



Automatic calibration and performance evaluation of a water quality model for a river greatly influenced by wastewater treatment plant effluent

Jae Heon Cho^{1*†} and Jong Ho Lee²

¹ Catholic Kwandong University, Gangneung-si, Korea

² Cheongju University, Cheongju-si, Korea

jhcho@cku.ac.kr, jhlee1013@cju.ac.kr

Abstract

When manually calibrating a water quality model, considerable time and attention are required to calibrate each water quality variable and model parameter. Hence, developing an automated model that allows for efficient and objective automatic calibration is highly desirable. The QUAL2Kw model calibrates the QUAL2K model automatically using a genetic algorithm (GA). As the calibration results of a GA vary strongly with the performance criteria used as the objective function in GA optimization problems, this study compares and analyzes auto-calibration results and selects the optimal criterion for each objective from among 6 performance criteria: CV(RMSE), R², NSE, PBIAS, RSR, and SSNR. Additionally, a multi-objective auto-calibration was conducted using two kinds of performance statistics as the objective function of the GA. The auto-calibration model was applied to the Youngsan River in Korea. The TMDL was established to achieve water quality goals at specific target points. Among the 6 auto-calibration results based on a single performance criterion, NSE was the best criterion for calculating fitness through auto-calibration. When the calibration accuracies of the TMDL target points and the entire river are considered simultaneously, an objective function using multiple performance criteria, specifically CV(RMSE) and RSR, was selected as the final auto-calibration of the model.

* Masterminded EasyChair and created the first stable version of this document

† Created the first draft of this document

1 Introduction

Manual calibration is commonly used in water quality modeling, but it may introduce errors and lacks objectivity. Thus, using an automatic calibration method with an optimization technique is preferable. Little and Williams calibrated the QUAL2E model using the least squares method [1]. Kim and Je determined the parameters of biochemical oxygen demand (BOD), dissolved oxygen (DO), and chlorophyll a content for QUAL2E using a nonlinear programming technique [2]. Li et al. demonstrated the importance of calibration data length when estimating the optimal parameters and uncertainty performance of conceptual rainfall-runoff models. The particle swarm optimization was used to optimize the parameters of the rainfall-runoff model [3]. Chlumecky et al. described a novel methodology and software for the optimization of a rainfall-runoff modeling using a genetic algorithm (GA) with the novel concept of a random number generator, which is the core of the optimization. Their goal was to optimize the calibration of the model with minimal user interaction [4]. Many additional mathematical optimization techniques have been used to estimate the parameters of models in various fields [5, 6, 7, 8].

Pelletier and Chapra developed the QUAL2Kw model by integrating the QUAL2K model and a GA. QUAL2K is one-dimensional river and stream water quality model intended to represent a well-mixed channel both vertically and laterally with steady state hydraulics, non-uniform steady flow, and diel heat budget and water-quality kinetics. Pelletier et al. auto-calibrated the parameters of a small stream using QUAL2Kw, and they studied the improvement in fitness related to the various crossover modes, population sizes, random number seeds, and evolutionary strategies [9, 10, 11]. In this study, the QUAL2Kw model was applied to a heavily polluted river in South Korea. The GA of QUAL2Kw was executed with an objective function with several kinds of criteria. The calibration results are compared and analyzed to identify the best performance criteria. Chlumecky et al. used various statistical indicators to validate a rainfall-runoff model. They focused on statistical methods and used the six most fundamental indicators, such as the Nash-Sutcliffe model efficiency (NSE), root mean squares error (RMSE) and correlation coefficient [4]. Li et al. used the popular NSE and water balance error percentage (WBE) indexes as objective functions in the auto-calibration model [3]. Mannina and Viviani [12] proposed a simplified river water quality model for small rivers and used generalized likelihood uncertainty estimation (GLUE) to apply this model to a small Italian river. The RMSE, NSE, and R^2 (coefficient of determination) values were reported as statistical performance indicators for four model state variables. The Soil and Water Assessment Tool (SWAT) was applied to a river catchment in England to quantify the long-term impacts of potential changes in agricultural management practices on river water quality [13]. Model performance was evaluated with three quantitative statistics, i.e., NSE, percent bias (PBIAS) and the ratio of the RMSE to the standard deviation of the observed data (RSR). In our study, auto-calibration was conducted using 6 performance criteria – the weighted average of the normalized RMSE (CV(RMSE)), R^2 , NSE, PBIAS, RSR, and the sum of the squares of the normalized residuals (SSNR); furthermore, multi-objective auto-calibration was also conducted using combinations of two kinds of performance criteria.

2 Method

2.1 Application of the water quality model to a polluted river

The QUAL2Kw model was applied to the Youngsan River located in southwestern Korea. The drainage area upstream of the total maximum daily load (TMDL) site Yeongbon B is approximately 533 km². The population density is high in the middle and lower reaches of the river. The effluent of

the Gwangju wastewater treatment plant (WWTP) is discharged directly into the Youngsan River, and the plant is located immediately above the confluence with the Gwangjucheon River. The effluent of the plant significantly affects the river quality and flow. The present water pollution level of the river is severe, and the river is one of the target areas of the government's TMDL program. The modeling area includes the drainage area upstream of Yeongbon B, and the TMDL target points in the study area are Yeongbon A, Yeongbon B, and Whangyong A. A map of the study area is shown in Figure 1.

Water quality and flow measurements for the TMDL target points were conducted over eight-day intervals by the Ministry of Environment, South Korea. For the water quality survey points excluding the TMDL target points, the Ministry of Environment measures water quality parameters on the main river channel each month. Model calibration was performed using the mean values of the five months of water quality data set (October 2016 to February 2017).

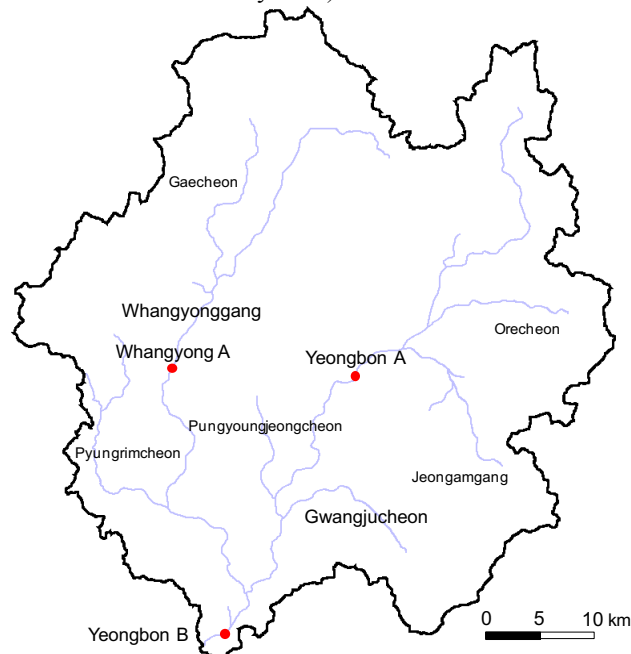


Figure 1: Study area of the Youngsan River basin

2.2 Evaluation of the performance criteria used as the objective function of a GA in auto-calibration

The calibration performance of the water quality model depends on the fitness of the GA in the auto-calibration model. The optimal objective function of a GA is the fitness calculation formula that minimizes the error between the calculated and observed water quality values. To analyze the model performance, it is necessary to consider the importance of each water quality variable and water quality survey site. The river in this study is currently a part of the Korean TMDL program, and the boundaries between local governments have been set as water quality target points. Specifically, BOD and total phosphorous (TP) are the target water quality variables for water quality management. Hence, among the numerous water quality variables measured at each water quality survey site, BOD and TP were assigned higher weights. Similarly, for the entire river, higher weights were given to target points in terms of achieving the water quality goals. To assess the fitness formula of the GA, CV(RMSE), R^2 , NSE, PBIAS, RSR, and SSNR were used as performance criteria, as they are often

used to analyze the performance of water quality and rainfall-runoff models. The calibration results using the six performance criteria are analyzed to identify the best performance criterion. The fitness formula for the GA in QUAL2Kw was modified so that the fitness values for NSE and R^2 are maximized while the errors of the other 4 performance criteria are minimized.

$$CV(RMSE) = \sum_{i=1}^n w_i \left[\frac{[\sum_{j=1}^m (P_{ij} - O_{ij})^2 / m]^{1/2}}{(\sum_{i=1}^m O_{ij} / m)} \right] \quad (1)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (P_i - \bar{O})^2} \quad (2)$$

$$PBIAS = \frac{\sum_{i=1}^n (O_i - P_i) \times 100}{\sum_{i=1}^n O_i} \quad (3)$$

Performance criteria for auto-calibration	NSE	R^2	CV(RMSE)	PBIAS	RSR	SSNR(site weighted)
CV(RMSE) (site weighted)	0.60	0.72	0.16	8.73	0.22	0.06
R^2	-2.35	0.73	0.80	49.46	0.56	1.00
NSE	0.68	0.79	0.18	4.74	0.20	0.06
PBIAS	-0.73	0.41	0.54	4.19	0.42	0.35
RSR	0.67	0.76	0.19	4.40	0.51	0.06
SSNR (site weighted)	0.59	0.72	0.16	9.02	0.01	0.05
SSNR (site weighted)+ RSR	0.57	0.72	0.17	10.18	0.22	0.05
CV(RMSE) (site weighted) + RSR	0.65	0.77	0.16	7.19	0.21	0.05

Table 1: Performance statistics determined from the auto-calibration using various performance criteria

$$RSR = \frac{[\sum_{i=1}^n (O_i - P_i)^2]}{[\sum_{i=1}^n (O_i - \bar{O})^2]} \quad (4)$$

$$\text{Sum of the squares of the normalized residuals} = \sum_{i=1}^n \left(\frac{O_i - P_i}{\frac{O_i + P_i}{2}} \right)^2 \quad (5)$$

Additionally, to improve the accuracy of calibration in terms of target points and the entire river, this study further examined the possibility of integrating two performance criteria as the objective function of the GA as follows.

$$\text{Objective function: Minimize } (\alpha(E) + \beta(RSR)) \quad (6)$$

where $E = CV(RMSE)$ or $SSNR$ and RSR is the ratio of the RMSE to the standard deviation of observed data, with the relationship $\alpha + \beta = 1$.

3 Results and discussion

3.1 Performance statistics calculated by auto-calibration using a single performance criterion

The auto-calibration performance statistics based on various performance criteria are shown in Table 1. Table 1 presents the calculation results of 6 different performance statistics, i.e., CV(RMSE), PBIAS, RSR, SSNR, NSE and R^2 values, for the auto-calibration results of water quality models using 8 performance criteria as the objective function.

The fitness of the model for CV(RMSE), PBIAS, RSR, SSNR was calculated as the inverse of each performance statistic using measured and calculated water quality values. And fitness of the model for the NSE and R^2 was calculated by its own statistics. High weights were given to carbonaceous BOD (CBOD) and TP based on the water quality goals established by the TMDL program. For rivers with established TMDLs in Korea, a higher calibration accuracy is required for TMDL target points. In the Youngsan River, Yeongbon A (17 km from the headwaters) and Yeongbon B (43 km from the headwaters), shown in Figure 1, are the main target points.

NSE is a normalized statistic that describes the degree of the 'goodness-of-fit' between model predictions and observations and can vary between $-\infty$ and 1, where a value of 1 represents a perfect fit. An NSE value between 0 and 1 is generally recognized to indicate acceptable model performance, and the model performance associated with an NSE value greater than 0.5 is satisfactory in simulating discharge, nitrate and TP loads at a monthly time step [13].

As shown in Table 1, among the 6 auto-calibration results based on a single performance criterion, the model's performance was the highest when NSE was used as the performance criterion to calculate fitness through auto-calibration, resulting in a generally low calibration error. The performances of NSE and R^2 were good, with values of 0.68 and 0.79, respectively. The values of PBIAS, RSR and SSNR were 4.74, 0.2 and 0.06, respectively, and the performances associated with these criteria were relatively good in comparison with the calibration results of the other performance criteria. When CV(RMSE) was used as the performance criterion, the model performance was greatest at the TMDL target points (Yeongbon A, 17 km from the headwaters; Yeongbon B, 43 km from the headwaters), as shown in Figure 2. Moreover, the calibration error was fairly small at the target points when SSNR was used as the performance criterion.

3.2 Model performance calculated by auto-calibration using multiple performance criteria

Because the TMDL target point with a designated water quality goal is situated at the border of two local government districts, the interests of two local governments are highly intertwined at this point. Since the Korean TMDL is being applied at the target point, the calibration accuracy at this point is the most important across all sections of the river. For the water quality target point, the fitness values and calibration accuracy tended to be high when CV(RMSE) and SSNR were used individually as the performance criteria. Therefore, this study examines a calibration method for the water quality model using the sum of these 2 performance criteria as the GA objective function to improve the calibration accuracy both at the target points and along the entire river. The value for both α and β in Equation (6) was set to 0.5.

If NSE is used, the overall calibration accuracy of the river tends to be higher. However, since the maximum value of NSE is 1, it is difficult to combine NSE with other variables, such as CV(RMSE) or SSNR, for use as the objective function to improve the accuracy at the target points. When calculating the fitness, CV(RMSE) (or SSNR) were given weights according to the importance of the water quality survey site. Therefore, the calibration accuracy of the water quality calculation is higher at target points of higher importance. For RSR, the overall calibration error is low, although no weights were given to different survey points. Hence, the performance of the water quality model after calibration using an objective function based on 2 performance criteria (Equation 6), namely CV(RMSE) (or SSNR) and RSR, was calculated and compared with the results with the aforementioned results based on the 6 single performance criteria.

The calibration accuracy of the target points was significantly higher when the objective function involved a combination of CV(RMSE) and RSR than when the objective function involved a single performance criterion. In this case, the CV(RMSE) and SSNR values, calculated with weighted values for each water quality survey point, were 0.16 and 0.05. The model calibration performance was good at not just the target points but throughout the entire river when no further weights were given to

important water quality survey points, as the model yielded an NSE of 0.65 and an R^2 of 0.77. The results of SSNR+RSR were also fairly good, although the calibration results of CV(RMSE)+RSR were far greater. When the auto-calibration results of CV(RMSE)+RSR was compared with the auto-calibration results of NSE based on a single performance criterion, the performance statistics of R^2 , NSE, and RSR were fairly similar. With values weighted per the importance of each TMDL target point, the auto-calibration of CV(RMSE)+RSR yielded slightly better results than the auto-calibration of SSNR+RSR. Therefore, this study chose the multiple performance criteria combination of CV(RMSE)+RSR as the final calibration.

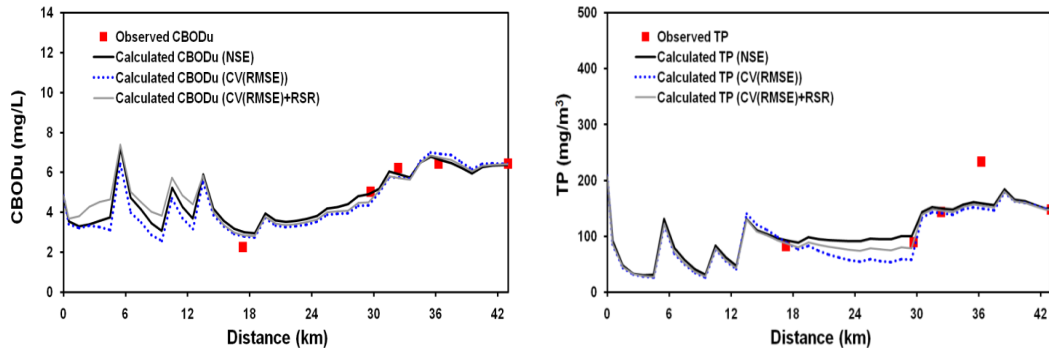


Figure 2: Calibration results for the main stream of the Youngsan River

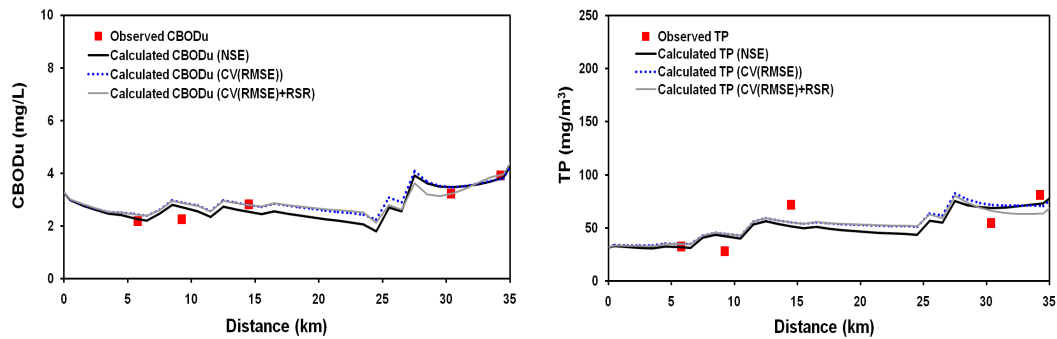


Figure 3: Calibration results for the Whangyonggang stream

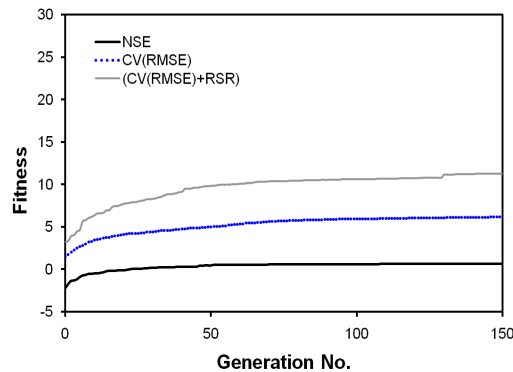


Figure 4: Changes in GA fitness according to the number of generations

The calibration results of the water quality model for the Youngsan River are shown in Figures 2 and 3. The calibration results for the two water quality variables (CBOD and TP) show good correspondence. Figure 4 shows the change in GA fitness when three objective functions, i.e., NSE, CV(RMSE), and CV(RMSE)+RSR, were used in the calculation for up to 150 generations. When NSE was used as the objective function, the final value was reached in a fairly early generation, whereas the calculation using CV(RMSE)+RSR and CV(RMSE) required at least 150 generations to reach the final value.

The water quality model with the auto-calibration using multiple performance criteria yielded slow CBOD hydrolysis rates of 0.08-3.88/d, slow CBOD oxidation rates of 0.51-4.96/d, fast CBOD oxidation rates of 0.02-4.90/d, organic P hydrolysis rates of 0.001-2.60/d, and organic P settling velocities of 0.0001-1.97 m/d (Table 2).

Model parameter	Range of calibrated values			Min.Max.	
	CV(RMSE)	NSE	CV(RMSE)+RSR		
Slow CBOD hydrolysis rate (/d)	0.52-4.97	0.02-4.23	0.08-3.88	0	5
Slow CBOD oxidation rate (/d)	0.003-4.98	0.06-4.78	0.51-4.96	0	5
Fast CBOD oxidation rate (/d)	0.07-4.99	0.12-4.99	0.02-4.90	0	5
Organic N hydrolysis rate (/d)	0.35-4.75	0.06-4.22	1.73-4.94	0	5
Organic N settling velocity (m/d)	0.27-1.75	0.05-1.58	0.16-1.83	0	2
Ammonium nitrification rate (/d)	0.88-7.89	0.02-7.61	1.41-9.02	0	10
Nitrate denitrification rate (/d)	0.75-1.99	0.45-1.52	0.43-1.38	0	2
Nitrate Sed. denitrification transfer coeff. (m/d)	0.03-0.83	0.02-0.96	0.16-0.95	0	1
Organic P hydrolysis rate (/d)	0.001-4.12	0.02-4.37	0.001-2.60	0	5
Organic P settling velocity (m/d)	0-1.97	0.001-1.67	0.0001-1.97	0	2
Inorganic P settling velocity (m/d)	0.001-1.99	0.001-1.99	0.0004-1.97	0	2
Detritus dissolution rate (/d)	0.08-4.69	0.60-4.99	0.15-4.33	0	5
Detritus settling velocity (m/d)	0.64-4.99	0.04-4.93	0.0001-4.98	0	5

Table 2: Estimated parameter values using QUAL2Kw and their minimum and maximum ranges

4 Conclusion

Auto-calibration of the parameters in a water quality model of a polluted Korean river in which TMDL limits have been implemented was conducted for water quality management purposes. The GA technique, along with various performance criteria, was used in the calibration process, and the calibration results were analyzed to identify the best performance criterion for use as the objective function of the GA. The QUAL2Kw was executed with several kinds of criteria as the objective function. Auto-calibration was conducted using the 6 performance statistics, and a multi-objective auto-calibration was also conducted through the integration of two kinds of performance criteria.

Among the auto-calibrations based on a single performance criterion, using the NSE criterion as the objective function yielded the highest model performance among all 6 auto-calibrated results for the entire stream. When CV(RMSE) was used as the performance criterion, the model performance was highest at the TMDL target points. Furthermore, when SSNR was used as the performance criterion, the calibration error was small at the target points.

When CV(RMSE) and RSR were used together as an objective function, the auto-calibration performance was good in terms of the CV(RMSE) and SSNR values calculated with weighted values for the target points, and the calibration accuracy was greater than that of the calibrations based on

other performance criteria. The NSE and R^2 values were satisfactory, and the performances of the model calibrations were good for all sections of the river. The performance of SSNR+RSR was fairly satisfactory, and it was close to the results of CV(RMSE)+RSR. Considering the importance of the TMDL target points, this study selected the auto-calibration results that utilized the multiple performance criteria combination of CV(RMSE)+RSR as the final calibration. When TMDL target points are not a concern, using the NSE performance criterion as the objective function for auto-calibration is more appropriate.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number: NRF-2017R1D1A1B03032816).

References

- [1] Little, K.W., Williams, R.E., 1992. Least-squares calibration of QUAL2E, *Water Environ Res* 64(2), 179-185.
- [2] Kim, K.S., Je, C.H., 2006. Development of a framework of automated water quality parameter optimization and its application, *Environ Geol.* 49, 405-412.
- [3] LI, C.Z., WANG, H., Liu, J., Yan, D.H., Yu, F.L., Zhang, L., 2010, Effect of calibration data series length on performance and optimal parameters of hydrological model, *Water Science and Engineering* 3(4), 378-393.
- [4] Chlumecky, M., Buchtele, J., Richta, K., 2017. Application of random number generators in genetic algorithms to improve rainfall-runoff modelling, *Journal of Hydrology* 553, 350–355.
- [5] Wood, D., Houck, M.H., Bell, J.M., 1998. Automated calibration and use of stream quality simulation model, *Journal of Environmental Engineering* 116(2), 236-248.
- [6] Je, C.H., Kim, K.S., 2004. Web-based application for estimating water quality impacts due to environmental dredging, *Environmental Geology* 46(2), 123-234.
- [7] Goktas, R.K., Aksoy, A., 2007. Calibration and verification of QUAL2E using genetic algorithm optimization, *J Water Res Pl-ASCE* 133(2), 126–36.
- [8] Zou, R., Lung, W.S., Wu, J., 2009. Multiple pattern parameter identification and uncertainty analysis approach for water quality modelling, *Ecol Model* 220, 621–9.
- [9] Pelletier, G.J., Chapra, S.C., 2006. QUAL2Kw theory and documentation(version 5.1): A modeling framework for simulating river and stream water quality, <http://www.ecy.wa.gov/programs/eap/models/>.
- [10] Pelletier, G.J., Chapra, S.C., Tao, H., 2006. QUAL2Kw - A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. *Environ Modell Softw.* 21, 419-425.
- [11] Chapra, S.C., Pelletier, G.J., Tao, H., 2007. QUAL2K: A Modeling Framework for Simulating River and Stream Water Quality, Version 2.07: Documentation and Users Manual, Civil and Environmental Engineering Dept., Tufts University, Medford, MA.
- [12] Mannina, G., Viviani, G., 2010. Water quality modelling for ephemeral rivers: Model development and parameter assessment, *Journal of Hydrology* 393, 186–196.
- [13] Taylor, S.D., He, Y., Hiscock, K.M., 2016. Modelling the impacts of agricultural management practices on river water quality in Eastern England, *Journal of Environmental Management* 180, 147-163.

