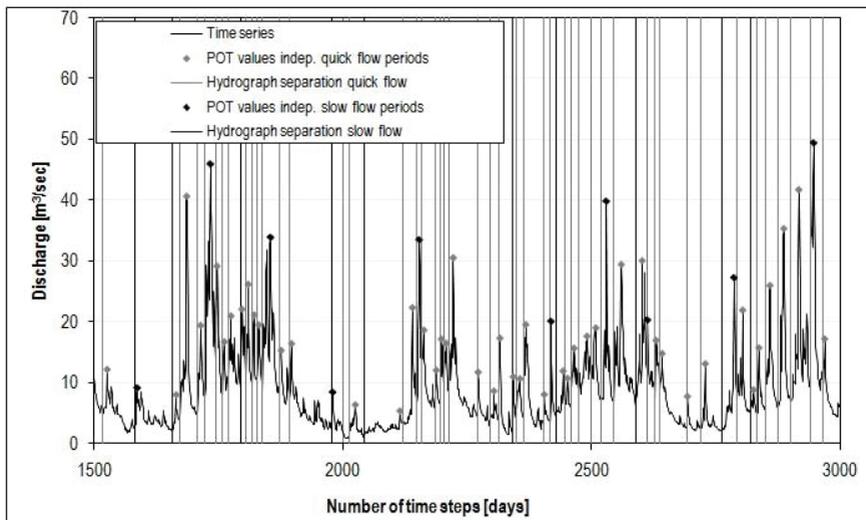


Figure-6: POT selection for quick and slow flow periods

4.4 PERFORMANCE EVALUATION PLOTS

Table-1 shows the results of performance evaluation in terms of MSE and NSE. Looking only on these goodness-of-fit statistics it is clear that the model is quite efficient on modelling slow flows than quick flows as indicated by higher NSE and lower MSE values for slow flow periods.

Table 1: Selected numbers of quick and slow flow periods with errors

Flow Periods	Number of POTs	MSE	NSE
		[M ³ /S]	[%]
Quick Flow	118	1.51	66
Slow Flow	34	1.12	79

Figure-7 and Figure-8 are the comparison plot of peak flows (maxima) and low flows (minima) periods. Plot for maxima (Figure-7) showed relatively higher scatter but almost zero bias (indicated by the overlapping of bisector and mean deviation line). In an average no systematic overestimation or underestimation is observed. Plot of minima (Figure-8) shows lower scatter but a clear negative bias (low flows or base flow components are systematically underestimated).

The mean ± standard deviation lines on both plots indicate the scattering of observed and simulated values for 68% confident interval.

Figure-9 indicates the performance of the model in extreme high events with different return periods. The plot reveals that extreme simulated discharge matches very closely to the observed and no systematic over and underestimation is observed for those extremes. For lower return periods, the model performance is quite good with slight overestimation on lower return period that indicates that the model can be used for high flow problems and applications.

Figure-10 shows the performance of the model in low flow events on different return periods. Due to small value of low flow, it is quite difficult to analysis by putting normal values of low discharge in y-axis in a figure. Therefore, inverse value of low discharge used to see lower value of discharge in this case. For return periods of 0.5 to 5 years, model tends to underestimate very low discharges. But lower discrepancies are observed towards lower and upper tail of the empirical extreme value distribution. So care should be taken while using the model for those events.

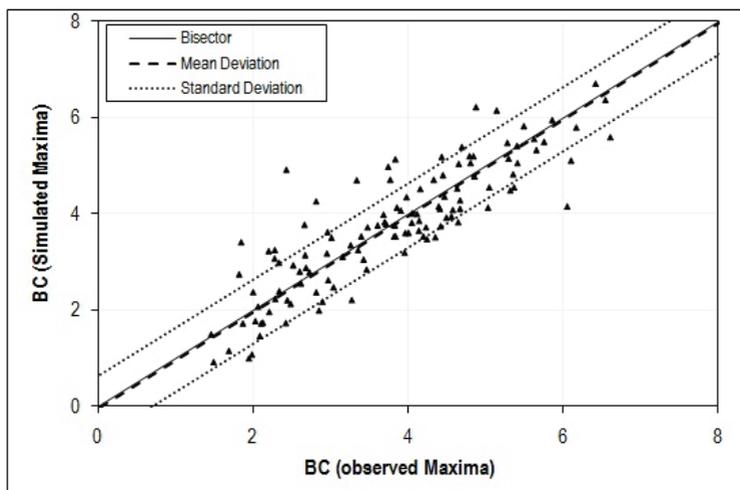


Figure-7: Comparison plot of Maxima after Box-Cox Transformation ($\lambda=0.25$)

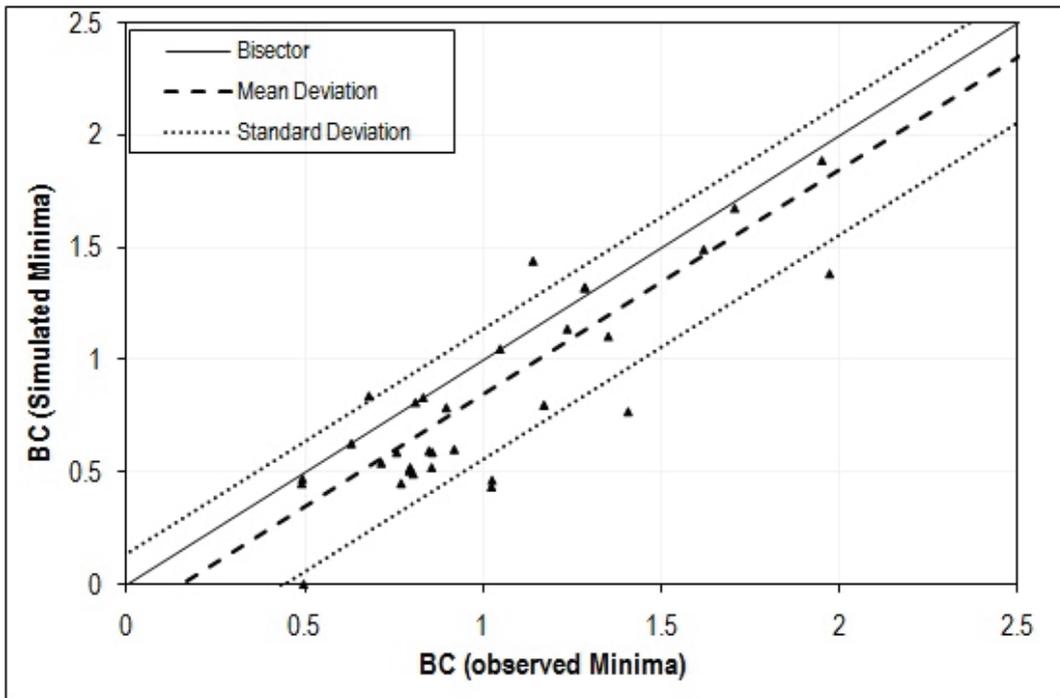


Figure -8: Comparison on plot of Minima after Box-Cox Transformation ($\lambda=0.25$)

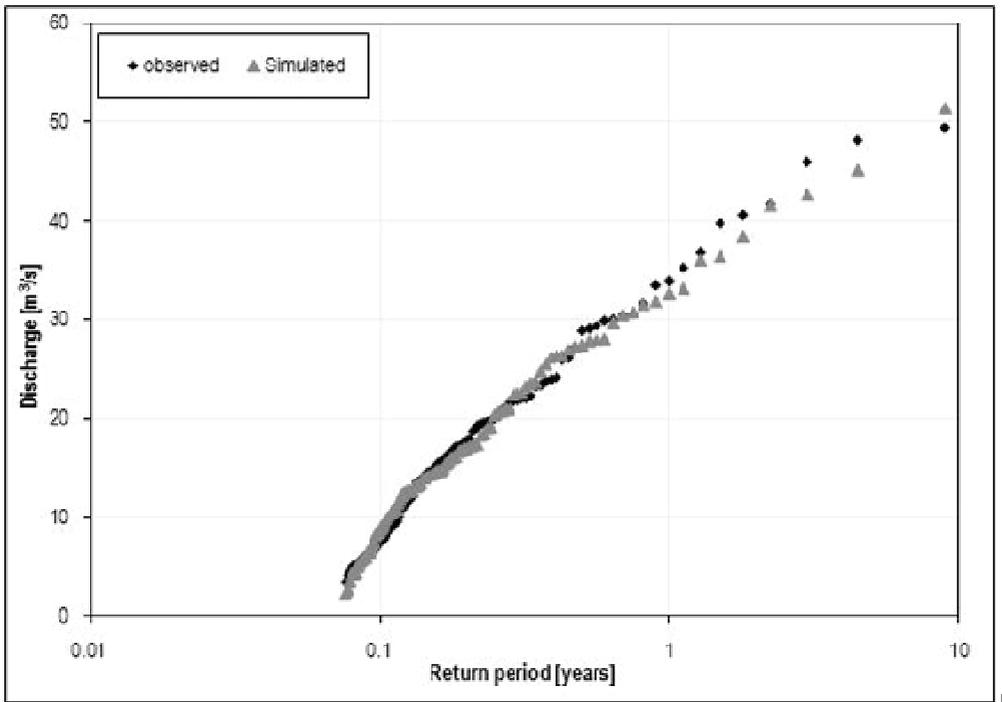


Figure 9: Comparison plot of high flow extreme value distribution

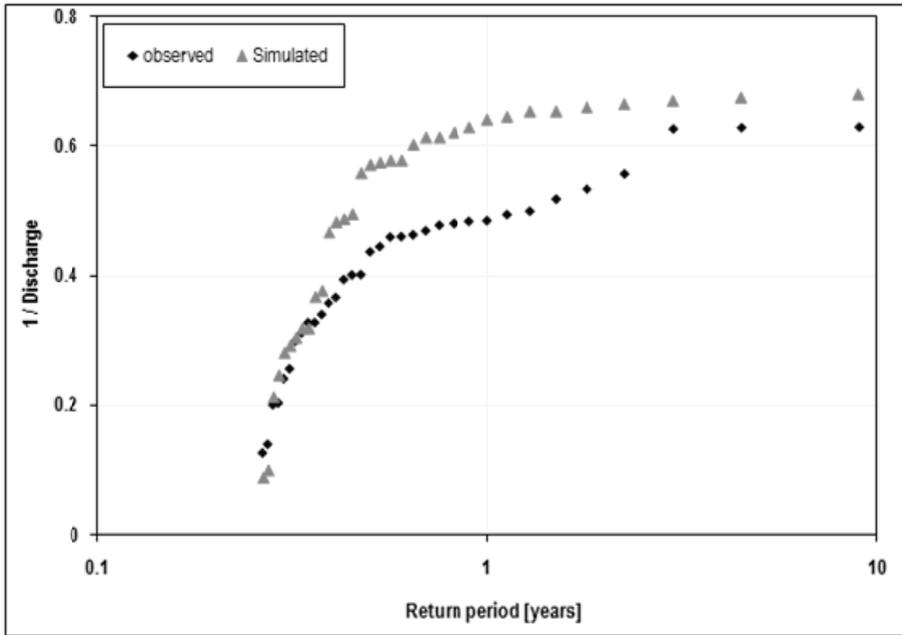


Figure-10: Comparison plot of low flow empirical extreme value

Figure-11 shows the comparison plot of cumulative discharges and it is clear that the model tend to overestimate the total flows. This is indication of high water yield through the simulation. This may be due to the effect of higher value of soil evaporation compensation factor (ESCO)

which allows less evapo-transpiration allowing substantial portion of water volume coming to the river stream. Another reason may be due to low threshold water depth in the shallow aquifer (GWQMN) releasing slightly greater amount of water from aquifer than it should.

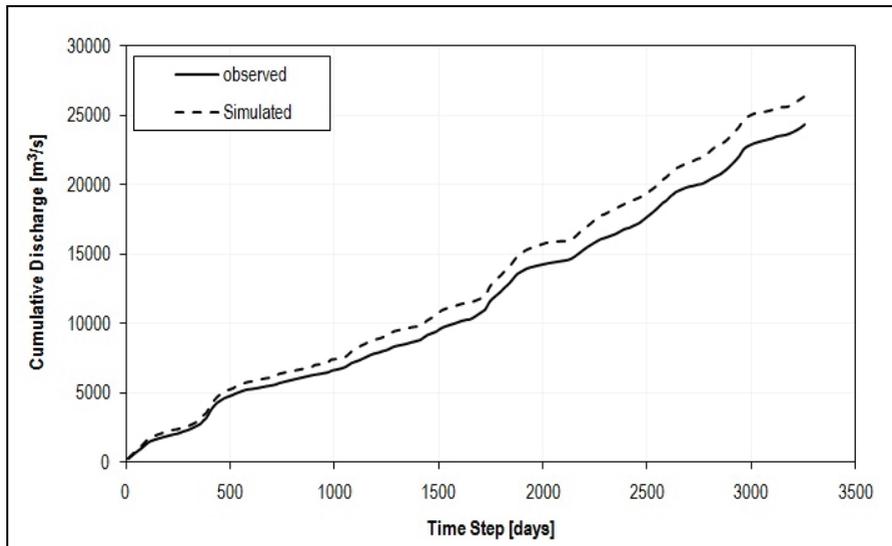


Figure 11: Comparison plot of cumulative flow volumes (total flows)

5. CONCLUSION

The paper demonstrates a methodology of model performance evaluation using WETSPRO. As it is clear that modeling is a necessary part for the decision makers so a robust model calibration and their performance evaluation should be thoroughly checked before the model are being used in real time. The trivial way of accessing the robustness of model output by using some goodness-of-fit statistical indicators is not always sufficient. Therefore, consideration on problems of serial dependency and heteroscedasticity should be addressed. Model performance on different flow components (peak and low flows), separate evaluation of different subflows, evaluation of peak and low flow extremes and evaluation on total flow volume clearly displays model's performance.

The performance evaluation made on this study is based on the calibration of SWAT model of Kliene Nete Catchment of Belgium by Shrestha et al., (2010). It is clear that model's performance on high flows is quite satisfactory with slight overestimation on lower return periods. Performance on low flows showed slight underestimation as can be seen from *Figure-8 and 10*. This indicates that the model can be used for high flow problems but care should be taken while using it in low flow problems. In terms of total volume, the model has tendency to overestimate the water balance. If one uses this model for designing a reservoir for a larger time span, then it might be that he/she can come up with slightly oversized reservoir design. This can be addressed by manipulating SWAT model parameters for example; decreasing ESCO (soil evaporation compensation factor) to allow a larger portion of water to evaporate or a higher depth for GWQMN (threshold water depth in the shallow aquifer for flow) can be adopted so as to allow less water to release from the aquifer.

6. ACKNOWLEDGEMENT

Authors wish to express their gratitude to Prof. P. Willems and Prof. W. Bauwens for their immense support during study period. Special thanks go to the Flemish Interuniversity Council (VLIR) for providing the scholarship for this International Master Program in Water Resources Engineering (2007-2009) and Katholieke Universiteit Leuven and Vrije Universiteit Brussel, Belgium for providing the platform. Lastly, appreciate the help of PhD researcher Mr. Nirman Shrestha from Katholieke Universiteit Leuven, Belgium for language correction of this paper.

REFERENCES

- Arnold, J.G., Srinivasan, P., Muttiah, R.S., Williams, J.R. (1998): Large area hydrologic modeling and assessment. Part I. Model development. *Journal of American Water Resources Association*. 34, pp. 73–89.
- Box, G.E.P., Cox, D.R., (1964): An analysis of transformations. *Journal of the Royal Statistical Society* 26, pp. 211–243.
- Chapman, T. (1991): Comment on Evaluation of automated techniques for base flow and recession analyses by Nathan R.J. and McMahon, T.A. *Water Resources Research* 27 (7), pp. 1783–1784.
- Dam J., Salvadore E., Van Daele T. & Batelaan, O., (2009): Case Kleine Nete: hydrologie, Wetenschappelijk rapport, NARA-2009. [INBO.R.2009.28]. Rapporten van het Instituut voor Natuur- en Bosonderzoek 2009 (28). Instituut voor Natuur- en Bosonderzoek, Brussel.

- Dmitri, K., George, K. and Stewart, W. F., (2006). Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology* Vol. 320, pp. 173-186.
- Green, C.H., Griensven, A.V. (2008): Autocalibration in hydrologic modeling: Using SWAT2005 in small-scale watersheds. *Environmental Modelling & Software*. Elsevier Science, Vol. 23, Issue 4, pp. 422 – 434.
- Nash, J. E., and Sutcliffe, J. V., (1970). River flow forecasting through conceptual models part I-A discussion of principles. *Journal of Hydrology*, 10 (3), pp. 282–290.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., and King, K.W. (2002): Soil and water assessment tool theoretical documentation. Version 2000, Texas water resources institute, Colleague station, Texas, USA.
- Shrestha, N.K., P.C., Shakti and Gurung, P., (2010). Calibration and validation of SWAT Model for low lying watershed: a case study on Kliene Nete Watershed, Belgium. *Hydro Nepal: Journal of Water, Energy and Environment* Vol. 6, 2010, pp.47-51.
- Srinivasan, R., Ramanarayanan, T.S., Arnold, J.G. and Bednarz, S.T. (1998): Large area hydrological modeling and assessment. Part II: Model application. *Journal of American Water Resources Association*, 34(1), pp. 91-101
- Willems, P., (2009). A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling and Software* 24(3), pp. 311-321.
- Willems, P., (2010). VHM approach: transparent, step-wise and data mining based identification and calibration of parsimonious lumped conceptual rainfall-runoff models. *Journal of Hydrology* (submitted).