HAWQS v1.0

Calibration Process

Version 1.0 – Released September 2017

Importance of Calibrating and Validating a Water Quality Model:

Recently there has been an increase in the use of water quality models to evaluate the impacts of climate, land use, and management practices on the quantity and quality of water resources. To assure that a model's results are sufficiently accurate for these "real world" applications, calibration and validation of a water quality model are necessary. Calibration is an iterative process of testing model performance by adjusting input parameters in a way that the output from the model is reasonably close to observed values. Validation is an extension of the calibration process where the calibrated model is evaluated for a different period to assess if the calibrated model can reasonably represent the wide range of events in field observations. Therefore, the whole process of model calibration/validation is regarded as a systematic evaluation of errors or differences between model estimates and field observations as seen in Figure 1. While there are several methods of validating a model, the most commonly used procedure is the split-sample calibration/validation procedure. For a split-sample calibration/validation procedure only a portion of the available record of observed values is used for calibration, and once the final parameter set is established through calibration, simulation is performed for the remaining period of observed values, and goodness-of-fit between observed and simulated values is reassessed (Donigian, 2002).

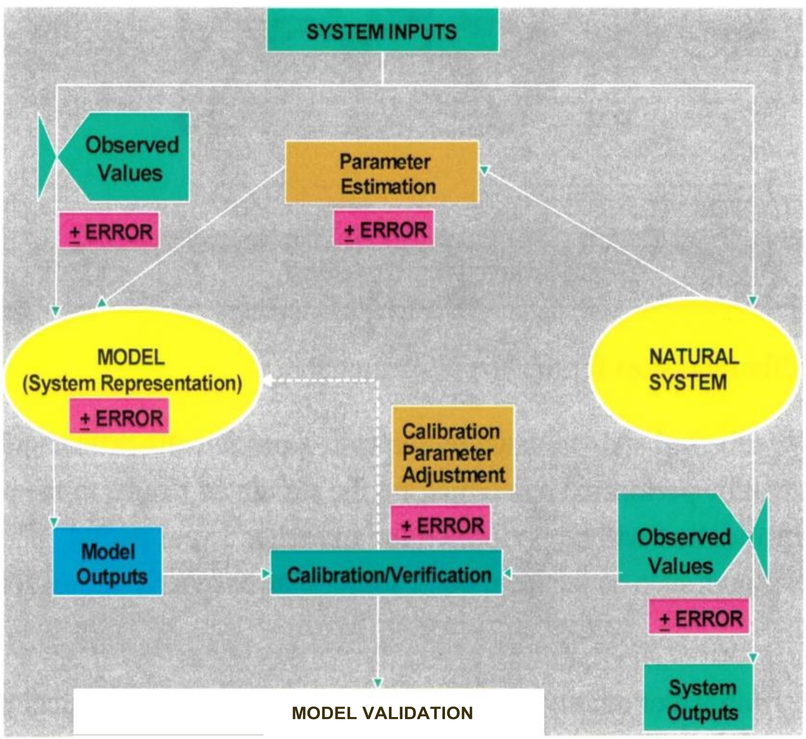

Figure 1: Calibration-Validation process (Donigian, 2002).

Methodology for Calibrating HAWQS

HAWQS was calibrated across the continental United States on a monthly basis. A total of 79 watersheds were selected in our initial calibration. The first watershed were chosen because gage stations are located near the outlet of the 8-digit watershed (if a calibrating site does not coincide with the outlet of an 8-digit watershed then simulated outputs are multiplied by the area ratio of HAWQS drainage area to observed USGS drainage area to make it comparable to the observed data set). The calibration sites were purposely chosen to exclude any reservoirs to prevent additional sources of uncertainty in calibration away from any immediate reservoir to avoid the influence of reservoir impacts on calibration. The sites selected for calibration are distributed across the contiguous United State as shown in Figure 2. Only 8-digit HUCs are calibrated, and 10 and 12-digit HUCs within the 8-digit HUCs received the same calibrated parameters for the corresponding calibrated 8-digit HUCs.

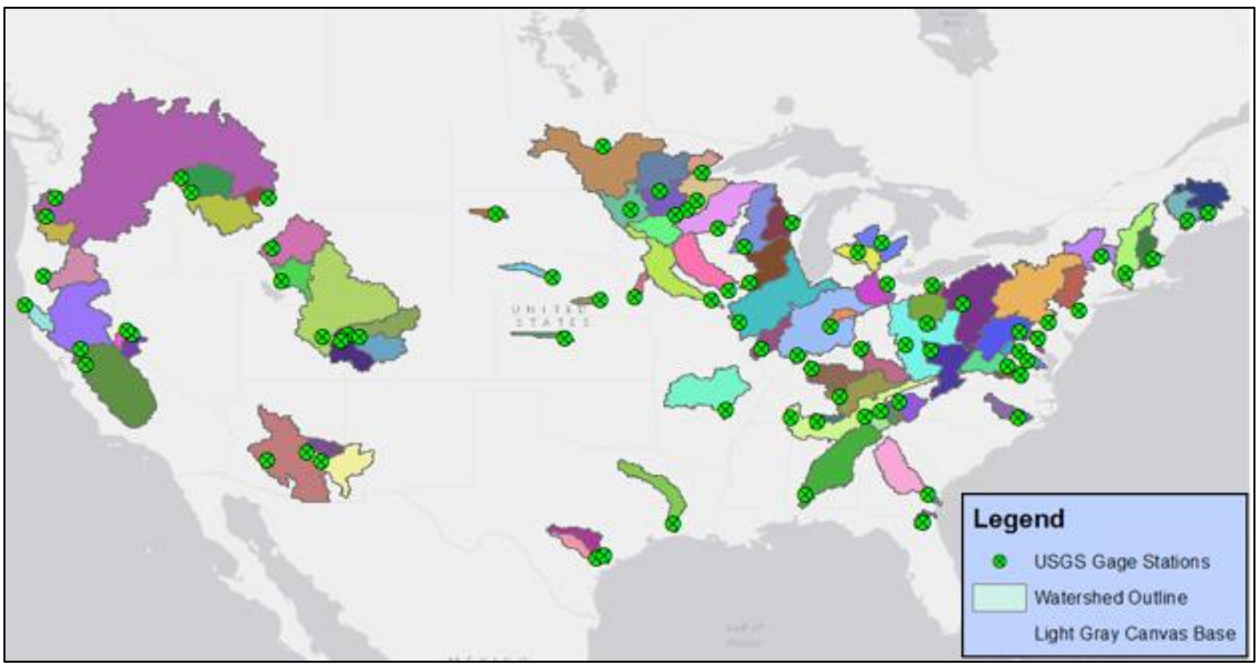

Figure 2: Calibration sites for the HAWQS

If two or more 8-digit HUCs in the same river were calibrated, the calibrated parameters were applied above each 8-digit HUC location until either head watershed was reached or another calibration 8-digit HUC was reached. The model was calibrated via the **split-sample calibration/validation** procedure as described in Figure 1, where calibration was performed by comparing the simulated results a minimum five years of observed flow data. However, water quality calibrations were done in some cases for selected gages where data was available for as few as two years. This is because some of the model inputs such as land use, point sources, atmospheric deposition, etc. are from the recent past.

Both hydrology and water quality were calibrated on a monthly basis through a hierarchical process beginning with hydrology calibration of streamflow, followed by sediment calibration (TSS), and finally calibration of water quality constituents such as total nitrogen (TN) and total phosphorus (TP) depending on the availability of observed data. Typically, model performance and calibration are evaluated through both qualitative and quantitative measures, involving both graphical comparisons and statistical tests. Graphical comparisons include:

1. Time series plots of observed and simulated values for streamflow, TSS, TP, and TN
2. Observed vs. simulated scatter plots, with a linear regression line, for streamflow, TSS, TP and TN
3. Cumulative frequency distributions of observed and simulated streamflow (e.g. flow duration curves)

Typical statistical criteria that were used to evaluate the relationship between the observed and simulated streamflow, TSS, TP, and TN are:

1. Coefficient of determination (R2),
2. Nash-Sutcliffe efficiency (NS or COE),
3. Root Mean Square Error (RMSE) and RMSE-observations Standard Deviation Ratio (RSE), and
4. Percent Bias (PBIAS) as recommended by Moriasi et al. (2007).

The R2 value indicates the consistency with which measured versus predicted values follow a best fit line, with 1.0 being optimal (Santhi et al., 2001). The NSE has been widely used to evaluate the performance of hydrologic models (Wilcox et al., 1990). The NSE value can range from negative infinity to 1, where a value of 1 indicates perfect model fit. The RSR value represents an error index recommended by Legates and McCabe (1999). RSR varies from the optimal value of 0 to l. Lower values of RSR indicate lower RMSE and better model simulation performance. PBIAS measures the average tendency of the simulated data to be larger or smaller than their observed counterparts. The optimal value of PBIAS is 0%, with low-magnitude values indicating accurate model predictions. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias (Gupta et al., 1999). Table 1 lists the performance ratings for the above-mentioned statistics as recommended by Moriasi et al. (2007).

Table 1: Performance statistics to evaluate hydrologic models (Moriasi et al. 2007)

			PBIAS (%)		
Performance Rating	*RSR*	*NSE*	*Streamflow*	*Sediment*	*TN and TP*
Very good	0 < RSR < 0.5	0.75 < NSE	PBIAS < ±10	PBIAS < ±15	PBIAS<±25
Good	0.5 < RSR < 0.6	0.65 < NSE < 0.75	±10< PBIAS < ±15	±15 < PBIAS < ±30	±25 < PBIAS < ±40
Satisfactory	0.6 < RSR < 0.7	0.50 < NSE < 0.65	±15 < PBIAS < ±25	±30 < PBIAS < ±55	±40 < PBIAS < ±70
Unsatisfactory	RSR> 0.7	NSE < 0.5	PBIAS > ±25	PBIAS >±55	PBIAS>± 70

The observational datasets that were used to calibrate/validate HAWQS are from long-term USGS benchmark monitoring stations. The calibrated parameters were applied to 10 and 12-digit watersheds within the calibrated 8-digit watershed, and no further calibration or statistics were computed at 10 and 12-digit watersheds. However, due to the uncertainties in weather, land use, land management, point sources, and atmospheric deposition, the statistics may vary across spatial resolutions.

Utilization of SWATCUP Program During Calibration Process:

The SWAT-CUP program is a stand-alone, public domain program that assists during calibration and validation of SWAT models. There are currently five optimization algorithms that the user can select to perform calibration and validation. The five algorithms are The Generalized Likelihood Uncertainty Estimation (GLUE), Sequential Uncertainty Fitting (SUFI2), Parameter Solution (ParaSol), Markov chain Monte Carlo (MCMC) and Particle Swarm Optimization (PSO). In this study, we used the SUFI2 algorithm to calibrate the SWAT model. The SUFI2 algorithm performs parameter sensitivity and uncertainty analysis to identify which parameters that contribute the most to the output variance relative to their input. In SUFI-2 parameter uncertainty accounts for all sources of uncertainty such as uncertainty in driving variables (e.g., rainfall), the conceptual model, parameters, and measured data. The simulation results are expressed by a 95 percent prediction uncertainty (PPU) band and as R-factor and P-factor. The P-factor is the percentage of the measured data bracketed by the 95PPU. This index provides a measure of the model's ability to capture uncertainty. The R-factor, on the other hand, is a measure of the quality of calibration and indicates the thickness of the 95PPU. Ideally, the P-factor should have a value of 1, indicating 100% bracketing of the measured data. A comprehensive description of the SUFI-2 algorithm can be found in Abbaspour et al. (2007).

The accompanying spreadsheet provides detailed statistics for each of the calibrated locations along with the parameters used for calibration. Overall the flow calibrations were good. However, in certain watersheds, there was limited observed water quality data available. In these cases, water quality data was modeled using the USGS LOADEST program. The monthly calibration was performed for 18 years in most locations. The water quality calibration varied from 2 to 18 years depending on the availability of the data. Table 2 provides typical parameters used for calibration. The individual calibration values and corresponding statistics for each site (Figure 2) were provided in a spreadsheet on the HAWQS website associated with this document.

Limitations of HAWQS Calibrations:

1. Transferring calibrated parameters between similar hydrologic regions may not always generate satisfactory results. In such cases expert opinion will be used to modify the calibration parameters.
2. Model outputs may not produce an acceptable match with observed data in locations downstream of reservoirs

Table 2: Typical parameters used during calibration (for detailed information please refer to SWAT_IO document and SWAT_CUP documents)

.bsn	ADJ_PKR	Replace
.bsn	CDN	Replace
.bsn	NPERCO	Replace
.bsn	PHOSKD	Replace
.bsn	PPERCO	Replace
.bsn	PRF	Replace
.bsn	PSP	Replace
.bsn	SDNCO	Replace
.bsn	SFTMP	Replace
.bsn	SMFMN	Replace
.bsn	SMFMX	Replace
.bsn	SMTMP	Replace
.bsn	SPCON	Replace
.bsn	SPEXP	Replace
.bsn	TIMP	Replace
.bsn	CNCOEF	Replace
.bsn	SURLAG	Replace
.bsn	ICN	Replace
.chm	LABP	Replace
.gw	ALPHA_BF	Replace
.gw	GW_DELAY	Replace/Add
.gw	GW_REVAP	Replace
.gw	GWQMN	Replace
.gw	RCHRG_DP	Replace
.gw	REVAPMN	Replace
.gw	SHALLST	Replace
.hru	CANMX	Replace
.hru	DEP_IMP	Replace
.hru	ESCO	Replace
.hru	SLSUBBSN	Relative %
.hru	EPCO	Replace
.hru	LAT_TTIME	Replace
.hru	HRU_SLP	Relative %
.mgt	CN2	Relative %
.mgt	DDRAIN	Replace
.mgt	USLE_P	Replace

.pnd	PND_ESA	Replace
.pnd	PND_EVOL	Replace
.pnd	PND_FR	Replace
.pnd	PND_PSA	Replace
.pnd	PND_PVOL	Replace
.pnd	PND_VOL	Replace
.res	NDTARGR	Replace
.res	IFLOD1	Replace
.res	IFLOD2	Replace
.res	STARG1	Relative %
.res	STARG10	Relative %
.res	STARG11	Relative %
.res	STARG12	Relative %
.res	STARG2	Relative %
.res	STARG3	Relative %
.res	STARG4	Relative %
.res	STARG5	Relative %
.res	STARG6	Relative %
.res	STARG7	Relative %
.res	STARG8	Relative %
.res	STARG9	Relative %
.res	RES_EVOL	Relative %
.rte	ALPHA_BNK	Replace
.rte	CH_K2	Replace
.rte	CH_N2	Replace
.rte	ch_cov1	Replace
.rte	ch_cov2	Replace
.swq	RS2	Replace
.swq	RS5	Replace
.sol	SOL_AWC	Relative %
.sol	USLE_K	Relative %
.sol	SOL_K	Relative %
.wwq	AI2	Replace

References:

Abbaspour K. C., M. Vejdani, and S. Haghighat. 2007. SWATCUP calibration and uncertainty programs for SWAT. In Proc. Intl. Congress on Modelling and Simulation (M0 DSIM' 07), 1603-1609. L. Oxley and D. Kulasiri , eds. Melbourne, Australia: Modelling and Simulation Society of Australia and New Zealand.

Donigian, A. S., Jr. 2002. Watershed model calibration and validation: The HSPF experience. In Proc. WEF Natl. TMDL Science and Policy, 44-73. Alexandria, Va.: Water Environment Federation.

Gupta, H. V., S. Sorooshian, and P. 0. Yapo. 1999. Status of automatic calibration for hydrologic models: Comparison with multi-level expert calibration. J. Hydrol. Eng. 4(2): 135-143.

Legates, D. R., and G. J. McCabe. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Res. 35(1): 233-241.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50(3): 885-900.

Santhi, C., J. G. Arnold, J. R. Williams, L. M. Hauck, and W. A. Dugas. 2001. Application of a watershed model to evaluate management effects on point and nonpoint-source pollution. Trans. ASAE 44(6): 1559-1570.

Wilcox, B. P., W. J. Rawls, D. L. Brakensiek, and J. R. Wight. 1990. Predicting runoff from rangeland catchments: a comparison of two models. Water Resour. Res. 26:2401- 2410.