Battle of the Water Calibration Networks

Avi Ostfeld1; Elad Salomons2; Lindell Ormsbee3; James G. Uber4; Christopher M. Bros5; Paul Kalungi6; Richard Burd7; Boguslawa Zazula-Coetzee8; Teddy Belrain9; Doosun Kang10; Kevin Lansey11; Hailliang Shen12; Edward McBean13; Zheng Yi Wu14; Tom Walski15; Stefano Alvisi16; Marco Franchini17; Joshua P. Johnson18; Santosh R. Ghimire19; Brian D. Barkdol20; Tiit Koppel21; Anatoli Vassiliev22; Joong Hoon Kim23; Gunhui Chung24; Do Guen Yoo25; Keong Diao26; Yuwen Zhou27; Ji Li28; Zilong Liu29; Kui Chang30; Jinliang Gao31; Shaojian Qu32; Yixing Yuan33; T. Devi Prasad34; Daniele Laucelli35; Lydia S. Vamvakeridou Lyroudia36; Zoran Kapelan37; Dragan Savić38; Luigi Berardi39; Giuseppe Barbaro40; Orazio Giustolisi41; Masoud Asadzadeh42; Bryan A. Tolson43; and Robert McKillop44

Abstract: Calibration is a process of comparing model results with field data and making the appropriate adjustments so that both results agree. Calibration methods can involve formal optimization methods or manual methods in which the modeler informally examines alternative model parameters. The development of a calibration framework typically involves the following: (1) definition of the model variables, coefficients, and equations; (2) selection of an objective function to measure the quality of the calibration; (3) selection of the set of data to be used for the calibration process; and (4) selection of an optimization/manual scheme for altering the coefficient values in the direction of reducing the objective function. Hydraulic calibration usually involves the modification of system demands, fine-tuning the roughness values of pipes, altering pump operation characteristics, and adjusting other model attributes that affect simulation results, in particular those that have significant uncertainty associated with their values. From the previous steps, it is clear that model calibration is neither unique nor a straightforward technical task. The success of a calibration process depends on the modeler’s experience and intuition, as well as on the mathematical model and procedures adopted for the calibration process. This paper provides a summary of the Battle of the Water Calibration Networks (BWCN), the goal of which was to objectively compare the solutions of different approaches to the calibration of water distribution systems through application to a real water distribution system. Fourteen teams from academia, water utilities, and private consultants participated. The BWCN outcomes were presented and assessed at the 12th Water Distribution Systems Analysis conference in Tucson, Arizona, in September 2010. This manuscript summarizes the BWCN exercise and suggests future research directions for the calibration of water distribution systems. DOI: 10.1061/(ASCE)WR.1943-5452.0000191. © 2012 American Society of Civil Engineers.

CE Database object headings: Water distribution systems; Calibration; Hydraulic models; Optimization.

Author keywords: Water distribution systems; Calibration; Model; Optimization; Network.
Introduction

The Battle of the Water Calibration Networks (BWCN) is the third in a series of “Battle Competitions” dating back to the Battle of the Network Models (BNM) (Walski et al. 1987) and, more recently, the Battle of the Water Sensor Networks (BWSN) (Ostfeld et al. 2008).

Calibration is a process of comparing model results with measured data and making the appropriate adjustments so that both results agree. It usually involves analyzing why the model does not agree with measured data and then making the adjustments. The calibration objective is typically set such that the calibration process adjusts the model coefficient values within their feasible domains in the direction of reducing the calibration objective. Calibration methods can involve formal optimization methods or manual methods in which the modeler informally examines alternative model parameters (Ormsbee and Lingireddy 1997).

Water distribution system hydraulic models are based on equations of mass continuity and energy. The calibration process “tunes” system demands, roughness of pipes, pump operation characteristics, and other model characteristics such that model predicted values match reliable system data over a set of operational conditions.

Calibration, however, is far from being a technical problem as stated by Whitemore (2001): “Model efficacy depends heavily on experience and, in this sense, can be described as an art. While model formulation is more related to science, calibration has experience and, in this sense, can be described as an art. Good calibrations are more than curve fitting exercises that simply align model simulations with observed constituent behavior.” This nature of calibration was also noted by Walski (1990) and Savic et al. (2009).

The BWCN called for teams/individuals from academia, consulting firms, and utilities to address the challenge of water distribution system calibration by proposing, and applying, a calibration methodology to a real water distribution system. The results of the BWCN were presented at a special session of the 12th Water Distribution Systems Analysis (WDSA 2010) conference in Tucson, Arizona, in September 2010.

The objective of this manuscript is to summarize the outcome of the BWCN effort and to highlight future directions for research on the calibration of water distribution systems. This paper describes the following: (1) the BWCN rules and data, (2) a synopsis of the teams’ calibration approaches, (3) a comparison of the calibration results, and (4) conclusions and future research directions.

Problem Description

The municipality of C-Town is in need of a calibrated hydraulic simulation model for its water distribution system. To accomplish this task, the city has performed fire-flow tests and gathered data, with due diligence given to accuracy.

Data

The following presents the available information for C-Town as provided to the participants. All data are incorporated in C-Town.inp (U.S. EPA 2002, version 2.00.12) and C-Town.xls as supplemental files to this manuscript.

Network topology and mode of operation: The network topology is extracted from the C-Town geographic information system (GIS). The C-Town district meter areas (DMAs) and the system mode of operation are described in Figs. 1 and 2, respectively.

Elevations, pipes, pump curves, control rules, control valves, tanks, and sources: Junction elevations are extracted from a recent field survey or digital elevation model with an accuracy of ±1 m; pipe diameters and lengths, as well as the original manufacturer pump curves, are taken from the municipality GIS and file archives; pipe type and age are provided as best estimates; control rules are programmed into the programmable logic controllers (PLCs) of the pumping stations; records of the control rules are available; there are several pressure-reducing valves (PRVs) in the system whose settings are checked annually; all of the tanks are cylindrical, with tank diameters and minimum and maximum levels as reported in the municipality master plan document; the network is supplied from a constant headwater source; and, for some pipes, the lengths are more than the Euclidean distance between the upstream and downstream nodes as intermediate vertices were not available through the GIS system.

Isolation valves: It may be assumed that every pipe in the network has an isolation valve. Unless explicitly stated, all valves may be assumed to be fully open. Construction activity was ongoing in DMA2 (Fig. 1) during the calibration tests, which required isolating portions of the system. It is entirely possible that one or more valves in this part of the system may not have been fully re-opened following the upgrades. Fire-flow data were collected after the upgrades.

Demands and supervisory control and data acquisition (SCADA): Monthly water demands taken from billing records are given at each junction; hourly tank levels and pumping station flows are available for a period of 168 h (1 week).

Fire-flow tests: Fire-flow tests were conducted at each of the DMAs separately during the evening. It can be assumed that the basic system demands (excluding the observed hydrant demands) at the time of the fire-flow tests approximately correspond to the demands during hour 1 of the provided 1-week SCADA time series.
Assessment

Participants were required to submit an EPANET (U.S. EPA 2002) input file with their calibrated system for the 168 h corresponding to the times recorded in the SCADA data file. The BWCN rules provided a variety of possible measures for calibration assessment:

1. Maximum sum of squared relative errors (SSRE) and standard deviation errors for the estimated pipe roughness values, node

---

**Fig. 1.** C-Town layout

**Fig. 2.** C-Town system operation mode: the system is fed from a large seasonal reservoir (R1); the water is pumped through pumping station S1 to the lower tanks (T1 and T2); flow to T2 is controlled by a valve (V2), which is operated according to T2 levels; two pumping stations (S2 and S3) draw water from T2 to two higher tanks (T3 and T4); the two other stations (S4 and S5) pump water from T1 to T5, T6, and T7
Fourteen teams participated in the BWCN. This section gives a brief description of each team’s calibration approach.

- Kang and Lansey (2010) used the weighted least squares (WLS) method based on the Jacobian matrix, which provides analytical expressions for the sensitivities of the measurements to input parameters. The primary benefit of WLS is computation speed compared with random-sampling optimization approaches, which would be critical especially for real-time demand estimation.
- Shen and McBean (2010) linked a genetic algorithm with Monte Carlo simulations. The Monte Carlo simulations identified parameters to feed the calibration process, and the genetic algorithm minimized the discrepancy between observed and simulated data.
- Wu and Walski (2010) combined engineering judgment with optimization tools in an overall progressive approach, which first utilized the fire-flow test data to identify roughness, pump curves, and partly closed valves; then used SCADA data to back out demand patterns for each DMA; and finally made adjustments based on the entire system. They relied heavily on visualization tools and experience to identify problems and evaluate solutions.
- Alvisi and Franchini (2010) combined an automatic heuristic optimization process based on the SCE-UA (Shuffled Complex Evolution–University of Arizona; Duan et al. 1992), manual refinement, and the Monte Carlo–Latin hypercube approach (McKay et al. 1979), with the latter identifying the most influential input parameters (e.g., pipe roughness values) on the measured output pressures, pipe flows, and tank water levels.
- Johnson et al. (2010) presented a flow-sequential sector-specific lumped algorithm (FSL) for calibrating the pipe roughness values and user demands.
- Koppel and Vassiljev (2010) utilized the Levenberg-Marquardt algorithm (LMA) (Levenberg 1944; Marquardt 1963) with proper parameter domain adjustments for partial derivative computations (Koppel and Vassiljev 2009).
- Kim et al. (2010) used the metaheuristic harmony search (HS) (Geem et al. 2001), which is based on imitating a musical process of searching for the best state of harmony.
- Diao et al. (2010) introduced a heuristic methodology coupled with engineering judgment based on identifying dominant scenarios whose influence on the hydraulic system performance is substantial.
- Chang et al. (2010) adopted a two-stage genetic algorithm (Goldberg 1989) process in which the nodal demands were first calibrated, followed by the pipe resistance coefficients.
- Prasad (2010) developed an artificial immune heuristic population-based algorithm entitled clonal selection algorithm (CLONALG), which is based on evolving a population of candidate solutions using a clonal selection principle.
- Laucelli et al. (2010) implemented preliminary topological analysis of the network to spatially decompose and simplify the C-Town model. This stage reduced the model size by approximately 50%, leading to substantial computational savings (without compromising mass and energy balance equations) and enabling easier analysis of the results. The BWCN problem was then solved using a multiobjective optimization approach with three objectives targeting the minimization of different types of absolute relative error measures. The parsimony of the calibration procedure is noteworthy, as only eight parameters (of pipe roughness, valve coefficients, pump speed factor, and value ranges) have been calibrated. This is important for avoiding potential model overfitting (Kapelan 2010).
- Asadzadeh et al. (2010) used optimization algorithms based on dynamically dimensioned search (DDS) in which the fire-flow test measurements were first fitted using the multiobjective Pareto archived DDS (PA-DDS) algorithm (Asadzadeh and Tolson 2009) after which the demand pattern multipliers were calibrated using the single objective DDS algorithm (Tolson and Shoemaker 2007).

### Calibration Results

The BWCN result assessments are presented in Figs. 3–11. Fig. 3 describes examples of demands and pressures matching for one of the junctions (Junction 9). The correspondence between the modeled (predicted) and true (measured) demands/pressures for the ith system’s node are quantified through

\[
SSRE_i = \sum_{t=1}^{T} \frac{(g^m_{i,t} - g_{i,t})^2}{(g_{i,t})^2} \quad \forall \ i \in N
\]

where \(SSRE_i\) = sum of squared relative errors of the ith system’s node; \(T\) = number of time steps, indexed \(t\); \(N\) = number of system nodes, indexed \(i\); \(g^m_{i,t}\) = modeled (predicted) demand/pressure at the ith node at time \(t\); and \(g_{i,t}\) = true (measured) demand/pressure at the ith node at time \(t\). The SSRE in Fig. 3 for the demands is equal to 6.903, and for the pressures, it is equal to 0.630.

Fig. 4 presents the SSRE demand node distribution curves for all teams. Each team’s curve is formulated by top-down sorting of all its node SSRE outcomes and by assigning each SSRE value its corresponding accumulated fraction. For example, referring to Kim et al., no SSRE values are above ~46 (i.e., ~46 is the maximum obtained SSRE value at any of the nodes), 0.6 of the SSRE values are above ~23, and all SSRE outcomes are above ~19. All node SSRE values for all teams were summed up. Fig. 5 shows the best three SSRE accumulated demand outcomes: Alvisi and Franchini (lowest, first), Diao et al. (second), and Koppel and Vassiljev (third). Fig. 6 describes the team SSRE pressures node distribution curves, Fig. 7 shows the three lowest SSRE accumulated team pressures matching, and Fig. 8 presents the accumulated tank SSRE water levels matching.

Three areas were selected to compare the calibration of the submitted models to the true model. Parameters were used to determine how well the models calibrated pipe roughnesses, identified throttled valves, and reproduced fire-flow test results (Figs. 9–11).

For the pipe roughness results, each pipe’s error was determined from the square difference between the submitted model’s roughness and the true model’s roughness for the individual pipe, normalized with respect to the square of the true model’s roughness.
The SSRE for each pipe was calculated for each submission (Fig. 9). Pipes along the path from pumping station S3 to tank T4 in DMA2 were handled differently because of the different methods used to identify the throttled valves along this path (see next section).

To assess how well each submission accounted for the throttled valves, the total head loss along that path was used as a metric. Head loss was calculated for each submission along the path from pumping station S3 to tank T4 in DMA2 and compared with the head loss along the same path in the true model. The SSRE was calculated for each submission (Fig. 10).

For each of the fire-flow tests, the collected data at the test points were compared with the values predicted by each submitted model. The data included fire-flow pressures, static pressures, tank flows, and pump flows. The SSRE for each parameter and flow condition was calculated (Fig. 11).

Table 1 summarizes the ranks of the top three teams, Alvisi and Franchini (1), Wu and Walski (2), and Koppel and Vassiljev (3), assuming equal weights for all calibration rank categories. The top three teams adjusted pipe roughnesses, pump curves, and pump state change times. Alvisi and Franchini have also added minor losses.

![Fig. 3. Examples of demands and pressures matching for Junction 9 with SSRE values of 6.903 and 0.630, respectively](image)

![Fig. 4. (Color) Nondimensional SSRE demands fitting distributions (truncated above 50)](image)

![Fig. 5. Top three nondimensional SSRE demands fitting distributions](image)
Observations

The BWCN was a vital experience to all participants. This section presents a summary of the BWCN insights:

1. One should always remember that a model is a simplification of a real system and not the real system itself. Hence, it is important that educators explain and reinforce the point to engineering students that models are not reality and that their results should always be validated (i.e., tested against new data/operating conditions) using intuitive common sense.

2. Part of the value of model calibration lies in the additional insights gained through the process about the system, not just a better model. From this perspective, an important aspect can be the analysis of the existing network topology to account for the actual observability of the different parts of the network (i.e., pipes, nodes, and segments) as a consequence of the available measurements and observations.

3. The necessary accuracy of the model calibration is dependent on the proposed model application and prediction variables of interest. For example, a more detailed model with accurate flow/velocity estimates is required for a water quality problem than for a general planning application.
4. The ultimate measure of whether a model is calibrated well enough is the extent to which the system operators have confidence in the results of the model and then actually use it.
5. Calibrated models can be useful in identifying system anomalies such as partially closed valves, bad telemetry data, and so forth.
6. Although automated calibration models (e.g., those based on some type of optimization approach like genetic algorithms) can be very effective in fine-tuning model parameters, human knowledge and experience are needed to get the water distribution model to a place that such technologies can be successfully applied. Identifying which parameter to adjust in the optimization is the key.
7. Models should always be both calibrated and validated before use.
8. Model calibration is an iterative process, involving iterations between model application, additional data collection, and

Table 1. Top Three Teams' Rankings

<table>
<thead>
<tr>
<th>Team</th>
<th>Demands</th>
<th>Pressures</th>
<th>Tanks</th>
<th>Roughness</th>
<th>Valves</th>
<th>Fire flows</th>
<th>Total score</th>
<th>Total rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alvisi and Franchini</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Wu and Walski</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Koppel and Vassiljev</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>27</td>
<td>3</td>
</tr>
</tbody>
</table>

Fig. 9. Roughness total SSRE (truncated above 50)

Fig. 10. Valve total SSRE (truncated above 2)

Fig. 11. Fire-flow total SSRE (truncated above 0.1)
The calibration process typically alters system demands, fine-tunes pipe roughnesses, and modifies pump operation characteristics until satisfactory matching is attained between measured and modeled data. Once such a solution is received, how can one tell that the system is really calibrated (i.e., perhaps the entire process was simply a “curve fitting” exercise), or in other words, how can one discriminate between two calibration solutions that resulted in the same matching value? A possible way to address this issue is to extend the model’s calibration matrix from data matching to, for example, the model’s ability to successfully predict the resultant pressure and flows associated with an independently applied demand pattern and operating conditions, and to effectively predict the resultant pressure and flows associated with random abnormal/failure scenarios. These two latter criteria were posed to the participants as possible assessment avenues for C-Town but were eventually not incorporated.

2. Uncertainty inclusion: Calibration models usually treat the unknown calibration parameters (i.e., state parameters, such as demands, and space parameters, such as roughnesses) as deterministic. Inclusion of uncertainty in the model parameters and exploring the influence on the calibrated model outputs is important as it enables estimating model prediction errors. The assessment of both model parameter and prediction uncertainties can be accomplished using some analytical (e.g., first-order second moment) or sampling-based (e.g., Monte Carlo simulation) methods, either as part of the postoptimization-based calibration procedure (Kapelan 2010) or as part of the optimization process itself (Kapel et al. 2007).

3. Calibration size problem reduction: This point can be perceived in different ways, such as by means of aggregation methodologies and preliminary topological analysis. Like aggregation methods for water distribution system operation (Ulanicki et al. 1996) and water quality (Perelman and Ostfeld 2008), similar methodologies for water distribution system hydraulic calibration are required for reducing the calibration size problems, which, in turn, eases their application and helps avoid model overfitting. This challenge has two dimensions: clustering of calibration data and construction of an equivalent reduced-size system for performing the calibration tasks. To accomplish meaningful calibration (i.e., determine pipe roughness coefficients and valve states) and avoid unnecessary simulations for large water distribution systems, a study of topological observability is considered to be necessary (Ozawa 1987; Carpenter and Cohen 1993; Giustolisi and Berardi 2011). In particular, a preliminary analysis of network observability (i.e., pipes, nodes, and segments), according to the available measurements, can significantly reduce the search space (roughness and valve states) while reducing the network simulation problem and, consequently, the required computational time. This is even more important when the calibration problem is solved by using multiobjective optimization that requires multiple network simulation runs for several candidate solutions.

4. Leakage data: Incorporation of leakage may be included in hydraulic calibration efforts because leakage directly affects nodal demand allocation and pump curve characterizations.

5. Calibration data value: The effect of different field data on model calibration should be investigated; for example, pipe flow data would be more valuable for demand prediction than pressure measurements, whereas pressures provide better information for pipe roughness estimation. The effect of instrumentation type, number, and location would be significant.
Conclusions

This paper provides a summary of the BWCN, the goal of which was to objectively compare the solutions derived using different methods and approaches to the problem of calibration of water distribution systems. Participants were asked to apply a calibration methodology to a real water distribution system. Fourteen teams from academia, utilities, and private consultants participated. The BWCN outcomes were assessed at the 12th Water Distribution Systems Analysis (WDSA 2010) conference in Tucson, Arizona, in September 2010.

Calibration results were assessed with respect to the state variables of demands, pressures, and tank water levels, and for system parameters such as roughness coefficients. Teams’ performances were ranked and equally weighted with regard to the evaluated calibration categories. The BWCN concluded with announcing the top (best) three teams’ rank outcomes.

The success of the teams is attributed to the experience and intuition of the modelers and to the adopted methodologies. As the methodologies were tested on only one case study, it is difficult to identify/recommend a particular approach over another.

As a result of the BWCN, teams’ observations and future research directions were identified. These directions are expected to serve as guidelines and benchmarks for future applications and research developments in water distribution system calibration.

Supplemental Data

The supplemental material data include the following files, which are available online in the ASCE Library (www.ascelibrary.org):
- C-Town.inp—EPANET input file (version 2.00.12) distributed to participants;
- C-Town.xls—data file distributed to participants;
- C-Town_True_Network.inp—EPANET (version 2.00.12) calibrated (true) network; and
- C-Town_True_Data.xls—calibrated (true) data.

In addition, the following changes were made to the “real” system:

1. INP file
   1.1. Set pipe’s roughness to 100
   1.2. Set duration to 0 and hydraulic and reporting time steps to 1
   1.3. Removed all patterns
   1.4. Removed unused curves
   1.5. Set status of all pumps to “OPEN” and opened valve V2.
   1.6. “Raised” the pump curves according to the affinity laws
      1.6.1. Curve 8 (DMA1)—10%
      1.6.2. Curves 10 and 11 (DMA2 and 5)—5%
      1.6.3. Curve 9 (DMA3 and 5)—2%
   1.7. Control rule changes
      1.7.1. “ON” level for pump PU2 set to 2.5 (was 1)
      1.7.2. “ON” level for pump PU8 set to 2.5 (was 1.5)
   1.8. Set all demands to 0

2. Excel file (SCADA data)
   2.1. Set two data blocks to zero (marked in yellow)
   2.2. Set one data block with fixed value (marked in green)

References


