International Conference on Hydroinformatics

2014

Parameter Calibration Method Based On GA Technique For Multi-Event

Yeonjoo Kim

Eun-Sung Chung

Kwangjai Won

Byungho Kong

Kyungil Gil

PARAMETER CALIBRATION METHOD BASED ON GA TECHNIQUE FOR MULTI-EVENT

Yeonjoo Kim(1), Eun-Sung Chung(2) *, Kwangjai Won(2), Byungho Kong(2), Kyungik Gil(2)

(1): Korea Environment Institute, 215, Jinheung-ro, Eunpyeong-gu, Seoul, 122-706, South Korea

(2): Deptartment of Civil Engineering, Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul, 139-743, South Korea

**Correspondent: Eun-Sung Chung (eschung@seoultech.ac.kr)*

This study developed a robust parameter estimation (ROPE) framework of a rainfall-runoff model to consider multi-events with combining the Pareto optimal parameter sets and a composite (CP) programming, one of multi-criteria decision making (MCDM) approaches. The Pareto optimal parameter sets based on the Nash-Shutcliffe coefficient (NSE) were derived for a rainfall-runoff model with a generic algorithm. Then a robust parameter set among the Pareto optimums was selected using the composite values with NSE and peal flow error with the CP programming. Our case study in a small watershed in South Korea shows that the combined framework between traditional optimization techniques such as the Pareto optimality, and MCDM techniques that are not commonly used for the parameter selection problems of hydrologic models, could be an alternative approach for such parameter selection practices that could consider multiple aspects of model simulations.

1. INTRODUCTION

Watershed models have been used by researchers and decision makers to understand hydrological, ecological, and biogeochemical processes and to examine effects of human activities and climate change or variability on water quantity and quality. These models, however, require careful calibration of a large number of parameters mostly due to measurement limitations and scaling issues [1].

Some model parameters can be experimentally determined, some others have little or no physical meaning and their values cannot be obtained directly from measurable quantities of catchment characteristics. Therefore, it is often required to perform the model calibration for determining the model parameters, which is normally performed through either manual or automatic procedures. In particular, the need for automatic calibrations has been widely recognized over many years and increasingly emphasized since the calibrations with multiple objective functions or criteria were widely adopted [2]. Therefore, several automatic routines that use a multi-objective formulation of the calibration problem have been introduced in the rainfall-runoff modeling in recent years [3, 4, 5, 6].

The performance of the model can be evaluated with different performance measures that can quantify the goodness-of-fit between the simulated and observed data. For the hydrologic rainfall-runoff models, the critical performance measures include runoff volume, shape of hydrograph, peak and low flow timing, rate and volume and others [7]. To solve the multi-objective functions including multiple performance measures, the calibration techniques have been advanced along with the advances in computation powers. However, as more objective functions were included in the calibration, the number of exact or near Pareto optimal parameter sets increased and thus the calibration became a decision making problem of selecting a set of suitable model parameters from numbers of Pareto sets [2].

Furthermore, large uncertainty exists in determining weighting value on each objective function and the selection of each storm event in the multi-objective and multi-event optimization problems. These problems can be generally solved using Pareto optimum, but it will be still impossible to select a parameter set for real rainfall-runoff simulation since non-linearity of the hydrologic models and of the objective functions lead to very complex optimization problems [8]. Due to this reason, Beven *et al.*[9] argued that there is no optimum parameter set, in fact that there is a large set of parameter vectors which all perform reasonably and one cannot easily distinguish between them. They call this an *equifinality* problem which leads to high uncertainties in the model predictions. Cullmann *et al.* [10] showed that robust parameter estimation (ROPE) performs better in validation of small to medium sized events.

ROPE can be addressed using general multi-criteria decision making (MCDM) techniques since the high uncertainty in model prediction is closely related to the traditional decision making problem in the operation research field. However, there have been few studies for robust parameter selection that combines MCDM techniques with rainfall-runoff simulation models. Therefore, this study developed an ROPE framework for a rainfall-runoff model with combining the Pareto optimal parameter sets and a composite (CP) programming, one of MCDM approaches and applied the framework to the parameter selection of the SWMM for a river basin in Korea. Here the CP programming was employed to consider the multiple performance measures and multiple events in evaluating the derived Pareto parameter sets and the genetic algorithm (GA) was used in the optimization process.

2. METHODS

2.1 FRAMEWORK

The parameter selection framework proposed in this study. It sequentially combines two main procedures: 1) deriving the Pareto optimal parameter sets of two events with the GA and 2) applying the CP programming for the MCDM problems with considering the multiple events and performance measures. This framework derives the limited number of Pareto optimal parameter sets with using the different combinations of weights for two events that can be translated into the alternative parameter sets for the MCDM problem. The performance criteria for the MCDM include the multiple performance measures of the Nash and Sutcliffe Efficiency (NSE) and the peak flow error (PFE) for multiple events. Then the constructed MCDM problem is solved with the CP programming. In particular, the proposed framework utilizes the limited number of model simulations rather than computationally intensive calibration procedures for considering the multiple events and performance measures for the parameter selection.

2.2 STUDY WATERSHED

The suggested approach was applied for the parameter selection problem of the SWMM with the 5 rainfall-runoff events in the Milyang Dam basin, Korea. The Milyang Dam basin includes the area of 95.40 km^2 and exists in the mountainous area. The basin was divided into the 24 sub-basins and the river channels were divided into the 26 sub-channels for the SWMM.

The five events were divided into two different sets randomly: 2 events for the parameter optimization with the GA and 3 events for the decision making processes with the CP programming. The rainfall data from two telemetry (TM) stations of the Milyang Dam and Seoli were used. As shown in Figs. 3 and 4, all the rainfall-runoff events with the one peak were used in this study.

2.3 SWMM MODELING

2.3.1 MODEL OPTIMIZATION

In this study, the parameter optimization was performed with the GA [11, 12]. The GA is widely used for the parameter calibration procedures of the rainfall-runoff simulation, as it is known to be one of most effective techniques. The multi-objective GA (MOGA) such as the Elitist non-dominated sorting genetic algorithm (NSGA-II) Deb *et al.* [13] is also applied in many studies. The GA is a heuristic global search algorithm, based on the idea of Darwin's evolutionary processes of natural selection and survival of the fittest. The GA is far from a particular model structure and only requires an estimate of the objective function for each decision set in order to proceed. The advantages of GAs over conventional parameter optimization techniques are that they are appropriate for the ill-behaved problem, highly non-linear spaces for global optima and adaptive algorithm [3].

For the objective function of the optimization, the NSE Nash *et al.* [14] was used as follows:

$$\mathrm{E} = 1 - \frac{\sum_{t=1}^{T}\left(Q_o^t - Q_s^t\right)^2}{\sum_{t=1}^{T}\left(Q_o^t - \overline{Q_o}\right)^2} \tag{1}$$

where Q_o^t is observed discharge, and Q_s^t is simulated discharge at time t. $\overline{Q_o}$ is the averaged observed discharge.

Here the NSEs for two events were combined with the weighting factor as it follows:

$$\max_{\mathbf{x}}\ w_1 E_1 + w_2 E_2 \tag{2}$$

where $\mathbf{x}$ means the parameter set of hydrologic model and E_1 and E_2 are NSEs for two events and w_1 and w_2 are the weighted values on two events for the calibration ($w_1 + w_2 = 1$). This study used 11 cases (1.0 and 0.0, 0.9 and 0.1, … , 0.0 and 1.0) on values of w_1 and w_2.

While the Pareto optimal solutions of two objective functions can be derived with many different techniques such as MOGA, here this study used 11 different combinations of weights that are the alternatives of MCDM problems in the later step of the procedure.

Furthermore, the additional 3 events were used in this study for evaluating the derived optimal parameter sets. For those events, not only the NSE but also the PFE were evaluated as the PFE is given below:

$$PFE = \left|\frac{p_o - p_s}{p_o}\right| \tag{3}$$

where p_o and p_s are the peak flows for observed and simulated. Such multiple events and multiple performance measures function as the decision criteria for the MCDM problem, which will be further explained later.

The 18 model parameters subject to the model optimized in this study. While the parameters related to groundwater and channel characteristics are uniform over the Milyang Dam basin, the parameters related to basin characteristics, such as the percent of impervious area, characteristic width of the overland flow and runoff curve coefficient, differ for each sub-basin in this study.

2.4 COMPOSITE PROGRAMMING

The CP programming, which is a multi-level and multi-objective programming method, was introduced as an empirical technique to resolve a geological exploration problem by Bardossy *et al.* [15]. A general multi-objective problem can be transformed to a single objective problem. This transformation is done via a step-by-step regrouping of a set of objectives into a single objective. It uses indicators from different categories to calculate a composite distance, which identifies the distance of the actual system from the ideal state. Hence, schemes with small composite distances are closer to the ideal state than those with large composite distances [16].

The CP programming employs a double-weighting mechanism. One set of weights are indicators which articulate the decision-maker's preferences regarding the relative importance of each indicator. The other set are balancing factors given to groups in which any numbers of indicators are involved. Unlike weights, balancing factors are associated with groups rather than with each indicator. While the choice of weights emphasizes the relative importance of the indicators to each other, selecting the balancing factors identifies how larger deviations in groups of indicators may affect the process. The purpose of high balancing factors is to give more emphasis to the indicators which have large negative values [17].

Once the relevant indicators, associated boundary values (ideal and worst values), actual values and weights are determined, the first step is to normalize the basic values (transposing them into the range of 0 ~ 1). This is undertaken to make all indicators comparable to each other, thereby avoiding differences in units. Given the ideal value (σ_{ideal}), and the worst value (σ_{worst}), the normalized value (s_i) of an actual indicator value (σ_i) can be calculated as follows:

$$s_i = \frac{\sigma_{ideal} - \sigma_i}{\sigma_{ideal} - \sigma_{worst}} \tag{4}$$

where the choice is made to ensure that the to be used in the following equation represents the relative position with respect to the best value. The next step is to calculate second-level composite distances for each second-level group of basic indicators using the following equation:

$$L_j = \left(\sum_{i=1}^{N_j} w_{ij} s_{ij}^{b_j} \right)^{1/b_j} \tag{5}$$

where i represents a basic indicator, j a certain group of basic indicators, s_{ij} the value of the basic index s_i within the second-level group j, L_j the distance from the ideal point in second-level group j, N_j the number of basic indicators in a second-level group j, w_{ij} the weights expressing the relative importance of the N_j basic indicators in group j, the sum of weights in any group being equal to one, b_j the balancing factor, which is equal or greater than 1, among indicators within the group j. The consecutive computations of higher-level composite indices are made in the same manner until a final composite distance for a system is reached. The additional information can be found in the literatures, including [16, 18, 19].

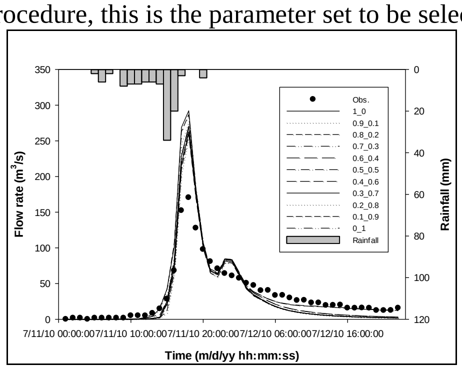
3. RESULTS AND DICUSSIONS

3.1 MODEL OPTIMIZATION FOR PARETO OPTIMAL SETS

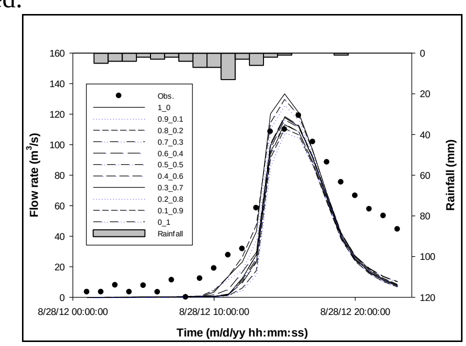
In this step, the Pareto optimal parameter sets of SWMM were derived with the different combinations of weights between two events. For 11 combinations of weights (Table 3), the GA is used to minimize the weighted sum of NSEs of two events. While the computationally expensive MOGA might work better for capturing the Pareto fronts, the limited number of combinations of different weights could also function to some extent with providing an opportunity to combine the Pareto optimal sets with the MCDM approaches for the parameter selection.

In Figure 1, the E1 dominant parameter set (i.e., the weight for E1 is larger than that for E2) presents higher flow rates in general. We also find a systematic error that the flow rate rapidly and markedly decreases after the peak. Therefore it is found that the peak flow is overestimated and the low flow in the tails of hydrograph is reasonable in E1 and in E2, the peak flow is reasonable and the low flow is underestimated.

Furthermore the NSEs of E1 and E2 clearly present the Pareto fronts. As pointed out already, it suggests that it is feasible to use the limited number of combinations of different weights in deriving the Pareto optimal sets in these problems. The best parameter set based on the NSEs of E1 and E2 is found in the case with NSEs of above 0.7 for both E1 and E2. In that case, the ratio of weight between E1 and E2 is 0.3:0.7. Without the CP programing in the next procedure, this is the parameter set to be selected.



(a) E1



(b) E2

Figure 1. Flood hydrograph of E1 and E2 with the 11 Pareto optimal parameter sets

3.2 PARAMETER SELECTION WITH THE COMPOSTIE PROGRAMMING

This procedure aims to select the best parameter set for the 3 events by combining the Pareto optimal parameter sets with the CP programming: the 2nd to 4th steps of the flow diagram.

First, SWMM simulations for the 3 events (E3, E4 and E5) were performed with the 11 Pareto optimal parameter sets (Figures. 2 and 3) and then the MCDM problem was constructed to select the best parameter set. The simulated flood hydrographs shows that all the Pareto parameter sets tend to underestimate the peak flows, which is apparent in E3 and E4. Furthermore, the low flows in the tails of flood hydrographs are underestimated as well and such underestimation is very clear in E3 and E5. As a result, the NSE and PFE of E5 are generally superior to those of other events.

Second, the first level composite solutions were calculated with evaluating the NSEs and PFEs of 3 events (Figure. 4). As expected, different performance measures led to different composite scores and rankings. Based on NSE, the parameter sets 7, 8 and 10 were ranked as the 3 most efficient parameter sets; based on PFE, the parameter sets 8, 10, 9 were among the best sets. For those top rankers, composite scores were lower, i.e., better for PFE than for NSE.

Third, the second level, final composite was estimated with assuming the equal importance among the 2 performance measures (Figure. 4). The parameter set 8, 10 and 9 were among the best with the CP scores of 0.609, 0.636 and 0.695, respectively. For the parameter set 8, selected as the best based on the CP programming, the ratio of weight between E1 and E2 is 0.3:0.7. The parameter set 2, 3 and 5 were among the worst with the CP scores of 1, 0.945 and 0.838, respectively.

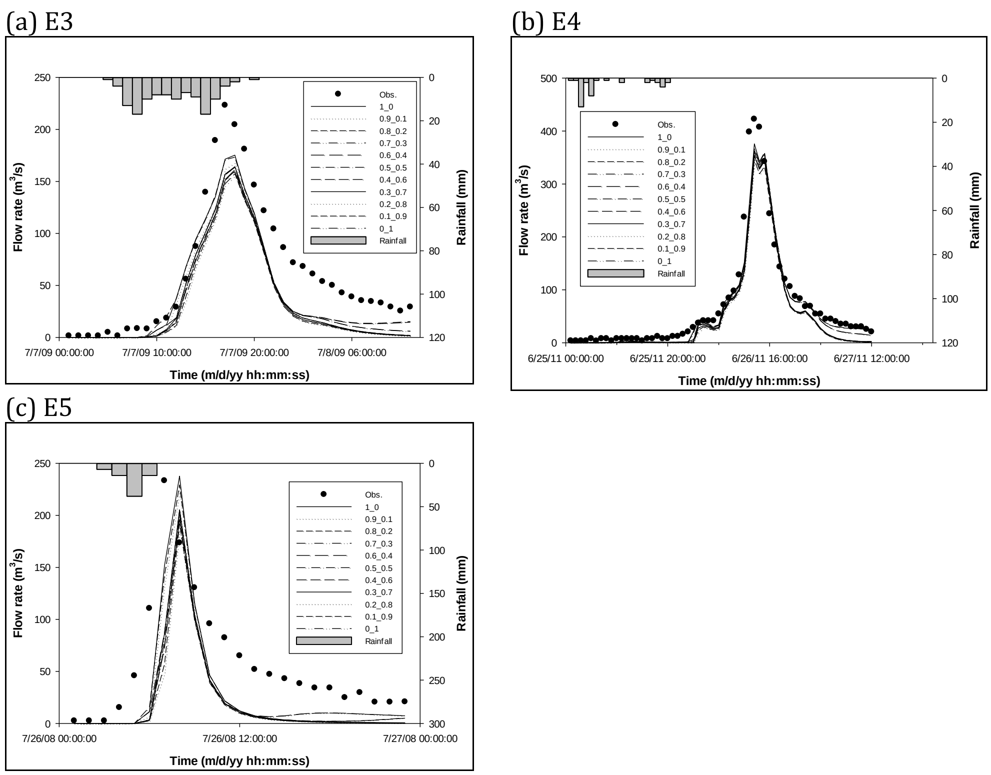


Figure 2. Flood hydrograph of E3, E4, E5 and E6 with the 11 Pareto optimal parameter sets

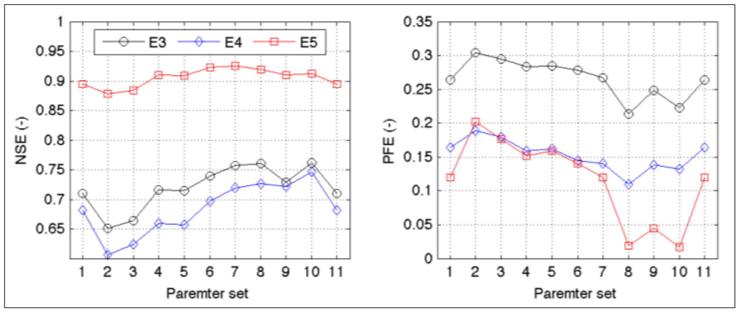


Figure 3. Calculated NSE and PFE values of each event with the 11 Pareto optimal parameter sets

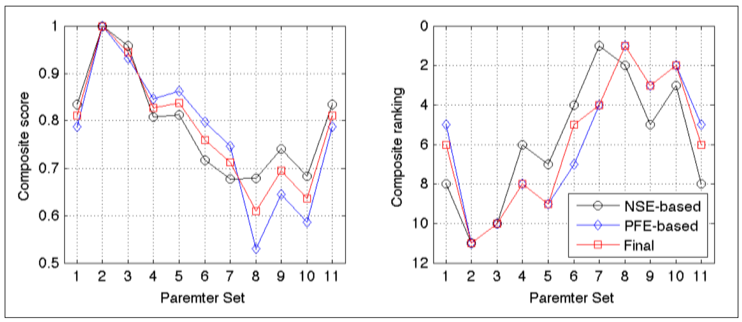


Figure. 4. Score and ranking of the composites

4. CONCLUSIONS

This study provided a robust parameter estimation (ROPE) framework that is the multi-objective optimization framework to select the robust parameter set with combining the Pareto optimal parameter sets and the CP programming. Pareto optimum can select all available parameter sets to any available weights on events and CP can give a way to derive a robust parameter set to other events and more performance criteria.

Our results show that the combined framework between traditional optimization techniques such as the Pareto optimality, and MCDM techniques that are not commonly used for the parameter selection problems of hydrologic models, could be an alternative approach for such parameter selection practices that could consider multiple aspects of model simulations. In the future, this procedure therefore can be extended to incorporate the multiple site observations for the robust rainfall-runoff calibration using multi-attribute decision analysis and multi-objective functions.

ACKNOWLEDGEMENT

This study was supported by the construction & transportation technology advancement research program by the KAIA under the ministry of Land, Infrastructure and Transport of Korea. This study was also supported by the funding from the Basic Science Research Program for the National Research Foundation of Korea.

REFERENCES

[1] Beven K. J., “Rainfall-Runoff Modeling: the Primer”, *John Wiley & Sons Press,* New York, (2000)

[2] Khu., S. T. and Madsen H., "Multiobjective calibration with Pareto preference ordering: an application to rainfall-runoff model calibration", *Wat. Res. Res*. 41, (2005), W03004, doi:10.1029/2004WR003041.

[3] Cheng Chun-Tian., Ou C P and Chau K W., "Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration", *J. Hydrol*. 268, (2006), pp 72–86.

[4] Hejazi M., Cai X and Borah D K., "Calibrating a watershed simulation model involving human interference: an application of multi-objective genetic algorithms", *Journal of Hydroinformatics*, 10(1), (2008), pp 97-111.

[5] Khu S T., Madsen H and di Pierro F., "Incorporating multiple observations for distributed hydrologic model calibration: An approach u sing a multi-objective evolutionary algorithm and clustering", *Advances in Water Resources*, 31, (2008), pp 1387-1398.

[6] Gong Y., Shen Z. Liu R. Hong Q. and Wu X., "A comparison of single- and multi-gauge based calibrations for hydrological modeling of the Upper Daning River watershed in China's Three Gorges reservoir region", *Hydrology Research*, 43(6), (2012), pp 822-832.

[7] Madsen H., "Automatic calibration of a conceptual rainfall-runoff model u sing multiple objectives", Journal of Hydrology, 235, (2000), pp 145-276.

[8] Bardossy A., and Sing S K., "Robust estimation of hydrological model parameters", *Hydrology and Earth System Sciences*, 12, (2008), pp 1273-1283.

[9] Beven K., and Freer J., "Equifinality, data assimilation, and data uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology", *Journal of Hydrology*, 249, (2001), pp 11-29.

[10] Cullmann J., Krausse T. and Saile P., "Parameterising hydrological models-Comparing optimization and robust parameter estimation", *Journal of Hydrology*, 404, (2011), pp 323-331.

[11] Holland J H., "Adaption in natural and artificial systems", Ann Arbor, MI: University of Michigan Press.

[12] Goldberg D E., "Genetic algorithms in search, optimization, and machine learning", Reading, MA: Addision-Wesley.

[13] Deb K., Agrawal S. Pratap A. and Meyarivan T., "A fast elitist nondominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In Proceedings of the Parallel Problem Solving Form Nature VI Conference (Lecture Notes in Computer Science) 1917, (ed. M. Schoenauer, et al.)", *Springer*, New York, (2000), pp. 849–858.

[14] Nash J E., and Sutcliffe J V., "River flow forecasting through conceptual models: Part 1 – A discussion of principles", *Journal of Hydrology*, 125, (1970), pp 221-241.

[15] Bardossy A., and Bogardi I., "Network Design for the Spatial Estimation of Environmental Variables", *Applied Mathematics and Computation*, 12, (1983), pp 339-369.

[16] Yurdusev M A., and O'Connel P E., "Environmentally-Sensitive Water Resources Planning: 1. Methodology", *Water Resources Management*, 19, (2005), pp 375-397.

[17] Goicoechea A., Stakhiv E Z. and Li F., "Experimental evaluation of multiple criteria decision models for application to water resources planning", *Water Resources Bulletin*, 28, (1982), pp 89-102.

[18] Hartmann L., Torno C. Bogardi I. and Higler L., "Methodological Guidelines for the Integrated Environmental Evaluation of Water Resources Development", Unep/Unesco Project FP/5201-85-01, Paris, (1987).

[19] Lee Y W., Bogardi I. and Stansbury J., "Fuzzy Decision Making in Dredged-material Management", *Journal of Environmental Engineering*, 122(4), (1991), pp 248-258.