Calibrating Water Distribution Model Via Genetic Algorithms

The paper is presented at 2002 AWWA IMTech Conference, April 14-17, Kansas City, Missouri, USA

Calibrating Water Distribution Model Via Genetic Algorithms

By Zheng Y. Wu, Thomas Walski, Robert Mankowski, Gregg Herrin, Robert Gurrieri, Michael Tryby

Haestad Methods, Inc., 37 Brookside Rd, Waterbury, CT06708, USA
zwu@haestad.com, Tel +1-203-755-1666, Fax +1-203-597-1488

ABSTRACT

Computer models have been built for the simulation of water distribution systems since the mid-1960s. However, a model needs to be calibrated before it can be used for analysis and operational study of a real system. Model calibration is a vitally important, but time consuming task. Over last two decades, several approaches using optimization techniques have been proposed for model calibrations. Although most of the methods can make the model agree with field observations, few are able to achieve a good level of calibration in terms of determining the correct model parameters (pipe roughness coefficients, junction demands and valve settings). The previously developed methods appear to be lacking versatility for users to accurately specify calibration task given real data for a real system.

This paper proposes a comprehensive and flexible framework for calibrating hydraulic network model. Calibration tasks can be specified for a water distribution system according to data availability and model application requirements. It allows a user to (1) flexibly choose any combination of the model parameters such as pipe roughness, junction demand and link (pipes, valves and pumps) operational status, (2) easily aggregate model parameters to reduce the problem dimension for expeditious calculation, and (3) consistently specify boundary conditions and junction demand loadings that are corresponding to field data collection. A model calibration is then defined as an implicit nonlinear optimization problem, which is solved by employing a powerful genetic algorithm (GA), a generic search paradigm based on the principles of natural evolution and biological reproduction. Calibration solutions are obtained by minimizing the discrepancy between the model predicted and the field observed values of junction pressures and pipe flows. With this methodology, a modeler can be fully assisted during a calibration process, thus it is possible to achieve a good model calibration with high level of confidence. As a result, calibrated models can be developed for conducting system analysis and operational management. Example application is presented to demonstrate the efficacy and robustness of the genetic-based methodology for calibrating water distribution model.

INTRODUCTION

Computer models have become an essential tool for the management of water distribution systems around the world. There are numerous purposes for use of a computer model to simulate the flow conditions within a system. A model can be employed to ensure the adequate quantity and quality service of the potable water resource to the community, to evaluate the planning and design alternatives, to assess the system performance and to verify a operating strategy for better management of the water infrastructure system, as well as to be able to perform vulnerability studies to assess risks that may be presented and affect the water supply. A model is constructed for these purposes in which data describing network elements of pipes, junctions, valves, pumps, tanks and reservoirs are assembled in a systematic manner to predict pipe flow and junction hydraulic grade lines (HGL) or pressures within a water distribution system.

Computer models that have been established over last twenty years and that are to be constructed in future are significant investments for water companies. To ensure a good investment return or correct usages of the models, the model must be capable of correctly simulating flow conditions encountered at the site. This is achieved by calibrating the models. A calibration involves the process of adjusting model characteristics and parameters so that the model predicted flows and pressures match actual observed field data to some desirable or acceptable level and is described in more detail in Walski, Chase and Savic (2001).

Calibration of a water distribution model is a complicated task. There are many *uncertain parameters* that need to be adjusted to reduce the discrepancy between the model predictions and field observations of junction HGL and pipe discharges. Pipe roughness coefficients are often considered for calibration. However, there are many other parameters that are uncertain and affect junction HGL and pipe flow rate. To minimize errors in model parameters and also eliminate the compensation error of calibration parameters (Walski 2001), all the model parameters such as junction demand and operation status of pipes and valves, along with pipe roughness coefficients, should be considered for calibration.

Calibrating water distribution network models relies upon field measurement data such as junction pressures, pipe flows, and water levels in storage facilities, valve settings and pump operating status (on/off) and speeds. Among all the possible field observation data, junction HGL and pipe flows are often used to evaluate goodness-of-fit of the model calibration. The other parameters of tank levels, valve settings and pump operating status/speed are used as boundary conditions that are recorded when collecting a set of calibration observation of junction pressures and pipe flow rates.

Field observation data are measured and collected at different time of day and also various locations on site, which may correspond to different demand loadings and boundary conditions. In order for that the model simulation results more closely represent the observed data, the simulation results must be resulted from the same demand loading and boundary conditions as the observed data is collected. Thus the calibration process must be conducted under multiple demand loading and operating boundary conditions.

Traditional model calibration of a water distribution model is based on a trial and error procedure, by which an engineer or modeler first estimates the values of model parameters, then runs the model to obtain a predicted pressure and flow and finally compares the simulated values to the observed data. If the predicted does not compare closely with the observed data, the engineer returns to the model, makes some adjustments to the model parameters, and runs it again to produce a new set of simulation results. This may have to be repeated many times to make sure that the model produces a close enough prediction of water distribution network in the real world. Thus the traditional calibration technique is, among other things, quite time consuming.

In addition, a typical network representation of a water network may include hundreds or thousands of links and nodes. Ideally, during a water distribution model calibration process, the roughness coefficient is adjusted for each link, and demand adjusted for each node. However, only a small percentage of representative sample measurements can be made available for the use of model calibration, due to the limited financial and labor resources for data collection. Therefore, It is of utmost importance to have a comprehensive methodology and efficient tool that can assist modeler and engineer to achieve a highly accurate model under practical conditions including various model parameters such as pipe roughness, junction demand and link status, and also multiple demand and boundary conditions. It is the objective of this paper to provide a modeler with a calibration methodology and demonstrate a software tool in which the calibration is automatically optimized. These calibration results can be refined and manually adjusted during the calibration process.

CALIBRATION METHODOLOGY

An optimization calibrator is developed for facilitating the calibration process of a water distribution model. The parameters are obtained by minimizing the discrepancy between the model predicted and the field observed values of junction pressures (hydraulic grades) and pipe flows for given boundary conditions such as tank levels, control valve setting and pump speeds. The optimized calibration is then defined as a nonlinear optimization problem with three different calibration objectives.

Calibration Objectives

The goodness-of-fit of model calibration is evaluated by the discrepancy between the model simulated and field measured junction HGL and pipe flow. The goodness-of-fit score is calculated by using a user-specified fitness point per hydraulic head for junctions and fitness point per flow for pipes respectively. This allows a modeler to flexibly weight the evaluation of both pipe flow and junction hydraulic head. Three fitness functions are defined as follows.

<u>Objective Type One: Minimize the sum of difference squares</u>

$$minimize \qquad \frac{\sum_{np=1}^{NH} w_{nh}\left(\frac{Hsim_{nh} - Hobs_{nh}}{Hpnt}\right)^2 + \sum_{nf=1}^{NF} w_{nf}\left(\frac{Fsim_{nf} - Fobs_{nf}}{Fpnt}\right)^2}{NH + NF} \qquad (1)$$

<u>Objective Type Two: Minimize the sum of absolute differences</u>

$$minimize \qquad \frac{\sum_{np=1}^{NH} w_{nh}\left|\frac{Hsim_{nh} - Hobs_{nh}}{Hpnt}\right| + \sum_{nf=1}^{NF} w_{nf}\left|\frac{Fsim_{nf} - Fobs_{nf}}{Fpnt}\right|}{NH + NF} \qquad (2)$$

<u>Objective Type Three: Minimize the maximum absolute difference</u>

$$minimize \qquad \max\left\{\max_{nh=1}^{NH} w_{nh}\left|\frac{Hsim_{nh} - Hobs_{nh}}{Hpnt}\right|, \max_{nf=1}^{NF} w_{nf}\left|\frac{Fsim_{nf} - Fobs_{nf}}{Fpnt}\right|\right\} \qquad (3)$$

where $Hobs_{nh}$ designates the nh-th observed hydraulic grade, $Hsim_{nh}$ is the nh-th model simulated hydraulic grade, $Hloss_{nh}$ is the head loss at observation data point nh, $Fobs_{nf}$ is the observed flow, $Fsim_{nf}$ is model simulated flow, $Hpnt$ notes the hydraulic head per fitness point while $Fpnt$ is the flow per fitness point, NH is the number of observed hydraulic grades and NF is the number of observed pipe discharges, W_{nh} and W_{nf} represent a normalized weighting factor for observed hydraulic grades and flows respectively. They are given as:

$$W_{nh} = \mathrm{f}(Hloss_{nh} / \sum Hloss_{nh}) \qquad (4)$$

$$W_{nf} = \mathrm{f}(Fobs_{nf} / \sum Fobs_{nf}) \qquad (5)$$

where f( ) is a function which can be linear, square, square root, log or constant. An optimized calibration can be conducted by selecting one of three objectives above and the weighting factors between head and flow. The model parameters are calculated by using a genetic algorithm while minimizing the selected objective function.

Genetic Algorithm Optimization

Genetic algorithm (GA) is a robust search paradigm based on the principles of natural evolution and biological reproduction (Goldberg, 1989). For optimizing calibration of a water distribution model, a genetic algorithm program first generates a population of trial solutions of the model parameters. A hydraulic network solver program then simulates each trial solution. The resulting hydraulic simulation predicts the HGL (junction pressures) and pipe flows at a predetermined number of nodes (or data points) in the network. This information is then passed back to the associated calibration module. The calibration module evaluates how closely the model simulation is to the observed data, the calibration evaluation computes a "goodness-of-fit" value, which is the discrepancy between the observed data and the model predicted pipe flows and junction pressures or

HGL, for each solution. This goodness-of-fit value is then assigned as the "fitness" for that solution in the genetic algorithm.

One generation produced by the genetic algorithm is then complete. The fitness measure is taken into account when performing the next generation of the genetic algorithm operations. To find the optimal calibration solutions, fitter solutions will be selected by mimicking Darwin's natural selection principal of "survival of the fittest". The selected solutions are used to reproduce a next generation of calibration solutions by performing genetic operations. Over many generations, the solutions evolve, and the optimal or near optimal solutions ultimately emerge. Many successful applications of GA to solving model calibration have been carried out for optimized calibration of water resource systems (Wang 1991; Wu 1994; Babovic etc. 1994; Wu and Larsen 1996). Over last decade, there are numerous variations of genetic algorithms. A competent genetic algorithm (also called fast messy GA by Goldberg et al. 1989 and 1993), which has been demonstrated the most efficient GA for the optimization of a water distribution system (Wu & Simpson 2001), has been used for the optimized calibration.

Integrated Calibration Capability

GA-based Darwin Calibrator is integrated into a modeling system WaterCAD. It offers a modeler with the maximum flexibility and the richest functionality. With Darwin Calibrator as a tool, a modeler is able to calibrate a water distribution model under practical conditions including the combination and aggregation of the model parameters, multiple demand loading conditions, various boundary system conditions, manual adjustment and sensitivity analysis of calibration solutions.

Due to the large number of pipes and junctions in a model, pipes that have the same physical and hydraulic characteristics are allowed to be grouped as one calibration link, which one new roughness coefficient or one roughness coefficient multiplier is assigned to all the pipes in the same group. The junctions that have the same demand patterns and within a same topological area can also be aggregated as one calibration junction, to which a same demand multiplier is calculated and assigned. Calibration parameters are bounded by prescribed upper and lower limits and adjusted with a user-prescribed incremental value. For examples, A Hazen William C value for a pipe or a group of pipes will be computed within a range of a minimum of 40 and maximum of 140 with an incremental of 5. Demand multiplier may range from 0.8 to 1.2 and increases by 0.1. Parameter aggregation is useful at reducing the calibration dimension, however, caution needs to be excised at pipe and junction grouping, which may affect the accuracy of the model calibration.

Darwin Calibrator allows a modeler to select any combination of three types of model parameters pipe roughness (based on HW, DW and CM head loss equations), junction demand and link (pipe and/or vales) status. Roughness values can be computed as a new value for a group of pipes or modified by a multiplication factor. Junction demand multipliers are adjusted for the spatial and temporary demand variation. Link status including valve and pipe are treated as binary variable that take a value of 1(OPEN) or 0 (CLOSE).

A good model calibration must be performed for multiple demand loading conditions to better simulate the demand variation when the observed data are collected and thus improve the accuracy of model calibration. Baseline demands for all the junctions are changed for the entire system to simulate a loading condition at a specific time. A junction demand can also be updated individually to calibrate the model for an extreme loading condition, e.g. fire flows. Boundary conditions, such as tank levels, pump speeds and valve settings, can be specified for each of demand loading conditions. This ensures that the calibration is performed to compare the observed values with simulated values at the same system conditions.

During the model calibration process, there exist a number of solutions that produce the same or very close goodness-of-fit due to the greater number of calibration parameters than the independent observed data set. A number of top solutions is kept and reported at the end of a calibration run. The top solutions can be used for many purposes such as sensitivity study and different modeling scenarios to verify the final calibration result, along with sound engineering judgments.

The sensitivity and verification studies can be easily conducted by using the manual calibration feature. The manual calibration feature allows that the calibration model parameters are to be set manually to the desirable values and the calibration run is performed without running GA optimization to evaluate how close the model simulation is to the field observation values.

APPLICATION EXAMPLE

Use of the calibrator can be illustrated using a simple system that captures many of the features of a real distribution system. The system is shown in Figure 1 below and is made up of old unlined cast iron pipes and new cement mortar lined ductile iron pipes. Water use at nodes can be described as residential or commercial depending on the type of customers and their diurnal water demand pattern.

For this example, a set of pressure measurement is made throughout the system and fire flow tests are conducted at two nodes (J-10 and J-31). Flow is only measured at the main pump station. C-factors for the cast iron pipes are initially estimated as 90 and the ductile iron pipes as 130.

Fitness is calculated using the minimum square error formulation above and the weighting given to head and flow measurements is 1 and 10 respectively. Initially, the model fitness is calculated as 20.1 using the initial estimates of C and demand. In general the simulated HGL values are too low during high flows indicating that C factors should be decreased. (High fitness means high discrepancies between measured and modeled values. Low fitness values are the goal.)

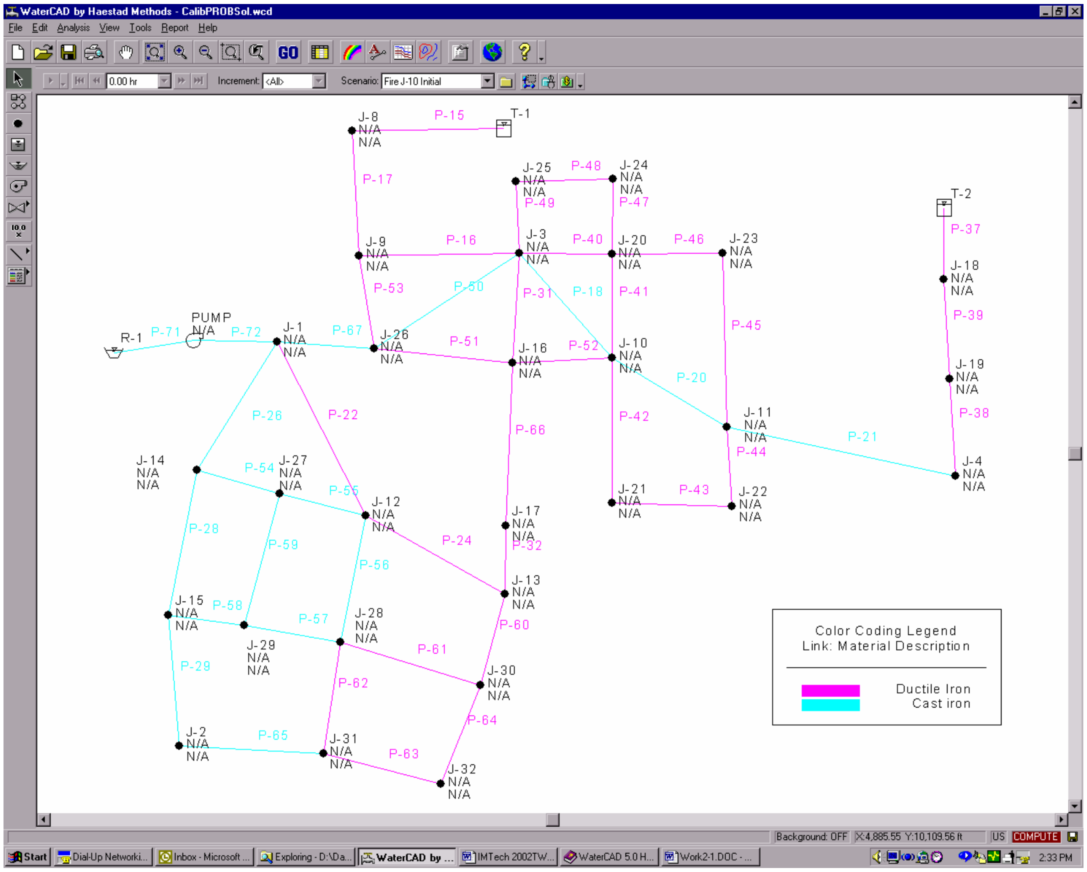

Figure 1. System Used in Example Application

In a manual calibration, the user is not sure what to adjust and by how much, so the C-factor is cut in half. This overestimates roughness and head loss and the fitness values increase to 476. The model produces graphs such as Figure 2 to show the relationship between observed and modeled heads. Obviously these C-factors are too low.

The user then utilizes the Darwin calibrator to find the optimal solution by adjusting C-factors. While this attempt reduces the Fitness function value to 4.36 but it is done by lowering the C-factor in the ductile iron pipe to 91 and raising the C-factor in the cast iron to 135. While the solution in terms of HGLs looks good, the results in terms of pipe roughness are not logical. Apparently, the assumption that all the error is due to C-factor is incorrect and something besides C-factor needs to be adjusted. (Some models only allow C-factor to be adjusted but Darwin is more flexible in allowing other parameters such as demand and pipe status to be unknowns.)

The user now allows Darwin to adjust demand multipliers for the commercial and residential nodes in addition to pipe roughness. In this case, the model arrives at a solution with a fitness of 0.10 by setting the C-factors to 130 and 54 for the ductile and cast iron pipes and using demand multipliers of 1.5 and 1.2 for the commercial and residential nodes. These results agreed exactly with the roughness and demands used to generate the "field data" used in the model. Figure 3 shows the agreement between model and observed HGL values for the correct solution.

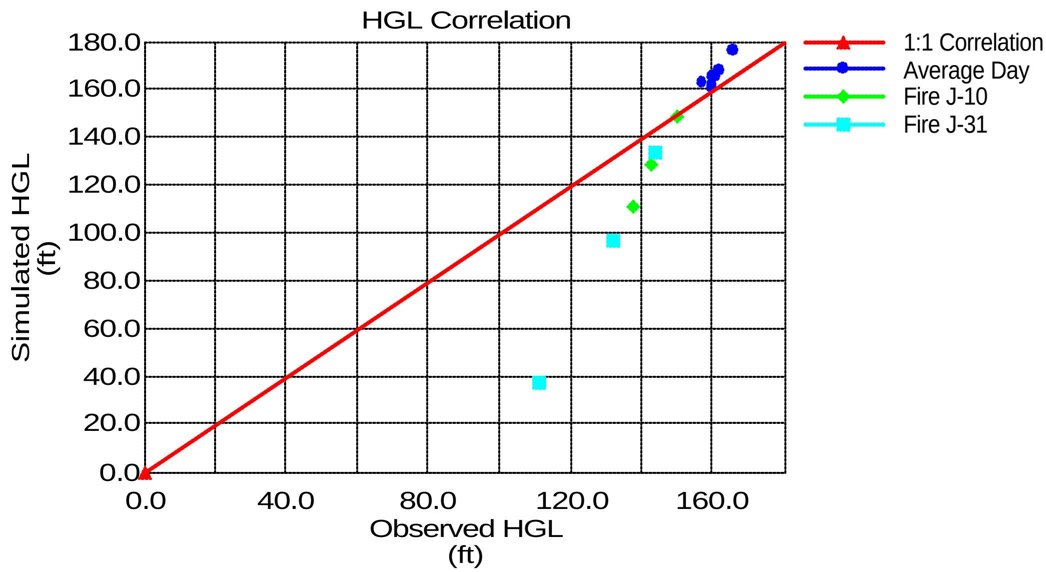

Figure 2. Error Plot for Initial Manual Trial

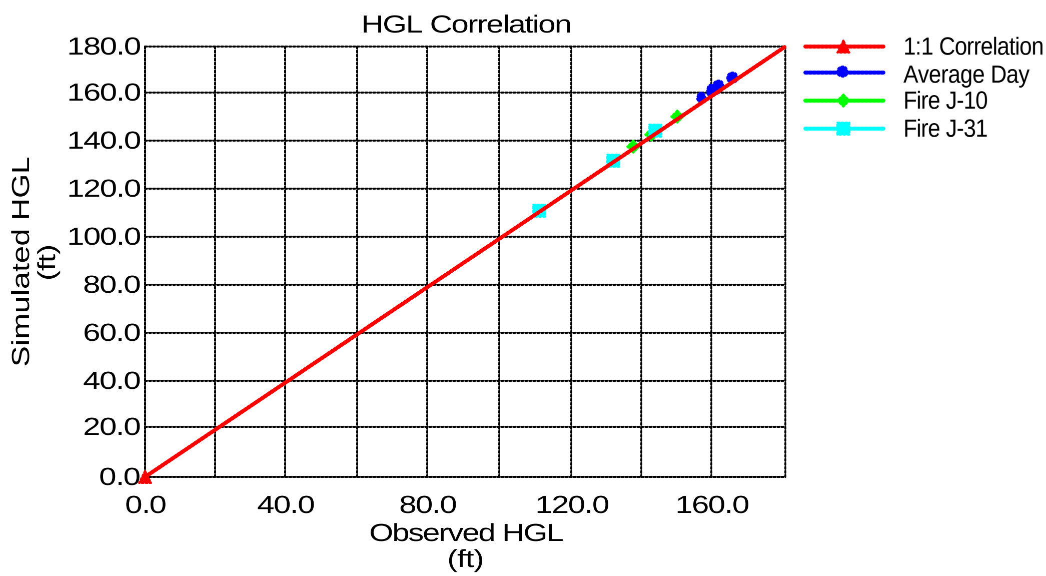

Figure 3. Error Plot for Correct Solution

The results discussed above were for the case where the HGL values were known to the nearest foot. In real systems errors of 2 to 5 ft are not uncommon. Walski (2000) demonstrated that when errors in measurement are of the same magnitude as head loss, accurate calibration may be impossible. So a run was made with the Darwin calibrator for the case where no fire flow data were available and Darwin found a solution with a fitness of 5.58. However, this solution was obtained by lowering the C-factor of the ductile iron pipe to an unrealistic value of 91. This illustrates the fact that optimized calibration can do a good job of matching HGL values in a system, but unless the data are of good quality, may not do as well at estimating pipe roughness or demands.

CONCLUSIONS

The proposed approach offers a powerful tool for water distribution model calibration process. Multiple parameters and corresponding boundary conditions are taken into account to provide an accurate representation of the network. The integrated system includes a software program that contains three integral parts: a genetic algorithm module, a hydraulic simulation module and a calibration module. These modules interact to provide an optimized calibration solution.

More specifically, the method of automatically calibrating a water distribution model gives a modeler a maximum flexibility at setting up a calibration under practical conditions. For example, a modeler can choose model parameters including the pipe roughness coefficient, junction demand, and pipe and valve operational status, or any combination of the parameters. Next, the user enters field observed data, namely amounts for pressure and pipe flow. The observed data can be weighted with a user-selected weighting function to focus the calibration on certain data points. Multiple demand and boundary conditions can also be simulated simultaneously for a calibration run. The demand loading condition can be modified globally for entire system for different times of day or updated for individual junctions such as fire flow testing corresponding to the time when the observed data is collected. Boundary conditions of storage tank levels, pressure control valve settings and pump operation speeds can also be taken into account. This improves the accuracy by providing a realistic snapshot of the network actually operating at each instant in time.

Finally the efficient genetic algorithm drives the search process for locating the optimal and a number of near-optimal model parameter solutions. The integrated calibration system includes optimized calibration and also manual calibration, which essentially enables the calibration task to be accomplished in a speedy manner, thus improves the productivity of the modeling process.

REFERENCES

Babovic V., Wu Z. Y. & Larsen L. C. (1994), Calibrating Hydrodynamic Models by Means of Simulated Evolution, in *Proceeding of Hydroinformatics'94*, Delft, the Netherlands, pp193-200.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, Reading, MA.

Goldberg, D. E., Korb, B., & Deb, K. (1989). "Messy genetic algorithms: Motivation, analysis, and first results," *Complex Systems, 3*, 493-530.

Goldberg, D. E., Deb, K., Kargupta, H., & Harik G. (1993). "Rapid, Accurate Optimization of Difficult Problems Using Fast Messy Genetic Algorithms," *IlliGAL Report No. 93004*, Illinois Genetic Algorithms Laboratory, University of Illinios at Urbana-Champaign, Urbana, IL 61801, USA.

Walski, T.M. (2000) "Model Calibration Data: The Good, The Bad and The Useless," *J AWWA*, 92(1), p. 94.

Walski, T. M. (2001) "Understanding the adjustments for water distribution system model calibration." *Journal of Indian Water Works Association*, April-June, 2001, pp151-157.

Walski, T.M., Chase, D.V. and Savic, D.A. (2001) *Water Distribution Modeling*, Haestad Methods, Inc.

Wang Q.J. (1991), The Genetic Algorithm and its Application to Conceptual Rainfall-Runoff Models, *Water Resources Research*, Vol.27, No.9, pp2467-2482.

Wu Z.Y. (1994), *Automatic Model Calibration by Simulating Evolution*, M.Sc. Thesis, H.H. 191, International Institute for Infrastructure, Hydraulic and Environmental Engineering, Delft, the Netherlands.

Wu Z.Y. and Larsen C.L. (1996). "Verification of hydrological and hydrodynamic models calibrated by genetic algorithms." *Proc. of the 2nd International Conference on Water Resources & Environmental Research*, Vol. 2, Kyoto, Japan, pp175-182.

Wu, Z. Y. and Simpson A. R. (2001) "Competent Genetic Algorithm Optimization of Water Distribution Systems." *J. of Computing in Civil Engineering*, ASCE, Vol 15, No. 2, pp89-101.