

Performance Evaluation of Single and Multi-Objective Calibration in Rainfall-Runoff Modelling

T.N. Wickramaarachchi and R.M.K. Rathnayake

Abstract: Hydrologic modelling is widely used as a tool for making decisions and predictions in planning and managing water resources. Successful application of a hydrologic model depends upon careful calibration and validation using appropriate performance measures in satisfying corresponding performance evaluation criteria. Performance of MIKE 11 NAM rainfall runoff model was assessed using four calibration objectives and their different combinations, as applied to upper Gin catchment, Sri Lanka. The four calibration objectives measured different aspects of hydrograph: good water balance, good overall agreement of the shape of the hydrograph, good agreement for peak flows, and good agreement for low flows. Their numerical performance was measured using four objective functions from which fifteen calibration schemes were formed (four single objective schemes and eleven multi-objective schemes). Using aggregated distance measure, equal weights were assigned to the four objective functions. Shuffled complex evolution algorithm was used to solve the multiple calibration objective problems. Model performance was evaluated using three criteria: Nash-Sutcliffe coefficient (NSE), percent bias (PBIAS) and coefficient of determination (R^2). Results revealed significant trade-offs between the objective functions, highlighting that no single calibration objective was able to depict all the aspects of the hydrograph simultaneously. However, multi-objective calibration yielded more accurate and consistent simulations covering different aspects of the hydrograph, simultaneously with overall best performance shown for combination of the four objective functions satisfying all the performance evaluation criteria (NSE=0.56, $R^2=0.56$, PBIAS=17%), compared to the single-objective calibration. Instead of using R^2 alone, use of the corresponding regression slope as a weighing factor of R^2 was recommended following further analysis of simulation results.

Keywords: Calibration objectives, MIKE 11 NAM, Objective functions, Performance criteria, Rainfall-runoff

1. Introduction

Hydrologic models have been widely used for decades as a tool applied in hydrological assessments, including impact assessments on water resources due to human activities and climate change [1], [2], [3], [4], [5], [6]. In distributed hydrologic modelling, the spatial variability of catchment characteristics is taken into account; however, distributed hydrologic models require huge amount of data and identification of considerable number of model parameters which possibly limit their applicability for operational purposes. Besides, there is no clear evidence to prove the distributed models' contribution towards additional model efficiency [7]. Thus, lumped rainfall-runoff models are widely applied as these models use well established physical laws at micro scale, include simple conceptual parameterization and efficient computation, and represent an appropriate alternative to understand hydrological system dynamics in areas with limited data [8], [9].

MIKE 11 NAM, a lumped, conceptual rainfall-runoff model, simulates the rainfall-runoff processes occurring at the catchment scale. It had been applied to a number of catchments around the world, representing many different hydrological regimes and accepted as a well-proven engineering tool that can be used to simulate the watershed physical processes [10], [11], [12], [13], [14], [15]. MIKE 11 NAM model was setup in this study to simulate the rainfall-runoff process.

In hydrologic modelling studies, proper model calibration is important as model performance

*Eng. (Dr.) (Mrs.) T.N. Wickramaarachchi, C.Eng., MIE(SL), B.Sc. Eng. (Hons) (Moratuwa), M.Phil. (Moratuwa), PhD (Yamanashi), Senior Lecturer, Department of Civil and Environmental Engineering University of Ruhuna, Sri Lanka.
Email: thushara@cee.ruh.ac.lk
ORCID ID: <https://orcid.org/0000-0002-0274-2169>
Eng. R.M.K. Rathnayake, B.Sc. Eng. (Ruhuna).*



highly depends on the selection of objective functions [16], [17]. Particularly, in conceptual hydrological models, since model parameters are conceptual representations of catchment processes, they cannot be determined through direct measurements and have to be estimated through matching observed and modelled catchment behaviour in the calibration process [18]. In most of the modelling approaches, objective functions are used to measure different aspects of the hydrograph which generally contain overall water balance, overall shape of the hydrograph, low flows and peak flows [19]. Single objective calibration is widely used in many practical applications, however, identification of an objective function that reproduces all the characteristics of a hydrograph may be daunting [20], [21]. To circumvent this problem, use of multi-objective functions for hydrologic model calibration has been assessed for different types of models and verified in providing better model performance: semi-distributed model [22], [23]; lumped conceptual model [18], [19]; and dynamic flood operation model [24]. This study incorporated the most widely used method for solving the multi-objective global optimization problem which transformed the multi-objective problem into a single objective problem by using aggregated distance measure, a method to assign equal weights to the different objectives in the optimization [19], [20], [25].

Performance evaluation criteria are defined as mathematical determination of how well a model simulation matches the available observations [26]. In general, many efficiency criteria contain a summation of the absolute or squared errors, to evade the cancelling of errors of opposite sign. Thus, the calibration attempt which aimed at minimizing these errors neglects smaller errors associated with low streamflow values while placing an emphasis on larger errors associated with high streamflow values leading to fit the higher portions of the hydrograph at the expense of the lower portions [27]. Accordingly, many studies demonstrated the advantages of using multiple evaluation criteria over a single criterion to cover entire hydrograph since the use of a single evaluation technique led to undue emphasis on matching one aspect of the hydrograph at the expense of other aspects [19], [27], [28], [29], [30]. In complying with them, this study included three performance evaluation criteria, namely Nash-Sutcliffe coefficient (NSE), percent bias (PBIAS) and coefficient of determination (R^2).

Aforementioned research focused mainly on the performance evaluation of selected calibration objectives and therefore, no comprehensive evaluation is available to measure and compare the performance of single objective functions and their different combinations together with corresponding multiple performance evaluation criteria as applied for lumped conceptual rainfall-runoff modelling. Moreover, there exists a significant need to evaluate the performance of single calibration objectives and their different combinations, because such a complete multi-criteria optimization only allows an analysis of trade-offs among the different objectives which enable to better understand the limitations of the model structure, thereby gaining insights into possible model improvements. Present study was conducted to fill up the lacuna in this area of research by applying MIKE 11 NAM lumped, conceptual rainfall-runoff model to the Gin catchment which provides flood control and water supply to the capital city of Galle in southern Sri Lanka.

2. Materials and Methods

2.1 Study Area

Gin catchment covers an area of about 930 km² and drains to Indian Ocean by Gin River with a total length of about 116 km. The catchment is located in southern Sri Lanka between 80°08" E to 80°40" E and 6°03" N to 6°26" N. It is characterized by bimodal rainfall pattern predominantly between the months of May and September (south west monsoon), and between November and February (north east monsoon) followed by inter monsoon rains during the remaining months of the year with mean annual rainfall between 2,500 mm (at the upstream) and 3,500 mm (at the downstream). The upper reaches of the river flow through hilly region, and the lower reaches of the river flow through alluvial plain with relatively flat slopes, most of which is flood prone. Land use in the catchment primarily consists of forests, cultivations and human settlements with most of the settlements located within the extensive floodplain of Gin River. Red Yellow Podzolic soils are predominant in the upland areas while the lowland floodplains are dominated by alluvial and bog soils [31].

The catchment receives heavy rainfall during the south west monsoon period resulting in frequent flooding which greatly threatens the safety of life and property in the catchment.

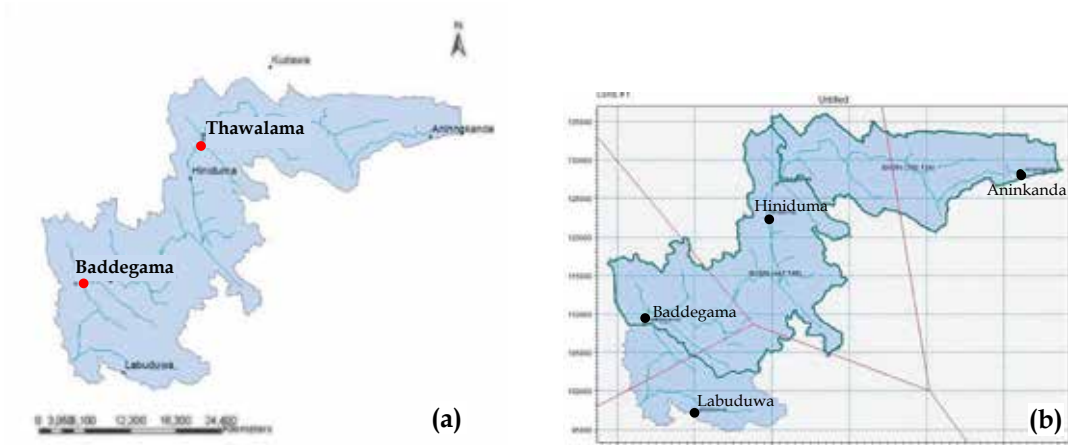


Figure 1 - (a) Gin Catchment (b) Rain Gauging Stations Weighted by Thiessen Polygon Method

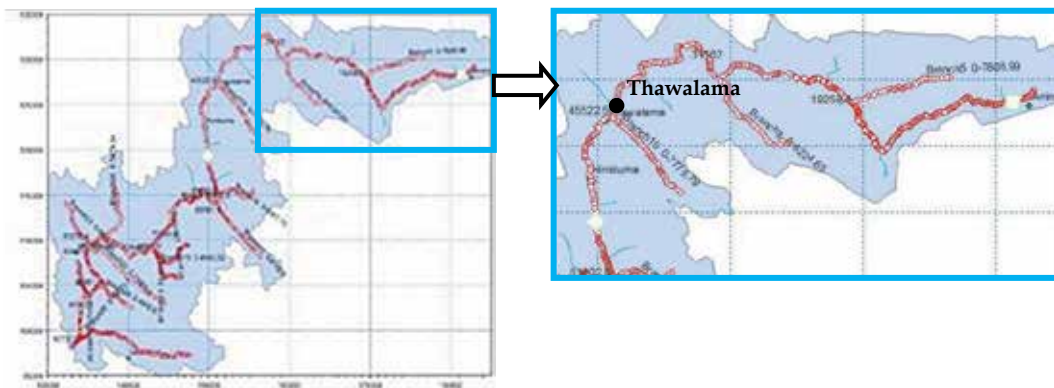


Figure 2 - River Network Generated in MIKE 11

May, 2017 flood event approximately conform to a flood with a return period of 200 years [32]. Gin River is the primary drinking water source from which pipe-borne water is supplied to Galle city, the capital of southern Sri Lanka. Thus, accurate rainfall-runoff simulations are vital to understand the catchment hydrological processes in minimizing the flood hazards and managing the water resources efficiently.

2.2 Data Sets and River Network Generation

Daily rainfall data at Aninkanda, Baddegama, Hiniduma, and Labuduwa rain gauging stations (from 2000 to 2016) and other meteorological data were obtained from the Department of Meteorology, Sri Lanka. Daily river discharge data at Thawalama river gauging station (from 2000 to 2016) and river cross-sections were obtained from the Department of Irrigation, Sri Lanka. The weighted areal rainfall was computed using Thiessen's polygon method. Gin catchment, and its rain gauging stations weighted by Thiessen polygon method, are shown in Figure 1 (a) and Figure 1 (b), respectively.

Delineation of the Gin catchment was carried out using Shuttle Radar Topography Mission (SRTM) data of 1 arc second resolution [33]. Based on the locations of the two river gauging stations at Thawalama ($6^{\circ}20'33''$ N, $80^{\circ}19'50''$ E) and Baddegama ($6^{\circ}11'23''$ N, $80^{\circ}11'53''$ E), the catchment was divided into sub-catchments. The upper most sub-catchment, upstream area of Thawalama, which covers 490 km^2 was considered for the study.

River network developed in MIKE 11 with the locations of cross sections is shown in Figure 2. The uppermost sub catchment was linked at Thawalama defining the location of the sub-catchment outlet.

2.3 MIKE 11 NAM Model Set-up

MIKE 11 NAM hydrological model was applied in this study to simulate the rainfall-runoff processes occurring at the catchment scale. MIKE 11 NAM, which is a lumped, conceptual rainfall-runoff model, considers the catchment as a single entity. The model consists of four different interrelated storages: groundwater storage, root zone storage, surface storage, and snow storage. Overland



flow, inter flow and base flow components were simulated as a function of the moisture contents in the four storages. Parameters related to surface, root zone and groundwater storages were calibrated in this study. The parameter values of the model corresponding to catchment characteristics were estimated using different mathematical algorithms. Schematic representation of the MIKE 11 NAM model is shown in Figure 3. Description of the model parameters and the range of each parameter used for the model calibration are shown in Table 1.

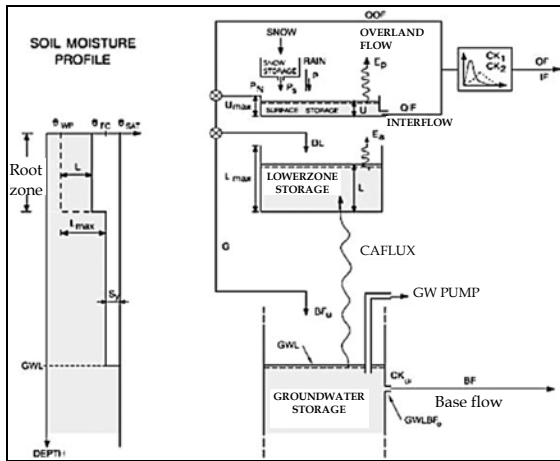


Figure 3 - Schematic Representation of the MIKE 11 NAM model [37]

2.4 Calibration Objectives and Objective Functions

Calibration objectives have to be formulated as numerical goodness-of-fit measures that are optimized in the model calibration. In the model calibration, the following four calibration objectives were considered: (i) a good agreement between the average simulated and observed catchment runoff (i.e. a good water balance); (ii) a good overall agreement of the shape of the hydrograph; (iii) a good agreement of the peak flows with respect to timing, rate and volume; and (iv) a good agreement for low flows.

The parameters which represent the catchment characteristics of lumped, conceptual hydrological models are not measured manually in the calibration process. They are estimated using mathematical equations (objective functions) such that the model simulated values match with observed values of the catchment [19], [34].

Table 1 - Description of Model Parameters and the Parameter Range used for Model Calibration

Parameter	Description (Unit)	Range
U_{max}	Maximum water content in surface storage (mm)	10 - 19.998
L_{max}	Maximum water content in lower zone/root storage (mm)	100 - 299.267
C_{QOF}	Overland flow coefficient	0.115 - 1
C_{KIF}	Interflow drainage constant (hrs)	200.06 - 1000
C_{KI}	Timing constant for overland flow (hrs)	10 - 49.998
T_{OF}	Overland flow threshold	0 - 0.989
T_{IF}	Interflow threshold	0 - 0.99
TG	Groundwater recharge threshold	0 - 0.988
C_{KBF}	Timing constant for base flow (hrs)	1000.507 - 3985.771

For the four calibration objectives stated above, there are four different objective functions defined in MIKE 11 NAM model to measure the numerical performances [35]:

- (i) agreement between the average simulated and observed catchment runoff: overall volume error, $F_1(\phi)$ (Equation 1);
- (ii) overall agreement of the shape of the hydrograph: overall Root Mean Square Error (RMSE), $F_2(\phi)$ (Equation 2);
- (iii) agreement of the peak flows: average RMSE of peak flow events, $F_3(\phi)$ (Equation 3); and
- (iv) agreement of the low flows: average RMSE of low flow events, $F_4(\phi)$ (Equation 4).

$$F_1(\phi) = \left| \frac{1}{N} \sum_{i=1}^N [Q_{obs,i} - Q_{sim,i}(\phi)] \right| \quad \dots(1)$$

$$F_2(\phi) = \left[\frac{1}{N} \sum_{i=1}^N [Q_{obs,i} - Q_{sim,i}(\phi)]^2 \right]^{\frac{1}{2}} \quad \dots(2)$$

$$F_3(\phi) = \frac{1}{M_p} \sum_{j=1}^{M_p} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{obs,i} - Q_{sim,i}(\phi)]^2 \right]^{\frac{1}{2}} \dots(3)$$

$$F_4(\phi) = \frac{1}{M_l} \sum_{j=1}^{M_l} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{obs,i} - Q_{sim,i}(\phi)]^2 \right]^{\frac{1}{2}} \dots(4)$$

where $Q_{obs,i}$ is the observed discharge at time i , $Q_{sim,i}$ is the simulated discharge at time i , ϕ is the set of model parameters to be calibrated, N

is the number of time steps in the calibration period, M_p is the number of peak flow events in Equation (3), M_l is the number of low flow events in Equation (4) in the calibration period and n_j is the number of time steps in event number j .

Both peak and low flow events were specified separately with respect to threshold values.

2.4.1 Flow-Duration Analysis and Threshold Values for Low Flows and High Flows

Flow-duration analysis was performed by using daily average flow during 2000-2016 to identify the high flow segment and the low flow segment. Weibull plotting position was used to develop the flow-duration curves as it provided an unbiased estimate of exceedance probability, regardless of the underlying probability distribution of the ranked observations [36]. Low flow was defined as the values falling above the 75th percentile in the distribution of streamflow values. High flows were the streamflow values below the 25th percentile. For 75% of the time, the streamflow

equalled or exceeded 13 m³/s and for 25% of the time, it equalled or exceeded 38 m³/s.

2.5 Model Calibration

Since most of the MIKE 11 NAM parameters are of an empirical and conceptual nature, generally it is not possible to determine their values on the basis of the physiographic, climatic and soil physical characteristics of the catchment [37]. Hence the final parameter estimation was performed by calibration against time series of hydrological observations which automatically generated the model parameters and assessed the performance of the model according to the model performance measures using the Shuffled Complex Evolution (SCE) algorithm [38], [39]. SCE algorithm which combined the global search method (genetic algorithm) with the local search method (simplex method) had been widely applied all over the world and proven to be consistent, effective, and efficient global optimization procedure in locating the optimal model parameters of a hydrological model [29], [40].

Table 2 - Calibration Schemes and Objective Functions

Calibration Scheme	Objective Functions	Description
1	$F_1(\emptyset)$	Overall volume error
2	$F_2(\emptyset)$	Overall RMSE
3	$F_3(\emptyset)$	Average RMSE of peak flow events
4	$F_4(\emptyset)$	Average RMSE of low flow events
5	$F_1(\emptyset)F_2(\emptyset)$	Combined overall volume error and overall RMSE
6	$F_1(\emptyset)F_3(\emptyset)$	Combined overall volume error and average RMSE of peak flow events
7	$F_1(\emptyset)F_4(\emptyset)$	Combined overall volume error and average RMSE of low flow events
8	$F_2(\emptyset)F_3(\emptyset)$	Combined overall RMSE and average RMSE of peak flow events
9	$F_2(\emptyset)F_4(\emptyset)$	Combined overall RMSE and average RMSE of low flow events
10	$F_3(\emptyset)F_4(\emptyset)$	Combined average RMSE of peak flow events and average RMSE of low flow events
11	$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)$	Combined overall volume error, overall RMSE and average RMSE of peak flow events
12	$F_1(\emptyset)F_2(\emptyset)F_4(\emptyset)$	Combined overall volume error, overall RMSE and average RMSE of low flow events
13	$F_1(\emptyset)F_3(\emptyset)F_4(\emptyset)$	Combined overall volume error, average RMSE of peak flow events and average RMSE of low flow events
14	$F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$	Combined overall RMSE, average RMSE of peak flow events and average RMSE of low flow events
15	$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$	Combined overall volume error, overall RMSE, average RMSE of peak flow events and average RMSE of low flow events



2.5.1 Optimization Algorithm

In the multi-objective calibration process, the optimization problem was stated as follows [37]:

$$\min \{F_1(\emptyset), F_2(\emptyset), \dots, F_n(\emptyset)\}, \emptyset \in \Theta \quad \dots(5)$$

where $F_i(\emptyset)$ ($i = 1, 2, 3, \dots, n$) are the different objective functions. In order to constrain the optimization problem, \emptyset was restricted to the feasible parameter space Θ . The parameter space was defined as a hypercube by specifying lower and upper limits on each parameter. The lower and upper limits were chosen according to mathematical and physical constraints in the model and from modelling experiences.

Due to various trade-offs between the different objectives, the solution of Equation (5) was not presenting a single unique set of parameters but consisted of the so-called Pareto set of solutions (non-dominated solutions). In order to conveniently solve the multi-objective calibration problem to get single point of the Pareto front, a single objective optimization problem was defined that aggregated the different objective functions $F_1(\emptyset) - F_4(\emptyset)$. The applied aggregate measure which is called the Euclidian distance is stated below:

$$F_{agg}(\emptyset) = [(F_1(\emptyset) + A_1)^2 + (F_2(\emptyset) + A_2)^2 + (F_3(\emptyset) + A_3)^2 + (F_4(\emptyset) + A_4)^2]^{1/2} \quad \dots(6)$$

where A_i are transformation constants. A balanced aggregated measure was incorporated by assigning the transformation constants in order to make the different objectives having equal weight in the optimization. The transformation constants were automatically calculated based on the initially generated population of parameter sets in the optimization loop [19], [37]. By minimising Equation (6) with respect to \emptyset , the optimal parameter set was found and by using the SCE algorithm [38], [39], the optimization was performed automatically.

2.5.2 Calibration Schemes

The rainfall-runoff model was established using single and combined objective functions in the calibration process. Fifteen different calibration schemes were used including the four single objective functions (Table 2). Data from year 2000 to 2015 was used for the model calibration. Calibration was carried out for each scheme using the sixteen single year's data for the purpose of detailed comparison of performance evaluation variations in the

calibration schemes. To ensure optimal performance for both the calibration period and the validation period, the validation period was also kept similar to the calibration period [41], [42]. Thus, the model was validated for the year 2016. Following the calibration and validation, the performance of the model was evaluated using the three performance criteria.

2.6 Performance Evaluation Criteria

It was shown in the previous studies that the model performance strongly depended on the selection of performance criteria [27], [43]. Both graphical and numerical performance measures should be applied in the performance evaluation; while the graphical evaluation includes comparison of the simulated and observed hydrograph, the numerical performance measure includes the overall water balance error (i.e. the difference between average simulated and observed runoff), and measure of the overall shape of the hydrograph based on coefficient of determination or Nash-Sutcliffe coefficient [37]. According to Moriasi et al. [41], three evaluation indexes which had been commonly used, accepted and recommended in published literature were selected to evaluate the model performance: percent bias, PBIAS [44] (Equation 7), Nash-Sutcliffe coefficient, NSE [45] (Equation 8), and coefficient of determination, R^2 [27] (Equation 9).

$$PBIAS = \left| 1 - \left(\frac{\sum_{i=1}^N Q_{s,i}}{\sum_{i=1}^N Q_{o,i}} \right) \right| \times 100 \quad \dots(7)$$

$$NSE = 1 - \left(\frac{\sum_{i=1}^N (Q_{o,i} - Q_{s,i})^2}{\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)^2} \right) \quad \dots(8)$$

$$R^2 = \left(\frac{\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)(Q_{s,i} - \bar{Q}_s)}{\sqrt{\sum_{i=1}^N (Q_{o,i} - \bar{Q}_o)^2} \sqrt{\sum_{i=1}^N (Q_{s,i} - \bar{Q}_s)^2}} \right)^2 \quad \dots(9)$$

where $Q_{o,i}$ is the observed discharge at time i , $Q_{s,i}$ is the simulated discharge at time i , \bar{Q}_o is the mean observed discharge, \bar{Q}_s is the mean simulated discharge and N is the total number of time steps.

3. Results and Discussion

Averages of absolute values of the three performance evaluation criteria obtained for the model calibration are shown in Table 3. In assessing the goodness-of-fit of the observed and the simulated hydrological processes,

Table 3 - Performance Evaluation of the Objective Function Schemes in the Calibration

Scheme	Objective Function	Performance Evaluation Criteria		
		PBIAS (%)	NSE	R ²
1	$F_1(\emptyset)$	0.13	0.32	0.46
2	$F_2(\emptyset)$	3.10	0.57	0.58
3	$F_3(\emptyset)$	6.96	0.50	0.55
4	$F_4(\emptyset)$	21.53	0.24	0.34
5	$F_1(\emptyset)F_2(\emptyset)$	0.12	0.57	0.56
6	$F_1(\emptyset)F_3(\emptyset)$	0.31	0.52	0.52
7	$F_1(\emptyset)F_4(\emptyset)$	0.30	0.44	0.51
8	$F_2(\emptyset)F_3(\emptyset)$	6.17	0.56	0.57
9	$F_2(\emptyset)F_4(\emptyset)$	13.54	0.55	0.57
10	$F_3(\emptyset)F_4(\emptyset)$	4.59	0.55	0.56
11	$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)$	0.40	0.55	0.56
12	$F_1(\emptyset)F_2(\emptyset)F_4(\emptyset)$	0.22	0.56	0.57
13	$F_1(\emptyset)F_3(\emptyset)F_4(\emptyset)$	0.36	0.53	0.55
14	$F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$	4.68	0.56	0.57
15	$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$	0.17	0.56	0.56

PBIAS value should be close to zero. For NSE and R², if the value equals to one, it denotes a perfect fit of the model simulations with the observations whereas a value less than zero is an indication of poor model performance [2].

3.1 Single Objective Function Calibration (Schemes 1- 4)

According to Table 3 and Figure 4, significant differences existed in the performance evaluation of the four single objective function schemes (schemes 1-4). The best value for PBIAS was given by $F_1(\emptyset)$ compared to the other single objective function schemes but it showed the highest variation (Figure 4). $F_1(\emptyset)$ underestimated the low flows and overestimated the high flows. Among all the single objective function schemes, $F_2(\emptyset)$ performed best for both the NSE and R² values, followed by $F_3(\emptyset)$. However, objective functions $F_2(\emptyset)$ and $F_3(\emptyset)$ were having the worst values and the highest variations for PBIAS compared to the $F_1(\emptyset)$. $F_4(\emptyset)$ which placed emphasis on simulation of the low flow did not perform well in matching the observed and simulated overall water volumes and overall shape of the hydrographs since NSE and R² were more sensitive to high flows which agreed with Krause et al. [27] and Zhao et al. [42]. To illustrate the differences of the hydrograph shapes calibrated using the four single objective function schemes, representative hydrographs for the observed and simulated discharge for year 2012 are

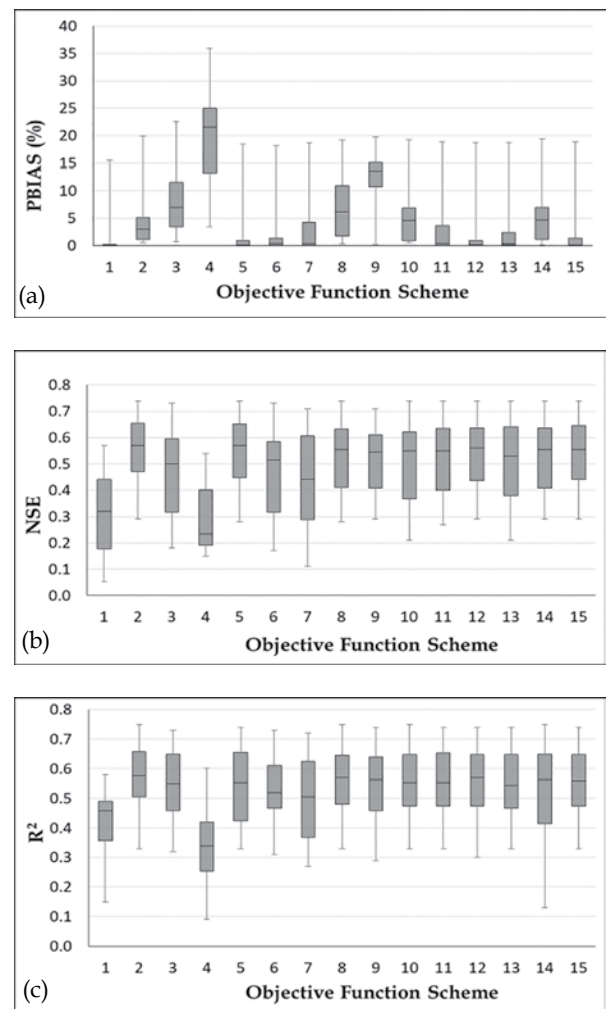


Figure 4 - Box Plots for the Performance Evaluation of Different Objective Function Schemes: (a) PBIAS; (b) NSE; and (c) R²

shown in Figure 5. Even though $F_1(\emptyset)$ showed a PBIAS value close to zero, according to



Figure 5, high and low flows were not simulated well in the calibration using $F_1(\emptyset)$; the model over predicted as much as it under predicted giving deceiving rating of model performance which agreed with Moriasi et al. [46]. Moreover, it is clearly shown in Figure 5 that the shape of the observed and simulated hydrographs were having reasonable agreement for $F_2(\emptyset)$ which measured the overall agreement of the shape of the hydrograph. However, owing to the comparatively high PBIAS, $F_2(\emptyset)$ did not show good agreement between the average simulated and observed catchment runoff. Therefore, it is evident significant trade-offs between the single objective functions, except for less significant trade-off shown between $F_2(\emptyset)$ and $F_3(\emptyset)$.

Thus, when single objective functions were used for the calibration, catchment characteristics were not represented well due to the trade-off between bias and variance which prevented the model to reach the optimal balance between two. This is where the multi-objective calibration approach is being used as shown by Gupta et al. [22] and Moriasi et al. [46] to represent the catchment characteristics while balancing all the objectives, since the objectives are of equal importance.

3.2 Multi-objective Function Calibration (Schemes 5 - 15)

In the multi-objective function calibration process, the four different calibration objectives were optimized using aggregated measure in which different objective functions were assigned equal weights. Table 3 lists the mean values of the performance evaluations

for different multi-objective functions (Schemes 5 - 15) and Figure 3 illustrates their variations. According to Table 3, the simulations by the multi-objective functions integrated the characteristics of the single objective functions and primarily produced either improved or balanced results for the performance evaluation criteria. For example, PBIAS value calibrated by the multi-objective function $F_1(\emptyset) F_2(\emptyset)$ was better than that by $F_2(\emptyset)$ and $F_1(\emptyset)$, while the value of NSE calibrated by the multi-objective function $F_1(\emptyset)F_2(\emptyset)$ was better than that by $F_1(\emptyset)$ and similar to that by $F_2(\emptyset)$. The value of R^2 calibrated by the multi-objective function $F_1(\emptyset) F_2(\emptyset)$ was better than that by $F_1(\emptyset)$ and a little worse than that by $F_2(\emptyset)$. As depicted by the model output, this slightly poor performance was attributed the oversensitivity of R^2 to outliers than to observations near the mean which was illustrated by Legates and Davis [47] as an obvious limitation of correlation-based measures.

From the box plots shown in Figure 3 which present the variation of the model's performance for different objective functions, it can be observed that schemes 5 - 15 have approximately similar mean values for NSE and R^2 , except for the schemes 6, 7 and 13. NSE showed high variation for the schemes 6 and 7, whereas R^2 showed high variation for the scheme 14. PBIAS showed significant differences in both mean value and variation for schemes through 5 to 15. The best PBIAS mean value was obtained by the scheme 5 followed by the scheme 15. Considering overall performance of all the evaluation

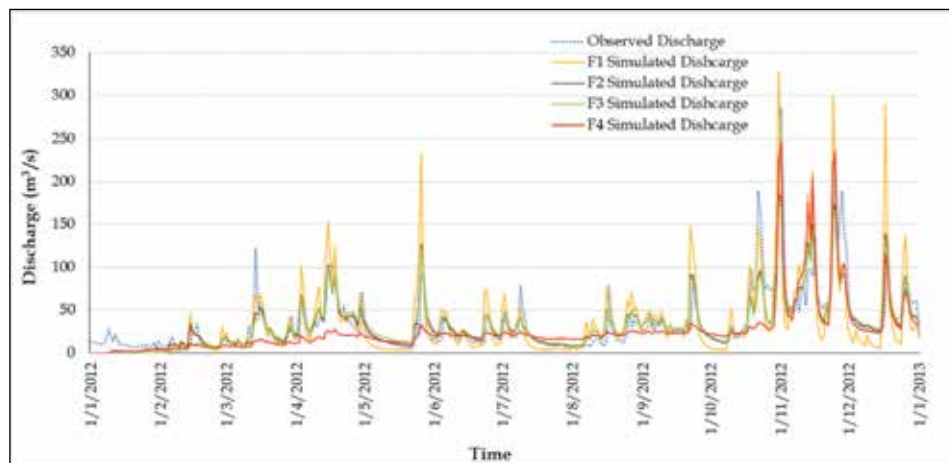


Figure 5 - Observed and Simulated Hydrographs for Year 2012 Calibrated by Single Objective Function Schemes (Schemes 1 - 4)

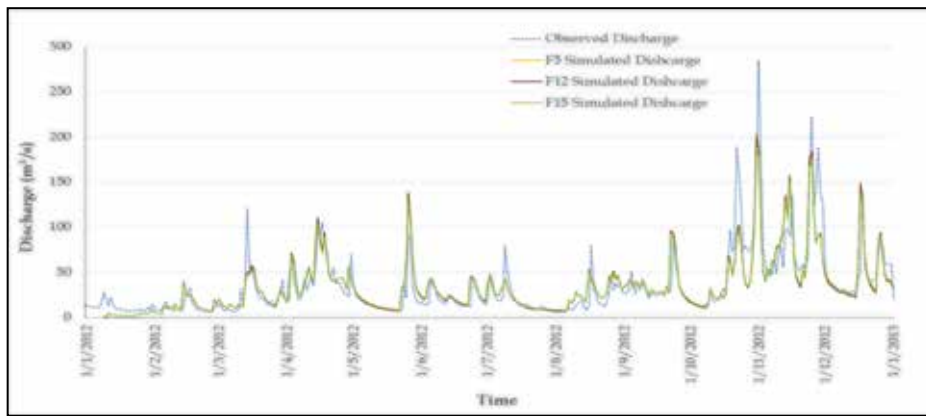


Figure 6 - Observed and Simulated Hydrographs for Year 2012 Calibrated by Multi-objective Function Schemes (Schemes 5, 12, and 15)

criteria of multi-objective calibration functions, schemes 5, 12 and 15 showed better values for NSE, R^2 and PBIAS in comparison to the other schemes. However, when schemes 5 and 12 were used for the calibration, a large variation was observed for NSE and R^2 .

Owing to the above aspects, scheme 15 [$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$] which includes combination of all the objective functions was recommended as the best. This further confirmed the findings by Jie et al. [20] and Moriasi et al. [46], who recommended synthesizing the performance of the multiple evaluation indexes in evaluating the model performance.

To verify the above, the simulated hydrographs which were calibrated by the multiple objective functions (scheme 5, 12 and 15) were compared with the observed hydrograph as shown in Figure 6 using representative hydrographs for year 2012. Simulated hydrographs showed better agreement with the observed (Figure 6), depicting more comprehensive performance in simulating the hydrological behaviour of the catchment (including high, intermediate and low flows) when calibrated using the multi-objective calibration functions, compared to the single objective calibration functions (Figure 5). Out of the three schemes, overall best agreement between the simulated and observed hydrographs was shown for the scheme 15.

Though the simulated hydrographs followed the trend of the observed flows, the highest flows ($> 90 \text{ m}^3/\text{s}$) were underestimated by the model. Underestimation of the highest flows might be due to the inadequate representation of spatial variability of rainfall owing to

limited number of rainfall gauging stations, particularly in the upper and middle reaches of the catchment [Figure 1(b)] [48], [49]. Moreover, according to Makungo et al. [13], MIKE 11 NAM model has generally been known to underestimate peak flows owing to lack of ability of rainfall-runoff models to simulate complex hydrological phenomena during high flow periods, accurately. Additionally, observational errors, technical errors associated with inaccurate calibration of recording devices and extrapolation of ratings beyond the range of gauged discharges, might have individually or collectively led to underestimation of the aforesaid flows.

3.3 Calibration vs. Validation Performance Criteria

Reliability and hydrological soundness of the calibrated MIKE 11 NAM model was evaluated by the validation process carried out using year 2016 data for the scheme 15 [$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$]. As recommended by Morasai et al. [46], the same performance evaluation criteria which were used for the model calibration, were applied for the model validation. Model validation performance data are stated in Table 4. PBIAS, NSE and R^2 values obtained in the multi-objective calibration and validation runs were acceptable since they were comparable or better than those obtained in the studies carried out by Morasai et al. [46] and Odiyo et al. [14] (“satisfactory” if $R^2 > 0.60$, $\text{NSE} > 0.50$, and $\text{PBIAS} \leq \pm 15\%$). This implied that, following the multi-objective calibration and validation processes, the MIKE 11 NAM model performed reasonably well in representing the hydrological characteristics of the catchment that influence the runoff generation process.



Table 4 - Performance Evaluation of the Objective Function Scheme 15 [$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$] in the Validation

Objective Function Scheme	Performance Evaluation Criteria		
	PBIAS (%)	NSE	R ²
15	8.36	0.68	0.69

3.3.1 Assessment of the systematic and dynamic behaviour of the model for the validated objective function scheme 15

For the validated scheme 15 [$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$], systematic and dynamic behaviour of the model was assessed. For the validated scheme 15, scattergram of observed and simulated discharge values with the 1:1 line indicating their equality of magnitudes are shown in Figure 7 while observed and simulated hydrographs are shown in Figure 8. Figure 7 suggests overall tendency of over prediction by the model, thus showing a positive bias. According to Figure 8, the dynamics of the observed hydrograph were simulated correctly while systematic over predictions and under predictions were apparent. This made the impression that the model sometimes overestimated or underestimated the flows, depending on the magnitude, giving a confused picture of the model performance.

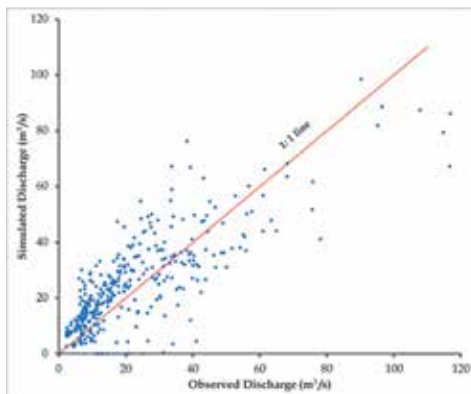


Figure 7 - Scattergram of Observed versus Simulated Values for the Validated Scheme 15

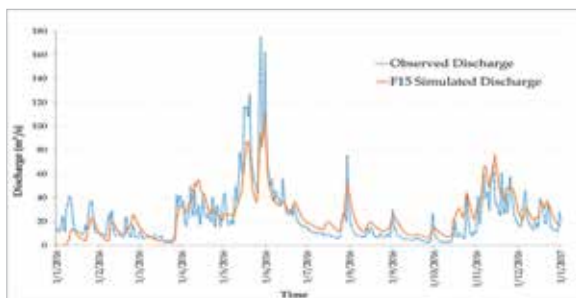


Figure 8 - Observed and Simulated Hydrographs for the Validated Scheme 15

To provide more insight into the model performance over different flow regimes, hydrologically oriented model diagnostics which are flow-duration curves were used (Figure 9) [46], [50]. The flow-duration curves showed that the model tended to underestimate the observed data for the highest flows (< 5% exceedance probability); had a good agreement for the high flows (> 5% and < 25% probability); overestimated during the intermediate and low flows (> 25% and < 90% probability); and had a good agreement for the lowest flows (> 90% exceedance probability) except for a few lowest flows exceeding 98% probability which was underestimated.

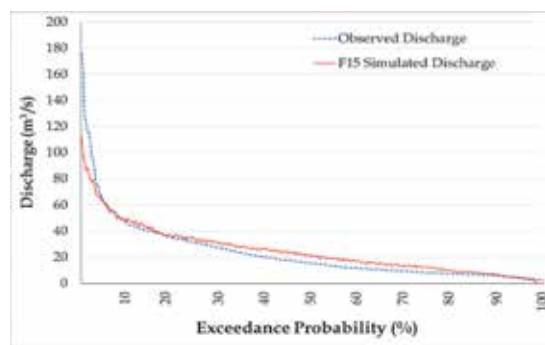


Figure 9 - Comparison of Observed and Simulated Flow-duration Curves for the Validated Scheme 15

For the validated scheme 15 [$F_1(\emptyset)F_2(\emptyset)F_3(\emptyset)F_4(\emptyset)$], the value of the NSE and R² were 0.68 and 0.69, respectively. Though NSE and R² were having high values, it was indicated that both NSE and R² were not very sensitive to the quantification of systematic under prediction or over prediction errors, which agreed with Krause et al. [27]. NSE and R² were based on squared differences between observation and prediction and, as a result, an emphasis was placed on larger errors (associated with high streamflow values) while smaller errors (associated with low streamflow values) tended to be neglected leading to fit the higher portions of the hydrograph (e.g., high flows) at the expense of the lower portions (e.g., low flows).

Previous studies revealed that, good R² values close to 1.0 could still be resulted from a model which systematically under predicts or over predicts all the time. R² standardizes the differences between the simulated and observed variances and it only evaluates linear relationships between the variables. Thus, R² is insensitive to additive and proportional

differences between the model simulations and observed data [27], [30]. To cope with that problem, Krause et al. [27] developed a weighted version of R^2 (ωR^2) by combining the R^2 and gradient of the regression on which R^2 was based. By weighting R^2 , model simulation results were reflected more comprehensively by quantifying over predictions or under predictions (Equation 10).

$$\omega R^2 = \begin{cases} |b|.R^2 & \text{for } b \leq 1 \\ |b|^{-1}.R^2 & \text{for } b > 1 \end{cases} \quad \dots(10)$$

where, ωR^2 is the weighted version of R^2 ; b is gradient of the regression on which R^2 is based.

According to Equation (10), the weighted R^2 value of 0.50 for the validated scheme 15 reflected the aforementioned poor simulation better than that of the R^2 value alone (0.69). Thus, it is evident that more reliable inferences could be obtained by weighting the coefficient of determination by the slope of the regression, which forms modified form of R^2 . This is stressing to further investigate for more conservative performance evaluation criteria for NSE as well, such as its modified form, by incorporating absolute values rather than squared differences that evaluate model performance without being influenced by extreme values.

4. Conclusion

Performance of fifteen calibration schemes which formed using four objective functions and their combinations were investigated using the most commonly used performance evaluation criteria for rainfall-runoff modelling using MIKE 11 NAM. SCE algorithm was used to calibrate the model while aggregated distance measure was used to set equal weights to different objective functions in obtaining a compromise solution. Study revealed that no single objective function was adequate to measure all the aspects of a hydrograph including the low flows, peak flows and shape of the hydrograph, and water balance simultaneously, since the different objective functions were adopted for different purposes. Single objective functions reduced the error with respect to simulation of some aspects of the hydrograph at the expense of other aspects; trade-off between overall root mean square error (overall agreement of shape of the hydrograph) and average RMSE of peak flow

events (agreement of peak flows) were less significant, while significant trade-offs between other objective functions were evident. Model calibration conducted using the multi-objective approach has been very successful in comparison to the single objective approach; simulations by the multi-objective functions integrated the characteristics of the single objective functions and primarily produced either improved or balanced results for the performance evaluation criteria, showing better agreement between observed and simulated flows. In particular, multi-objective calibration scheme 15 produced the best combination of objective functions to capture the catchment hydrological response. Weighting the R^2 by integrating the slope of the regression was recommended to obtain more reliable inferences, reflecting the systematic biases in the simulated flow.

Though this study was performed on one catchment by using one conceptual hydrologic model, the findings provide useful reference for rainfall-runoff simulation studies in tropical region catchments, since the choice of calibration objective is governed by the particular aspect of the hydrograph rather than by the catchment and the model. Further research is needed to identify additional performance evaluation criteria, like modified form of NSE, for more comprehensive reflection of model results.

Acknowledgement

Authors gratefully acknowledge Eng. J. M. M. U. Jayapadma and officials of the Lanka Hydraulic Institute, Sri Lanka for the technical assistance extended for the study.

References

1. Burlando, P., & Rosso, R., "Effects of Transient Climate Change on Basin Hydrology. 2. Impacts on Runoff Variability in the Arno River, Central Italy", *Hydrol. Process.*, Vol. 16, No. 6, 2002, pp. 1177-1199.
2. Douglgeris, C., Georgiou, P., Papadimos, D. & Papamichail, D., Evaluating Three Different Model Setups in the MIKE 11 NAM Model. In Lambrakis, N., Stournaras, G., & Katsanou, K., (Eds.), *Advances in the Research of Aquatic Environment. Environmental Earth Sciences*, Springer, 2011, pp. 241-249.



3. LΦrup, J. K., Refsgaard, J. C., Mazvimavi, D., "Assessing the Effect of Land Use Change on Catchment Runoff by Combined Use of Statistical Tests and Hydrological Modelling: Case Studies from Zimbabwe", *J. Hydrol.*, Vol. 205, No. 3-4, 1998, pp. 147-163.
4. Sadeghi, S., Saghafian, B., & Najarchi, M., "Assessment of Impacts of Change in Land Use and Climatic Variables on Runoff in Tajan River Basin", *Water Supply*, Vol. 20, No. 7, 2020, pp. 2779-2793.
5. Wickramaarachchi, T. N., Ishidaira H., Magome J. & Wijayaratna, T. M. N., "Projecting Turbidity Levels in Future River Flow: A Mathematical Modeling Approach", *Engineer: Journal of the Institution of Engineers, Sri Lanka*, Vol.48, No. 1, 2015, pp. 61-70.
6. Yan, R. H., Huang, J. C., Wang, Y., Gao, J. F. & Qi, L. Y., "Modeling the Combined Impact of Future Climate and Land Use Changes on Streamflow of Xinjiang Basin, China", *Hydrol. Res.*, Vol. 47, No. 2, 2016, pp. 356-372.
7. Refsgaard, J. C. & Knudsen, J., "Operational Validation and Intercomparison of Different Types of Hydrological Models", *Water Resour. Res.*, Vol. 32, No. 7, 1996, pp. 2189-2202.
8. Badjana, H. M., Fink, M., Helmschrot, J., Diekkrϋger, B., Kralisch, S., Afouda, A. & Wala, K., "Hydrological System Analysis and Modeling of the Kara River Basin (West Africa) using a Lumped Metric Conceptual Model", *Hydrol. Sci. J.*, Vol. 62, No. 7, 2017, pp. 1094-1113.
9. Croke, B. F. W., Merritt, W., & Jakeman, A. J., "A Dynamic Model for Predicting Hydrologic Response to Land Cover Changes in Gauged and Ungauged Catchments", *J. Hydrol.*, Vol. 291, No. 1-2, 2004, pp. 115-131.
10. Clilverd, H. M., Thompson, J. R., Heppell, C. M., Sayer, C. D., & Axmacher, J. C., "Coupled Hydrological/Hydraulic Modelling of River Restoration Impacts and Floodplain Hydrodynamics", *River Res. Appl.*, Vol.32, No. 9, 2016, pp. 1927-1948.
11. Gan, T. Y., Dlamini, E. M., & Biftu, G. F., "Effects of Model Complexity and Structure, Data Quality, and Objective Functions on Hydrologic Modeling", *J. Hydrol.*, Vol. 192, No. 1, 1997, pp. 81-103.
12. Jayapadma, J. M. M. U., Wickramaarachchi, T. N., Silva, G. H. A. C., Ishidaira, H., & Magome, J., "Rainfall-Runoff Modelling Using MIKE 11 (NAM Model): A Case Study of GIN River Basin", *Proceedings of the International Symposium on Advances in Civil and Environmental Engineering Practices for Sustainable Development (ACEPS-2018)*, Galle, Sri Lanka, March, 2018, pp. 231-238.
13. Makungo, R., Odiyo, J. O., Ndiritu, J. G., & Mwaka, B., "Rainfall-Runoff Modeling Approach for Ungauged Catchments: A Case Study of Nzhelele River Subquaternary Catchment", *Phys. Chem. Earth*, Vol. 35, No. 13-14, 2010, pp. 596-607.
14. Odiyo, J. O., Phangisa, J. I., & Makungo, R., "Rainfall-Runoff Modelling for Estimating Latonyanda River Flow Contributions to Luvuvhu River Downstream of Albasini Dam", *Phys. Chem. Earth.*, Vol. 50-52, 2012, pp. 5-13.
15. Shamsudin, S., & Hashim, N., "Rainfall-Runoff Simulation using Mike 11 NAM", *J. Civil Eng.*, Vol. 15, No. 2, 2002, pp. 1-13.
16. Engel, B., Storm, D., White, M., Arnold, J., & Arabi, M., "A Hydrologic/Water Quality Model Application Protocol", *JAWRA*, Vol. 43, No. 5, 2007, pp. 1223-1236.
17. Huang, Y., B ardossy, A., & Zhang, K., "Sensitivity of Hydrological Models to Temporal and Spatial Resolutions of Rainfall Data", *Hydrol. Earth Syst. Sci.*, Vol. 23, No. 6, 2019, pp. 2647-2663.
18. Kim, H. S. & Lee, S., "Assessment of a Seasonal Calibration Technique using Multiple Objectives in Rainfall-Runoff Analysis", *Hydrol. Process.*, Vol. 28, No. 4, 2014, pp. 2159-2173.
19. Madsen, H., "Automatic Calibration of a Conceptual Rainfall-Runoff Model using Multiple Objectives", *J. Hydrol.*, Vol. 235, No. 3, 2000, pp. 276-288.
20. Jie, M., Chen, H., Xu, C., Zeng, Q. & Tao, X., "A Comparative Study of Different Objective Functions to Improve the Flood Forecasting Accuracy", *Hydrol. Res.*, Vol. 47, No. 4, 2016, pp. 718-735.
21. Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W. & Sorooshian, S., "Effective and Efficient Algorithm for Multiobjective Optimization of Hydrologic Models", *Water Resour. Res.*, Vol. 39, No. 8, 2003, pp. 1214-1232.
22. Gupta, H. V., Sorooshian, S., & Yapo, P. O., "Towards Improved Calibration of Hydrological Models: Multiple and Noncommensurable Measures of Information", *Water Resour. Res.*, Vol. 34, No. 4, 1998, pp. 751-763.
23. Her, Y., & Seong, C., "Responses of Hydrological Model Equifinality, Uncertainty, and Performance to Multi-objective Parameter Calibration", *J. Hydroinform.*, Vol. 20, No. 4, 2018, pp. 864-885.

24. Prakash, O., Srinivasan, K. & Sudheer, K. P., "Adaptive Multiobjective Simulation-Optimization Framework for Dynamic Flood Control Operation in a River-Reservoir System", *Hydrol. Res.*, Vol. 46, No. 6, 2015, pp. 893-911.
25. Yapo, P. O., Gupta, H. V. & Sorooshian, S., "Multi-objective Global Optimization for Hydrologic Models", *J. Hydrol.*, Vol. 204, 1998, pp. 83-97.
26. Beven, K. J., *Rainfall-Runoff Modelling - The Primer*, 2nd ed., Wiley-Blackwell, 2012, 488 p.
27. Krause, P., Boyle, D. P., & Båse, F., "Comparison of Different Efficiency Criteria for Hydrological Model", *Adv. Geosci.*, Vol. 5, 2005, pp. 89-97.
28. ASCE, "Criteria for Evaluation of Watershed Models", *J. Irrig. Drain. Eng.*, Vol. 119, No. 3, 1993, pp. 429-442.
29. Boyle, D. P., Gupta, H. V. & Sorooshian, S., "Toward Improved Calibration of Hydrological Models: Combining the Strengths of Manual and Automatic Methods", *Water. Resour. Res.*, Vol. 36, No. 12, 2000, pp. 3663-3674.
30. Legates, D. R., & McCabe, G. J., "Evaluating the Use of Goodness-of-fit Measures in Hydrologic and Hydro-Climatic Model Validation", *Water Resour. Res.*, Vol. 35, No. 1, 1999, pp. 233-241.
31. Wickramaarachchi, T. N., *Hydrologic Modeling of the Gin River Basin Sri Lanka in Assessing the Environmental Effects due to Catchment Change*, PhD Dissertation, University of Yamanashi, Japan, 2014.
32. Wickramaarachchi, T. N., Flood Discharge Estimation in Baddegama using Pearson Type III and Gumbel Distributions. In Babel, M. S., Haarstrick, A., Ribbe, L., Shinde, V., Dichtl, N., (Eds.), *Water Security in Asia*, Springer Water, 2021, pp. 329-337.
33. Reuter, H. I., Nelson, A., & Jarvis, A., "An Evaluation of Void Filling Interpolation Methods for SRTM Data", *Int. J. Geogr. Inf. Sci.*, Vol. 21, No. 9, 2007, pp. 983-1008.
34. Hafezparast, M., Araghinejad, S., Fatemi, S. E., & Bressers, "A Conceptual Rainfall-Runoff Model Using the Auto Calibrated NAM Models in the Sarisoo River", *Hydrol. Curr. Res.*, Vol. 4, No. 1, 2013, pp. 1-6.
35. DHI, *MIKE 11 - A Modelling System for Rivers and Channels - Reference Manual*, Danish Hydraulic Institute (DHI), Hørsholm, Denmark, 2014, 512 p.
36. Vogel, R. M., & Fennessey, N. M. "L Moment Diagrams should Replace Product Moment Diagrams", *Water Resour. Res.*, Vol. 29, No. 6, 1993, pp. 1745-1752.
37. DHI, *MIKE 1D - DHI Simulation Engine for 1D River and Urban Modelling - Reference Manual*. Danish Hydraulic Institute (DHI), Hørsholm, Denmark, 2017, 334 p.
38. Duan, Q., Sorooshian, S., & Gupta, V. K., "Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models", *Water Resour. Res.*, Vol. 28, No. 4, 1992, pp.1015-1031.
39. Duan, Q. Y., Gupta, V. K., & Sorooshian, S., "A Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization", *J. Optim. Theory Appl.*, Vol. 76, No. 3, 1993, pp. 501-521.
40. Khakbaz, B., Imam, B., Hsu, K., & Sorooshian, S., "From Lumped to Distributed via Semidistributed: Calibration Strategies for Semidistributed Hydrologic Models", *J. Hydrol.*, Vol. 418-419, 2012, pp. 61-77.
41. Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T., "Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulation", *Trans. Am. Soc. Agric. Biol. Eng.*, Vol. 50, No. 3, 2007, pp. 885-900.
42. Zhao, B., Mao, J., Dai, Q., Han, D., Dai, H. & Rong, G. "Exploration on Hydrological Model Calibration by Considering the Hydro-Meteorological Variability", *Hydrol. Res.*, Vol. 51, No. 1, 2020, pp. 30-46.
43. Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F., "Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modeling", *J. Hydrol.*, Vol. 377, No. 1-2, 2009, pp. 80-91.
44. Gupta, H. V., Sorooshian, S., & Yapo, P. O., "Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration", *J. Hydrol. Eng.*, Vol. 4, No. 2, 1999, pp. 135-143.
45. Nash, J. & Sutcliffe, J., "River Flow Forecasting through Conceptual Models Part I - A Discussion of Principles", *J. Hydrol.*, Vol. 10, No. 3, 1970, pp. 282-290.
46. Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P., "Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria", *Trans. ASABE*, Vol. 58, No. 6, 2015, pp. 1763-1785.



47. Legates, D. R., & Davis, R. E., "The Continuing Search for an Anthropogenic Climate Change Signal: Limitations of Correlation-based Approaches", *Geophys. Res. Lett.*, Vol. 24, No. 18, 1997, pp. 2319-2322.
48. Vaitieknuiene, J., "Application of Rainfall-Runoff Model to Set Up the Water Balance for Luthuanian Basin", *Environ. Res. Eng. Manag. J.*, Vol. 31, No. 1, 2005, pp. 34-44.
49. Wickramaarachchi, T. N., Ishidaira, H. & Wijayaratna, T. M. N., "An Application of Distributed Hydrological Model, YHyM/BTOPMC to Gin Ganga Watershed, Sri Lanka", *Engineer: Journal of the Institution of Engineers, Sri Lanka*, Vol.45, No. 2, 2012, pp. 31-40.
50. Pfannerstill, M., Guse, B., & Fohrer, N., "Smart Low Flow Signature Metrics for an Improved Overall Performance Evaluation of Hydrological Models", *J. Hydrol.*, Vol. 510, 2014, pp. 447-458.