Continuous Calibration

Uzair (Sam) Shamsi¹ and Joe Koran²

¹Jacobs Engineering Group, Pittsburgh, Pennsylvania; ²Metropolitan Sewer District of Greater Cincinnati, Cincinnati, Ohio.

Abstract

Generally, event based rainfall—runoff model calibration and verification is a laborious and expensive activity. With the availability of long term continuous flow meter and rainfall data and faster computers and software, models can now be calibrated to a long term flow record spanning from few months to several years. The practice of model calibration for large and complex regional systems is evolving over the years away from the conventional event based two step calibration and verification process to a more efficient single step continuous calibration process. This paper describes and compares the event and continuous calibration methods. It also presents information and comparison of various statistical methods to objectively quantify the goodness-of-fit statistics for model calibration that can be used in continuous calibration. Examples are provided from the Metropolitan Sewer District of Greater Cincinnati (MSDGC) SWMM model which is one of the largest SWMM model in the world.

1 Introduction

Model calibration and verification have become increasingly more challenging as rainfall—runoff models have increased in size and complexity. Rigid interpretation of numerical model calibration criteria as a model pass or fail test rather than a QA/QC measure causes expensive project delays. This paper presents new methods and innovative ideas to overcome model calibration challenges. For example, instead of applying the same calibration and verification criteria to all models, different criteria can be applied to different types of models (e.g. planning vs design models). This flexibility is expected to prevent under-calibration and over-calibration of models.

1 Definitions

Often, validation and verification terms are used synonymously (Duda et al. 2012; Donigian 2002). However, some organizations have different meaning for the two terms. To standardize the terminology and avoid confusion, the authors suggest the following definitions.

Model validation is the process of testing the accuracy and validity of an existing or previously calibrated model using recent rainfall and flow data (MSDGC 2013). Validation results are used to determine if the existing model can predict the current conditions or if it should be updated and re-calibrated to reflect the current system. Re-calibration is not required if the existing model meets the validation criteria. Model validation does not involve adjusting model input parameters as that is done during model calibration. Model validation is not applicable to new models.

Model calibration is the process of adjusting model input parameters to make modeled hydrographs reasonably match

the observed hydrographs using model calibration criteria. Any changes to the model should be made only where this reflects the physical state of the sewer system and not solely to make the model fit the observed data (USEPA 1979; USEPA 1999).

Model calibration should be followed by model verification using a different set of data to that used in the calibration (USEPA 1999; WaPUG 2002).

Model verification is the process of checking a model after calibration against independent data to determine model accuracy using model verification criteria. If model verification criteria are not met, calibration should be repeated by readjusting the previously calibrated model parameters.

Model validation, calibration and verification criteria may be the same or different.

2 Current Practice: Event Calibration

In the past when the computers and modeling software were too slow for continuous long term simulations, an event-by-event model calibration and verification (C&V) approach was used to meet specific model accuracy criteria. This approach continues to be the most common practice today. Event based model calibration consists of adjusting model input parameters, such as roughness, imperviousness or soil permeability until the difference between the modeled and observed event quantities is within acceptable accuracy criteria. Model verification consists of running the calibrated model with one or more independent events to verify the accuracy of the calibrated model. The accuracy is determined from the difference between observed and modeled flow depth, flow volume, peak flow rate, time-to-peak, and hydrograph shape. Ideally, the adequacy of model calibration can be assessed by comparing these five hydrograph properties.

In sewer design and capacity analysis applications, peak flow is more important. Therefore the model should be calibrated to minimize the difference in modeled and observed peak flow. In combined sewer overflows (CSO) consent decree compliance (e.g. 85% CSO volume capture) and storage sizing applications (e.g. deep tunnels and storage basins), flow volume is more relevant than the peak flow. Therefore the model should be calibrated to minimize the difference in modeled and observed flow volume. Generally, 10% accuracy is acceptable during calibration, and 25% accuracy is acceptable during verification. If model verification does not produce accurate results, the model calibration is refined further. Model C&V steps are repeated several times (usually 10 to 20 times) until satisfactory results are obtained (Shamsi 2016).

2.1 WaPUG Criteria

A popular source of model calibration criteria is the *Wastewater Planning Users Group Code of Practice for the Hydraulic Modeling of Sewer Systems*. This document can be found at the WaPUG website. At the time of writing, the latest version is 3.001, published in November 2002 and amended December 2002 (WaPUG 2002). This document provides only event based model verification criteria, summarized in Table 1, which are commonly also used for model validation and calibration.

Table 1 WaPUG model verification criteria.

Criteria	Maximum Depth	Volume	Peak Flow	Time to Peak
Dry weather	Not specified.	±10%	±10%	±1 h
flow				
Wet weather	Depth of surcharge $= +1.64$ ft	+20% to	+25% to	Not specified
flow	(0.5 m) to -0.33 ft $(-0.1 m)$.	-10%	-15%	
	Unsurcharged depth = ± 0.33 ft			
Minimum storm	(100mm). Monitoring: three storms and			
events	two dry days (page 36).			
	Calibration: not specified.			
	Verification: two of the three			
	monitored events (page 43).			

According to a 2013 American Water Works Association (AWWA) survey based on 209 survey responses from utilities, model calibration is the most technically challenging aspect of hydraulic modeling (AWWA 2014). A disproportionate amount of resources is often applied to the building, development and calibration of models compared to the analysis and interpretation of results. The cost of collecting calibration data is substantial, perhaps more than the modeling (James 2005). Using calibration criteria blindly and rigidly may result in unnecessary model re-calibration and re-verification efforts.

Sometimes, it may not be possible to meet all the model C&V criteria in all events at every calibration site. Quantitative and strict model C&V standards like WaPUG's were intended to be used as a quality assurance and quality control measure, or yardstick, by which the success of the C&V effort could be judged when combined with engineering judgment. Model C&V should

not be perceived as a model pass–fail test. Non-compliance with the calibration criteria does not necessarily equate to failure in some special situations. In fact Section 6.7, *Non-Compliance*, of WaPUG guidelines states that (WaPUG 2002):

It may still be possible to consider the model sufficiently verified in some circumstances provided that:

- the reasons for the non-compliance have been determined but cannot be modeled and have been assessed as not being important to the subsequent use of the model;
- the cause of the discrepancy cannot be isolated but an assessment of the effect of likely causes on the accuracy of the model has shown that this will not be detrimental to the purpose of the model; and
- infiltration is the cause of the discrepancy and this will be taken into account in other ways in subsequent use of the model.

This flexibility encourages the use of professional engineering judgment which is unfortunately not always practiced. Therefore, model C&V criteria should not be used as a rigid and inflexible mechanism to reject an entire model. Rather, model C&V status should be quantified (e.g. % calibrated by area or partially calibrated) rather than labelling an entire model simply as uncalibrated or unverified. For example, when a model or a portion of a model (e.g. a metershed) does not meet the model C&V criteria, the modeler can document the extent (e.g. percent area) to which the model is calibrated and verified, plans to calibrate and verify the model (e.g. in the design phase of the project), and arguments for and against proceeding with the project using the current model. The reasons for continuing with the project given the current model status may include the relative size of the areas that don't qualify as fully calibrated and verified, or the degree to which the model (or parts thereof) are out of C&V criteria.

Table 2 shows Shamsi's model application criteria which can be used to make the best use of the partially calibrated and/ or verified models (Shamsi 2016; Shamsi et al. 2016).

Table 2 Model application criteria.

Model C&V Status	Planning	Preliminary Design	Final design
Calibrated and verified	Yes	Yes	Yes
Calibrated and partially verified	Yes	Yes	No
Partially calibrated and verified	Yes	Yes	No
Calibrated (not validated)	Yes	Yes	No
Verified (not calibrated)	Yes	Yes	No
Partially calibrated	Yes	No	No
Not calibrated and verified	No	No	No

In 2014, the WaPUG website acknowledged that their modeling code of practice was some twelve years old and in need of a major update to reflect the current needs of the industry. A revised version is expected in 2017.

2.2 USEPA Guidelines

The United States Environmental Protection Agency (USEPA) provides model calibration guidelines in a 1999 CSO guidance document (USEPA 1999). This document provides only event based model C&V guidelines rather than a WaPUG-like numerical criteria. They key points of the USEPA guidelines are listed below.

- Calibration is the process of running a model using a set of input data and then comparing the results to actual measurements of the system. Validation (elsewhere called verification) is the process of testing the calibrated model using one or more independent data sets. If validation fails, the modeler must recalibrate the model and validate it again using a third independent data set. Validation is important because it assesses whether the model retains its generality; that is, a model that has been adjusted extensively to match a particular storm might lose its ability to predict the effects of other storms.
- An uncalibrated model may be acceptable for screening purposes, but without supporting evidence the uncalibrated result may not be accurate.
 To use model simulation results for evaluating control alternatives, the model must be reliable.
- For calibration, the most important comparisons are total volumes, peak flows, and shapes of the hydrographs.
- 4. An adequate number of storm events (usually 5 to 10) should be monitored and used in the calibration. The monitoring period should indeed cover at least that many storms, but C&V are frequently done with 2 to 3 storms each.
- Common practice employs both judgment and graphical analysis to assess a model's adequacy.
 However, statistical evaluation can provide a more rigorous and less subjective approach to validation.
- It is desirable to calibrate the model to a continuous sequence of storms if it is to be applied to a continuous rainfall record.
- 7. Figure 1 (Exhibit 7-11 in the USEPA publication) provides a calibrated model example. It compares an observed and modeled hydrograph to conclude that the peak flow, shape of the hydrograph, and the total volume of overflow for this calibration run are very close to the measured values even though the peak flow differs by about 20 cfs (0.57 m3/s) or 28%.

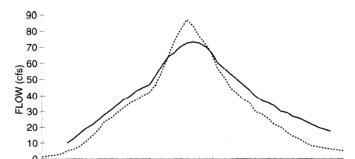


Exhibit 7-11. Flow Hydrograph

TIME OF DAY (hours)

7:00 7:30 8:00

----- Predicted ······· I Measured

8:30

Figure 1 Event calibration example: observed (measured) vs modeled (predicted) hydrographs (USEPA 1999).

6:00 6:30

2.3 Disadvantages of Event Based Model Calibration

- 1. Event based model C&V is generally laborious and expensive, especially for large models with tens of thousands of nodes and links.
- 2. Visual comparison of hydrograph shape is subjective and cannot be quantified.
- Often the three hydrograph parameters (volume, peak flow and time-to-peak) do not calibrate equally well. If the volumes are comparable, peaks are different, or vice versa.
- 4. Often there is no exit strategy if strict model C&V criteria are not met mostly due to inadequate observed data causing costly consent decree compliance delays. An endless cycle of model re-calibration and re-verification in turn impacts future projects.
- Sometimes additional flow monitoring is recommended but there is no guarantee of eventual model calibration or verification.
- 6. Most consent decree wet weather projects require compliance (e.g. 85% CSO volume capture or 4 CSOs per year) under typical year hydrologic conditions. A model best represents the typical year (average) condition when a range of events are used for both calibration and verification. Larger wet weather events generally used in the event based C&V can bias the model results to extreme rather than the typical year conditions. Smaller wet weather events are generally not used in the event based C&V process even though they account for nearly 90% of the typical year annual flow at most places.

3 Alternative Approach: Continuous Calibration

The practice of model development for large and complex regional systems is evolving over the years away from the two step calibration and verification process to a single step calibration process. Historically, because of limited flow monitoring data, a few (two to four) storm events would be used for calibration and one or two storms for verification. During the calibration process, the model parameters would be adjusted to best fit all the monitored data. With the availability of long term continuous flow meter data and faster computers, models can be now calibrated and verified to a long term flow record. As an alternative to the current event based model calibration approach, a continuous model calibration approach is suggested which uses statistical criteria as described below.

There are a variety of objective functions and statistical measures that may be used to measure the goodness-of-fit between a long term continuous measured and a modeled hydrograph. Statistical measures such as integral square error (ISE), Nash–Sutcliffe efficiency (NSE), and coefficient of determination (R^2) can be used as a single, non-subjective, statistical measure of model C&V rating. CHI's PCSWMM software automatically calculates the following eight error function values within the calibration plots avoiding laborious spreadsheet analysis after each model run.

- 1. Integral square error (ISE);
- 2. Nash-Sutcliffe efficiency (NSE);
- 3. Coefficient of determination (COD or R2);
- 4. Standard error of estimate (SEE);
- 5. Least squares error (LSE);
- 6. Least squares error dimensionless (LSED);
- 7. Root mean square error (RMSE); and
- 8. Root mean square error dimensionless (RMSED).

The error function values are updated on the fly as the user pans and zooms to different parts (i.e. events) of the continuous calibration plot.

3.1 Integral Square Error

The integral square error, ISE, integrates the square of the error e over time. That is:

$$ISE = \int \varepsilon^2 dt \tag{1.1}$$

ISE is calculated using Equation 1.2:

$$ISE = \frac{\left(\sum_{i=1}^{N} (O_i - M_i)^2\right)^{1/2}}{\sum_{i=1}^{N} O_i} \times 100$$
(1.2)

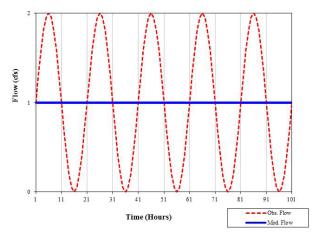
where:

 O_i = observed hydrograph value at time i,

 $M_i = \text{modeled hydrograph value at time } i$, and

N = number of hydrograph values.

ISE penalizes large errors more than smaller ones since the square of a large error is much bigger. This is demonstrated in Figure 2 which compares two hypothetical modeled and observed hydrographs with large (top) and small (bottom) peak flow differences. The top hydrograph with large peak flow differences has an ISE value of 7. The bottom hydrograph with smaller peak flow differences has a smaller ISE value of 1.75. The top hydrograph also shows that positive and negative differences do not cancel each other out to give a perfect ISE value of zero.



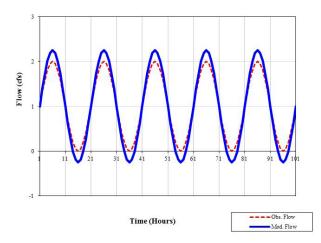


Figure 2 Effect of peak differences on ISE calibration.

Marsalek et al. (1975) examined three urban runoff models, namely the Road Research Laboratory Model (RRLM), the Storm Water Management Model (SWMM), and the University of Cincinnati Urban Runoff Model (UCURM), by comparing the modeled and observed hydrographs on several urban watersheds. This comparison was done for the hydrograph peak points as well as for the entire hydrographs using such statistical measures as the correlation coefficient, the special correlation coefficient and the integral square error. ISE was found to be a good measure of goodness-of-fit between observed and modeled hydrographs. Other sources supporting ISE include Smith and Vidmar (1994), Singhofen (2001), Shamsi (2002), James (2005), CDM (2007), and Shamsi and Ciucci (2013).

The second column of Table 3 (Marsalek et al.1975) provides goodness-of-fit or calibration ratings for different ISE ranges.

Table 3 ISE goodness-of-fit ratings for model calibration.

ISE Range	Calibration Rating	Model Application
0 to 3	Excellent	Planning, Preliminary Design, Final Design
3.1 to 6	Very good	Planning, Preliminary Design, Final Design
6.1 to 10	Good	Planning, Preliminary Design
10.1 to 25	Fair	Planning
>25	Poor	Screening

As discussed in Section 2, model calibration criteria should be based on the model application. For example, the following ISE calibration criteria can be used.

- planning applications: ISE ≤ 25 (fair to excellent);
- preliminary design applications: ISE ≤ 10 (good to excellent); and
- final design applications: ISE ≤ 6 (very good to excellent).

If the ISE rating of a calibrated model is excellent, with an ISE value between 0 and 3, the model can be considered suitable for all applications of the project, thus eliminating the need for costly additional monitoring and modeling.

3.2 Nash-Sutcliffe Efficiency (NSE)

This method has gained wide acceptance and is a good choice for a dimensionless measure of fit (Green and Stephenson,1986, in Pretorius et al., 2013). NSE is calculated using Equation 2:

$$NSE = 1 - \frac{\sum_{i=1}^{N} (M_i - O_i)^2}{\sum_{i=1}^{N} (O_i - \overline{O})^2}$$
 (2)

where:

 O_i = observed hydrograph value at time i,

 M_i = modeled hydrograph value at time i,

N = number of hydrograph values, and

 \overline{O} = mean of observed values.

NSE can range from $-\infty$ to 1. An NSE value of 1 represents a perfect match between the observed and modeled hydrographs. An NSE value of 0 has the same predictive power as the mean of observed values. Table 4 shows NSE ratings suggested by the authors.

Table 4 NSE goodness-of-fit ratings for model calibration.

NSE Range	Calibration Rating	Model Application
0.5 to 1.0	Excellent	Planning, Preliminary Design, Final Design
0.4 to 0.49	Very good	Planning, Preliminary Design, Final Design
0.3 to 0.39	Good	Planning, Preliminary Design
0.2 to 0.29	Fair	Planning
<0.2	Poor	Screening

3.3 Coefficient of Determination (COD or R^2)

COD is the square of correlation coefficient *R* which measures the strength of a linear relationship between two variables (observed and modeled values). *R* is calculated using Equation 3:

$$R = \frac{\sum_{1}^{N} O_{i} M_{i} - \frac{\left(\sum_{1}^{N} O_{i}\right) \left(\sum_{1}^{N} M_{i}\right)}{N}}{\sqrt{\left(\sum_{1}^{N} O_{i}^{2} - \frac{\left(\sum_{1}^{N} O_{i}\right)^{2}}{N}\right) \left(\sum_{1}^{N} M_{i}^{2} - \frac{\left(\sum_{1}^{N} M_{i}\right)^{2}}{N}\right)}}$$
(3)

If observed and modeled values have a strong positive linear correlation, R is close to 1. An R value of exactly 1 indicates a perfect positive fit. The coefficient of determination, R^2 , is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. COD is a measure that determines how certain one can be in making predictions from a certain modeled hydrograph.

3.4 Standard Error of Estimate (SEE)

SEE represents the accuracy of model predictions. The square root of the average squared error of prediction is used as a measure of the accuracy of prediction (Equation 4). This measure is called the standard error of the estimate. Recall that the regression line is the line that minimizes the sum of squared deviations of prediction (also called the sum of squares error). The standard error of the estimate is closely related to this quantity and is defined below.

$$SEE = \sqrt{\frac{\sum_{i=1}^{N} (O_i - M_i)^2}{N}}$$
 (4)

3.5 Least Squares Error (LSE)

Least square fitting is a mathematical procedure for finding the best fitting curve to a given set of points by minimizing the sum of the squares of the offsets (the *residuals*) of the points from the curve. Simple least square regression (LSE) is a method for finding a line that summarizes the relationship between the two variables (modeled and observed values in this case), at least within the domain of the explanatory variable (PCSWMM 2016). The formula for calculating the LSE is given in Equation 5:

$$LSE = \sum_{i=1}^{N} (O_i - M_i)^2$$
 (5)

3.6 Least Squares Error Dimensionless (LSED)

Dividing LSE by squared observed values results in a dimensionless measure of relationship between the observed and measured values. It is calculated using Equation 6:

$$LSED = \sum_{i=1}^{N} \left(\frac{O_i - M_i}{O_i} \right)^2$$
 (6)

3.7 Root Mean Square Error (RMSE)

RMSE is a frequently used measure of the differences between observed and modeled values. It represents the sample standard deviation of the differences between modeled and observed values. The RMSE serves to aggregate the magnitudes of the errors in predictions into a single measure of predictive power. RMSE is calculated using Equation 7:

$$RMSE = \sqrt{\sum_{i=1}^{N} \left(\frac{\left(O_i - M_i \right)^2}{N} \right)}$$
 (7)

3.8 Root Mean Square Error Dimensionless (RM-SED)

Dividing RMSE by the mean of observed values results in a dimensionless measure of the differences between observed and modeled values. The RMSE dimensionless term is between 0 and infinity. A smaller value (closer to zero) indicates a better model. RMSED is calculated using Equation 8:

$$RMSED = \frac{\sqrt{\sum_{i=1}^{N} \left[\frac{\left(O_{i} - M_{i}\right)^{2}}{N} \right]}}{\overline{O}_{I}}$$
 (8)

Goodness-of-fit ratings similar to ISE method (Table 3 above) and NSE method (Table 4 above) could not be found for all the error functions and should be developed to support continuous calibration.

3.9 Advantages of Continuous Model Calibration

- 1. Statistical measures of goodness-of-fit provide a single and non-subjective measure of model validation, calibration, and verification.
- 2. Rather than calibrating and validating the model to few observed events, calibrating to the entire observation period can be accomplished.
- Event calibration can provide an excellent match between observed and modeled data at a few events but does not guarantee it for every event. Continuous calibration can provide a reasonable match for all the events.
- 4. Continuous calibration can ensure that there is no overall or seasonal bias in the simulations and that over-simulation and under-simulation of individual storms balance out over the course of the long term simulation. Best of all, because the model is calibrated to the entire observation period, there is no need to verify the model with independent events. In fact, when all observed events have already been used in calibration, no events are left for verification.
- 5. Continuous calibration provides better representation of the typical year flow conditions commonly used in CSO control planning as all events, small and

- large, are used for C&V. There is no systematic bias in the model output caused by using only larger events.
- 6. Continuous calibration using software like PCSWMM, which automatically computes calibration error on the fly, reduces labour hours as no cumbersome spreadsheet pre-processing of rainfall data (event selection) or post-processing of model results (% difference in depth, flow, and volume) is required outside the model. This time saving is usually sufficient to offset the increased model run time for continuous simulation.
- 7. Multiple and flexible C&V criteria provide potential cost savings from not having to continually re-monitor, re-calibrate and re-run the models.

4 Case Study

MSDGC provides sewerage collection and treatment services to approximately 230 000 residential and commercial and 250 industrial users in Hamilton County, Ohio. It serves a population of about 855 000 in an area covering approximately 290 mi² (751.09 km²) through approximately 3 000 mi (4 828.03 km) sanitary and combined sewers. The system includes seven major wastewater treatment plants treating 70 000 MG/y (264 978.82 ML/y), three package treatment plants, >120 pump stations, and five real time control (RTC) facilities. The treatment system has a dry weather capacity of 200 MGD (757.08 ML/d). The largest Mill Creek plant has a wet weather capacity of 440 MGD (1 665.58 ML/d). Figure 3 shows the MSDGC sewer system and seven sewage treatment plants and corresponding watersheds (basins).

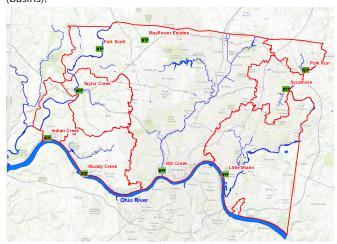


Figure 3 MSDGC service area showing seven sewage treatment plants and watersheds.

In 2003, MSDGC entered into a federal consent decree for control of combined sewer overflows (CSO) and sanitary sewer overflows (SSO). The consent decree requires completion of 116 Phase 1 projects totalling \$1.14 billion by December 2018. In addition, more than 260 Phase 2 projects totalling \$2 billion

should start in 2019. Compliance with the consent decree is based upon a *modeled problem with a modeled solution* approach which means that both the problem identification and the recommended solutions (projects) are based on a model. A separate calibrated SWMM model was created for each of the seven watersheds using USEPA's SWMM4 and SWMM5 software during the period 2001 to 2003. Combination of all the seven models is referred to as the *system wide model* (SWM). With >50 000 nodes and >30 000 subcatchments, MSDGC SWM is one of the largest SWMM model in the world. For additional information on MSDGC consent decree and model development history, please refer to the lead author's 2016 paper (Shamsi et al. 2016).

SWM was calibrated and verified during the period 2001 to 2003 using an extensive flow monitoring program and commonly practiced C&V criteria. In 2011, MSDGC adopted new modeling guidelines (MSDGC 2013) with more stringent model C&V criteria mostly based on event based WaPUG standards (WaPUG 2002). The 2001 and 2011 C&V criteria were different. For example, MSDGC used ±15% depth criterion because the absolute value of WaPUG depth criteria was difficult to meet for large sewers. The difference in the 2001 and 2011 model C&V criteria necessitated another round of model re-calibration and re-verification. Substantial flow monitoring and model re-calibration and re-validation has been conducted since 2011, but many sewersheds cannot appear to meet the WaPUG model C&V standards, which has resulted in frustration and delays in moving some consent decree projects to design and construction. WaPUG is expected to revise their model calibration standards for continuous simulation in 2017. MSDGC is also revising their modeling guidelines to use alternative calibration methods, such as those described above.

In 2011, the lead author requested Computational Hydraulics Int. (CHI) to implement automatic ISE calculation in the PCSWMM software; see Figure 4 (Event), Figure 5 (1 month continuous), and Figure 6 (5 month continuous for the entire flow survey period) for ISE calibration examples in PCSWMM for the MSDGC Bloody Run model. All figures indicate excellent ISE ratings of <3. ISE improves as the calibration period is increased from single event to long term continuous simulation indicating that it is easier to meet the ISE calibration criteria for longer calibration periods.

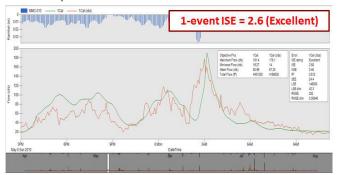


Figure 4 Single event ISE calibration.

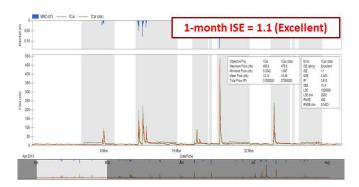


Figure 5 One month ISE calibration.

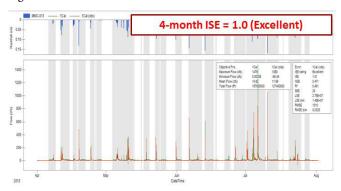


Figure 6 Four month ISE calibration.

Figure 7 shows various statistical error function values computed by PCSWMM for a single event (2013-05-08) which can be compared with WaPUG event criteria values shown in Table 5. The results are summarized below in Table 5.

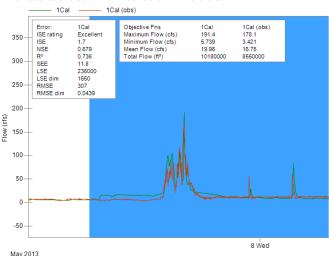


Figure 7 Comparison of various methods.

Table 5 WaPUG calculations.

Parameter	Modeled	Observed	% Difference
Peak Flow	191.4 cfs (5.4 m³/s)	178.1 cfs (5.0 m ³ /s)	7.5%
Volume	10 180 000 CF (288 265 m ³)	8 550 000 CF (242 109 m ³)	19.1%

- Peak flow difference = 7.5%, which meets the Wa-PUG peak flow criteria (+25% to -15%);
- WaPUG volume difference = 19.1%, which meets the WaPUG volume criterion (+20% to −10%);
- Integral square error (ISE) = 1.7, which meets the ISE calibration criteria with excellent calibration rating;
- Nash–Sutcliffe efficiency (NSE) = 0.679, which meets the NSE calibration criteria with excellent calibration rating;
- Coefficient of determination (R^2) = 0.736 (or R = 0.86, which is good);
- Standard error of estimate (SEE) = 11.8;
- Least squares error (LSE) = 236 000;
- Least squares error dimensionless (LSED) = 1 650;
- Root mean square error (RMSE) = 307; and
- Root mean square error dimensionless (RMSED) = 0.0439, which is very good.

These results indicate that error function values are consistent with the WaPUG results. That is, both the WaPUG and the error function criteria are satisfied. Because error functions are automatically computed, using them in lieu of or in conjunction with numerical criteria like WaPUG can reduce the model calibration time.

These results also indicate that error function goodnessof-fit ratings are reasonably but not perfectly consistent across all error functions. That is, different error functions can result in different calibration ratings especially because ratings have not yet been established for all the methods. Therefore, error functions should be used and interpreted carefully. Developing calibration ratings for all error functions is highly recommended.

6 Conclusions

Thanks to the advent of powerful personal computers and robust modeling software, it is now possible, and even more efficient, to calibrate the models to the entire flow monitoring period spanning from few months to several years. A model can be considered calibrated when a user-selected calibration criterion has been met which is flexible and can be adapted to meet the user's model application (planning vs design), schedule and budget. Rather than calibrating and validating the model to few observed events, calibrating to the entire observation period can be accomplished. Continuous calibration can ensure that there is no overall or seasonal bias in the simulations and that over-simulation and under-simulation of individual storms balance out over the course of the long term simulation. Best of all, because the model is calibrated to the entire observation period, there is no need to verify the model with independent events. Continuous calibration using software like PCSWMM, which automatically computes calibration error on the fly, reduces labour hours as no cumbersome spreadsheet pre- and post-processing of observed rainfall and flow data is required outside the model. Goodness-of-fit ratings based on statistical error functions have been developed to evaluate the adequacy of calibration for some but not all error

functions. Developing these ratings for all error functions is highly recommended.

References

- AWWA. 2014. "Committee Report: Trends in Water Distribution System Modeling." *Journal—American Water Works Association* 106 (10): 51–9. https://doi.org/10.5942/jawwa.2014.106.0145
- CDM. 2007. *Model Recalibration*. Tallahassee, FL: Blueprint 2000, Camp Dresser & McKee.
- Donigian, A. S., Jr. 2002. "Watershed Model Calibration and Validation: The HSPF Experience." In *Proceedings of the Water Environment Federation, National TMDL Science and Policy 2002*, 44–73. Alexandria, VA: Water Environment Federation. https://doi.org/10.2175/193864702785071796.
- Duda, P. B, P. R. Hummel, A.S. Donigian Jr. and J. C. Imhoff. 2012. "BASINS/HSPF: Model Use, Calibration, and Validation." *Transactions of the ASABE* 55 (4): 1523–47.
- Green, I. and D. Stephenson.1986. "Criteria for Comparison of Single Event Models." *Hydrological Sciences Journal* 31 (3): 395–411.
- James, W. 2005. *Rules for Responsible Modeling*, 4th ed. Guelph: CHI Press.
- Marsalek, J., T. M. Dick, P. E. Wisner and W.G. Clarke. 1975. "Comparative Evaluation of Three Urban Runoff Models." *Water Resources Bulletin* 11 (2): 306–28.
- MSDGC. 2013. Modeling Guidelines and Standards Volume I System Wide Model, revision 3. Cincinnati, OH: Metropolitan Sewer District of Greater Cincinnati.

 https://www.msdgc.org/downloads/customer_care/forms_and_documents/modeling/modeling_guidelines_and_standards.pdf
- PCSWMM Support. 2016. Error Functions. Guelph: Computational Hydraulics Int. https://support.chiwater.com/
- Pretorius, H., W. James and J. Smit. 2013. "A Strategy for Managing Deficiencies of SWMM Modeling for Large Undeveloped Semi-Arid Watersheds." *Journal of Water Management Modeling* R246. https://doi.org/10.14796/JWMM.R246-01.
- Shamsi, U. M. 2002. *GIS Tools for Water, Wastewater, and Stormwater Systems*. Alexandria, VA: ASCE Press.
- Shamsi, U. 2016. *MSDGC Model Review*. Cincinnati, OH: Metropolitan Sewer District of Greater Cincinnati, Jacobs Engineering Group.
- Shamsi, U. and R. Ciucci. 2013. "Innovations in CSO and SSO Long Term Control Plans." *Journal of Water Management Modeling* R246-25. https://doi.org/10.14796/JWMM.R246-25.
- Shamsi, U., B. Gamble and J. Koran. 2016. "Cincinnati's SWMM Model: A Journey Through Time." *Journal of Water*

- *Management Modeling* C398. https://doi.org/10.14796/JWMM.C398.
- Singhofen, P. J. 2001. "Calibration and Verification of Stormwater Models." Florida Association of Stormwater Utilities 2001 Annual Conference, June 20–22, 2001. http://www.streamnologies.com/support/pdfs/Calibration. pdf
- Smith, M. and A. Vidmar. 1994. "Data Set Derivation for GIS-Based Urban Hydrological Modeling." *Photogrammetric Engineering & Remote Sensing* 60 (1): 67–76.
- USEPA. 1979. Proceedings Stormwater Management Model (SWMM)

 Users Group Meeting, May 24–25, 1979. Washington, DC:

 United States Environmental Protection Agency. Report No 600/9-79-026.

- USEPA. 1999. Combined Sewer Overflows Guidance For Monitoring and Modeling. Washington, DC: United States Environmental Protection Agency. Report No 832-B-99-002.
- WaPUG. 2002. Code of Practice for Hydraulic Modeling of Sewer Systems Version 3.001 November 2002, Amended December 2002. London: Wastewater Planning Users Group. http://www.ciwem.org/wp-content/uploads/2016/05/Code-of-Practice-for-the-Hydraulic-Modelling-of-Sewer-Systems.pdf.