Analysis of Model Sensitivity and Uncertainty for Chlorine Transport and Decay in a Water Distribution System

W. J. Dawsey[1], B. S. Minsker[2], and A. Ostfeld[3]

[1]Illinois State Water Survey, Institute of Natural Resources Sustainability, University of Illinois; 2204 Griffith Drive, Champaign, IL 61820-7495; Phone: 217-244-8856; dawsey@illinois.edu.
[2]Department of Civil and Environmental Engineering & National Center for Supercomputing Applications, University of Illinois; 3230d Newmark Lab, MC-250 205 N. Mathews Ave., Urbana, IL 61801; Phone: 217-265-5293; minsker@illinois.edu.
[3]Faculty of Civil and Environmental Engineering, Technion–Israel Institute of Technology, Haifa 32000, Israel; Phone: +972-4-8292782; ostfeld@tx.technion.ac.il.

ABSTRACT

There are a number of sources of uncertainty in drinking water distribution system modeling. Uncertain parameters include pipe diameters, consumer demands, hydraulic energy loss coefficients, reaction coefficients and others. Understanding the relative importance of these sources of uncertainty can improve the allocation of resources for model refinement and calibration, as well as, aid knowledge inference from monitoring data. This paper presents an analysis of uncertainty and model sensitivity for chlorine transport and decay in a water distribution system. A clustering and global variance-based sensitivity methodology is proposed to account for spatial inconsistencies found in the results of previous studies of this problem. Results are presented from small and large scale case studies.

This methodology is then used to explore the occurrence of intrusion events in a water distribution system, and the potential to detect such events through online monitoring of chlorine residual concentrations. Noise present in the chlorine monitoring signal has the potential to overwhelm the detection of an upstream intrusion and its associated chlorine demand. Results are presented from simulated intrusion events of varying magnitude and duration.

INTRODUCTION

There are a number of sources of uncertainty in drinking water distribution system modeling. These include, pipe diameters, consumer demands, hydraulic energy loss coefficients, solute reaction coefficients, and others. These parameters can be grouped into two categories. Aleatory uncertainty is inherent to stochastic processes such as rainfall in Hydrology or in this case, consumer demands on a drinking water system. Epistemic uncertainty occurs when processes are unknown or poorly represented by mathematical models. An example of epistemic uncertainty is the variability that occurs in bulk first-order decay coefficients for free within a water distribution system. In this case, a single species first-order decay model does not accurately represent interactions with other species and pipe walls. Thus, the analysis of uncertainty in any system is highly dependent on the model form and sophistication. Understanding of model uncertainty can be useful in allocating resources to model calibration and exploring novel approaches to modeling.

The objective of this paper is to better understand the sources of uncertainty in water distribution system modeling so that better modeling and monitoring methodologies can be developed. One potential monitoring approach is to use chlorine residual as an indicator of intrusion events. Through the introduction of chlorine demanding species, intrusion events can change the residual chlorine profile in a pipe network. If such a change were detected by a monitoring system, then appropriate response actions could be taken by water system operators to protect public health and restore network integrity. However, variability in network hydraulics and chlorine decay could obscure the detection of such a chlorine demand event. Previous studies have not fully characterized the relative importance of the various sources of uncertainty in this modeling problem.

A methodology is proposed for evaluating the relative effects of uncertain inputs on residual chlorine concentration at a monitoring node. Nodes are divided into upstream and downstream observational pairs that are then clustered into experimental units by mean travel time and node location. These experimental units define spatially and hydraulically similar regions of the network. Latin hypercube sampling and simulation of network hydraulics and chlorine transport using EPANET are proposed for both the undisturbed model condition and a chlorine demand event condition. Simulations are performed for a set of randomly sampled node pairs within each experimental unit. Independent uncertain inputs include nodal demand, pipe diameter, roughness, bulk chlorine reaction coefficient, and wall reaction coefficient. The first step of the sensitivity analysis is to screen input parameters using the Morris method (Saltelli, 2004) to eliminate those that negligibly influence output uncertainty and identify potential interaction effects. The second step is to apply a variance-based sensitivity analysis to determine the sensitivity indices for each input parameter. First-order indices and total sensitivity indices are calculated for each parameter. Higher order indices representing interactions between multiple parameters are calculated as suggested by the results of the Morris screening test. Indices are determined for each experimental unit and aggregated to characterize the total system sensitivity. The chlorine demand event simulations will explore a range of different magnitude chlorine reductions and the effect on downstream monitoring nodes relative to other uncertainties.

BACKGROUND

Uncertainty in water distribution system modeling has for the most part been a secondary topic of study in the research literature. There are several studies that utilize one or two uncertain inputs in developing models of hydraulic reliability (Wagner et al. 1988a and 1988b), optimal design (Lansey et al, 1989), solute fate and transport (Khamal, 2006), optimal monitoring (Jankovic et al., 2007), and to evaluate sensitivity methodologies (Kang et al. 2007). These studies contain some information about the output response to uncertainty in the selected input(s). However, there are very few studies that attempt to quantify the relative importance of multiple uncertain inputs, and only one Pasha and Lansey (2005) attempt to do so for the case of fate and transport.

Table 1 presents a matrix of model characteristics and the relevant research literature. It is evident from this table that most of the studies have investigated uncertainty in hydraulic input parameters including nodal demand and pipe roughness coefficient. Simulation techniques such as Monte Carlo sampling and analytical approaches such as first order second moment (FOSM) are equally represented in the literature. There are a limited number of papers that have investigated uncertainty in water distribution systems water quality.

Wagner et al (1988a) and Wagner et al. (1988b) present a simulation and an analytical methodology, respectively, for modeling distribution system reliability subject to randomized mechanical failures of network components. Lansey et al. (1989) proposed a chance constrained model to minimize design cost taking into account uncertainties in pipe roughness, demand, and mechanical failure. Xu and Goulter (1999) also proposed a model to optimize design cost and reliability using the same uncertain inputs. Both of these papers utilized first order analytical methods for approximating model output response to input uncertainties. Neither paper makes assertions about the relative importance of input uncertainties.

Bao and Mays (1990) utilized Monte Carlo simulations to determine the effects of uncertainty in demand, pipe roughness, and pressure heads on reliability. Nodal reliability was defined as the probability that flow at a node will meet demand flows at a minimum pressure. The authors found that the choice of probability distribution for input parameters had a significant effect on nodal reliability, and that demand exerted a greater influence on outputs than did pipe roughness (Bao and Mays, 1990). Kapelan (2005) compared a sampling method with a deterministic integral approximation of the uncertain model response for the least-cost design problem. Demand was the only uncertain input for this study. Babayan (2005) proposed a cross entropy approach for multiobjective optimization of distribution system design with uncertain demand and roughness parameters. Babayan (2006) compared an integration based method with Latin hypercube sampling for solving the stochastic optimization design problem for water distribution systems. Again, demand and roughness were uncertain input parameters.

Kang et al. (2007) compared Monte Carlo and Latin hypercube sampling methods with a first order second moment analytical approach for estimating chlorine response to uncertain demand inputs. Sampling methods were found to be more accurate for unsteady flow conditions and the nonlinearities that occur due to flow reversals. First order methods were unable to approximate such nonlinear behavior. Demand was the only uncertain input parameter for this paper. Branisavljevic and Ivetic (2006) proposed a fuzzy approach for uncertainty analysis in a water distribution system. This method does not require the specification of an exact input probability distribution, but rather uncertainty is represented by fuzzy set mapping of input parameters. Outputs were found to be sensitive to the interval over which the fuzzy input set was specified. Pipe roughness was the only uncertain input parameter, and pressure head was the output.

Jankovic et al. (2007) describe a stochastic model for estimating network flows and heads from uncertain demand inputs in order to identify locations for leakage monitoring. Monte Carlo simulations were used with EPANET to determine output response. Kapelan et al. (2007) proposed a novel analytical method for mapping uncertain input parameters to a set of output calibration data. The probability distributions for a set of input parameters were approximated using Bayesian recursive algorithm. The only uncertain inputs considered by Kapelan et al. (2007) were roughness coefficients.

Two papers in the water distribution system research literature have considered uncertainty analysis for water quality parameters. Khamal et al. (2006) modeled the injection and transport of a conservative solute and introduced an exposure index to quantify the downstream effects for consumers. A generalized sensitivity analysis was performed to determine system response to uncertainty in base demand, storage capacity, injection mass and injection duration. Base demand and injection mass were identified as the input parameters with the greatest influence on exposure index. Latin hypercube sampling coupled to EPANET simulation were used to evaluate system response. Lansey et al. (2005) modeled decay and transport in a water distribution system subject to uncertain demands, bulk decay coefficient, wall decay coefficient, pipe roughness, and pipe diameter. Monte Carlo sampling and EPANET simulation were used to determine the output pressure and chlorine concentration at each node. Results shown at three monitoring nodes suggest that wall decay coefficient, bulk decay coefficient, demand and pipe roughness all exhibit a large influence on output values, with each input taking precedence at different monitoring locations.

The following system of equations is utilized by the EPANET model for hydraulics and water quality in water distribution systems. Additional details may be found in Rossman (2002). Continuity and conservation of energy in a pipe network require that,

$$\sum_j Q_{ij} = D_i \tag{1}$$

and,

$$H_i - H_j = h_{ij} \tag{2}$$

where,
Q_{ij} = pipe flow between nodes i and j,
D_i = demand at node i,
H_i and H_j = total head at nodes i and j respectively, and
h_{ij} = head loss in the pipe between nodes i and j.

The head loss may be estimated using the Hazen-Williams equation (US customary units),

$$h_{ij} = K\left(\frac{Q_{ij}}{C}\right)^{1.852}\frac{L_{ij}}{D_{ij}^{4.87}} \tag{3}$$

where,
K = constant,
C = Hazen-Williams roughness coefficient,
L_{ij} = length of pipe between i and j, and
D_{ij} = diameter of pipe between i and j.

Table 1. Matrix of water distribution system uncertainty analysis research literature

		Wagner et al., 1988a	Wagner et al., 1988b	Lansey et al., 1989	Bao and Mays, 1990	Xu and Goulter, 1999	Kapelan et al., 2003	Barkdoll and Didigam, 2004	Babayan et al., 2004	Babayan et al., 2005	Pasha and Lansey, 2005	Babayan et al., 2006	Branisavljevic and Ivetic, 2006	Khamal et al.,2006	Kang et al.,2007	Kapelan et al., 2007	Jankovic et al., 2007
Solute	Chlorine							X			X				X		
	Conservative													X			
	None	X	X	X	X	X	X		X								
	Other																
Uncertain Input Hydraulic Parameters	Nodal demand			X	X	X	X	X	X	X	X	X		X	X		X
	Roughness			X	X	X		X		X	X	X	X			X	
	Pipe diameter							X			X						
	Solute loading condition													X			
	Storage capacity													X			
	Mechanical failure	X	X	X		X											
Uncertain Input WQ Parameters	Wall react. coef.										X						
	Bulk react. coef.							X			X						
	Wall Mass transfer coef.																
	Loading condition																
Input Spatial independence	Full	X	X		X	X	X				X				X		
	Some													X			
	None																
Methodology	Simulation	X			X		X	X	X	X	X	X		X	X	X	X
	Analytical			X		X	X	X	X	X		X			X		
	Other		X							X			X				
Output	Solute Concentration							X			X			X	X		
	Pressure	X	X		X	X		X	X	X	X	X	X		X	X	X
	Reliability	X	X	X	X	X											
	Cost					X	X			X		X					

Solute fate and transport in the distribution system model can be represented by the following set of relationships for a first order decay model with bulk and wall reaction components (Rossman, 1994).

$$\frac{\partial C_i}{\partial t} = -u_i \frac{\partial C_i}{\partial x} + r(C_i) \tag{4}$$

where,
C = concentration
u = average flow velocity
x = longitudinal direction of pipe
t = time

i = pipe index
$r\,(C_i)$ = reaction rate ($M \cdot L^{-3} \cdot T^{-1}$) as a function of concentration, sum of wall and bulk reactions

$$r_{wall} = -\frac{k_w k_f C_i}{R_h\left(k_w + k_f\right)} \tag{5}$$

where,
k_w = wall decay constant ($L \cdot T^{-1}$)
k_f = fluid mass transfer coefficient ($L \cdot T^{-1}$)
R_h = hydraulic radius, pipe volume/area, or 2/radius

$$r_{bulk} = -\,k_b C_i \tag{6}$$

where,
k_b = bulk decay constant ($L \cdot T^{-1}$)

METHODOLOGY

The approach presented here proposes to accomplish two objectives: 1. extend the results from Lansey et al. (2005) to better understand model sensitivity for the fate and transport problem , and 2. model the occurrence of a chlorine demand event and system response within the context of other sources of uncertainty, a problem that has not been addressed previously in the research literature.

Sensitivity analysis (SA) and uncertainty analysis (UA) are terms that are often confused in the research literature. UA is a forward process to quantify the effect of uncertain inputs on output uncertainty, while SA works backward to ascribe uncertainty in the output to the different sources of uncertainties among the inputs. SA attempts to identify the hierarchy of influence that inputs exert on output variability.

There are a large number of sensitivity analysis methodologies (Saltelli, 2000) that have been applied to varying degrees in the Environmental Engineering domain. Variance-based approaches have been shown to be robust for non-monotonic and non-linear problems such as distribution system modeling (Saltelli, 2000). Storage tanks can create pipe flow reversals in nearby regions of a system, thereby causing abrupt nonlinear changes in the chemical profile of water at affected nodes. Variance-based sensitivity analysis is global in that it characterizes model sensitivity over the entire input parameter space.

Lansey et al. (2005) reported that the relative influence of uncertain input parameters varied with choice of monitoring node, and a different parameter was identified as the most important for the three cases shown. This result suggests that an additional spatially heterogeneous parameter may exist that would allow this source of variance to be separated and explained. In addition, Lansey et al. (2005) performed single fixed and single uncertain parameter simulations for each input, and made inferences about input parameter influence based on the standard deviation of chlorine concentration at monitoring nodes. Using the standard deviation is similar to variance-based methods, however, first or higher order sensitivity indices that are normalized by the output variance were not calculated.

For this study, input probability distributions are consistent with those from Lansey (2005). Pipe diameter will be sampled from a uniform distribution with an interval range of 0.5 - inches, and all other input parameters will be normally distributed with a coefficient of variation equal to 0.1 (Lansey, 2005). The initial case study will be example network 2 from Rossman (2002). This network has been used extensively in the water distribution system research literature for developing novel methodologies.

Spatial Clustering and Sampling

In order to efficiently characterize the spatial variability of the system, a randomized sample of chlorine injection node and downstream monitoring node pairs is proposed. In the language of experimental design, each node pair is considered to be an observational unit. These node pairs are then clustered using the following characteristics:

t_t , mean travel time between the pair of nodes,
(x_1, y_1), coordinate location of node 1,
(x_2,y_2), coordinate location of node 2, and
r_{ud}, upstream-downstream relationship between nodes.

Upstream-downstream relationship refers to the set of three possible flow states that may occur between nodes: 1. no relationship – flow does not pass between nodes, 2. one node is always downstream throughout the simulation period, and 3. upstream and downstream node switch caused by a reversal of flow between the nodes. A schematic is shown in Figure 1 to demonstrate this idea.

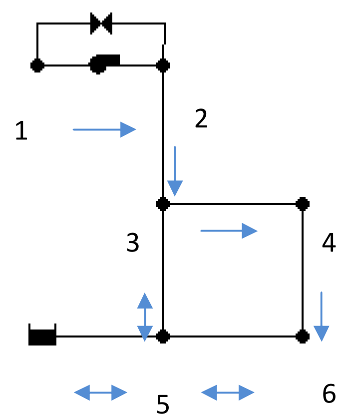

Figure 1. Example network schematic

The matrix shown in Table 2 demonstrates the full set of relationships between nodes in Figure 1. The reason for grouping observational pairs using this criteria is that pairs containing intermediate pathways with flow reversal may exhibit non-monotonic relationships between input and output parameters. Clusters of these six parameters define regions of the node pair space that are hydraulically and spatially similar. These regions define experimental units that are randomly sampled in order to efficiently characterize the effects of spatial heterogeneity on the results.

Table 2. Matrix of monitoring node pair relationships

	1	2	3	4	5	6
1	n	c	s	c	s	s
2	n	n	s	c	s	s
3	n	n	n	c	s	s
4	n	n	n	n	s	s
5	n	n	s	s	n	s
6	n	n	s	s	s	n

where,
n = no relationship,
c = consistent downstream node
s = switching upstream and downstream node

Morris Screening Method

Screening methods are used to reduce model complexity prior to running exhaustive simulations over uncertain parameters. The Morris method is a means of initially evaluating the influence of inputs to determine if they exert negligible influence, exhibit linearity, or combine to form interaction effects (Saltelli, 2004). The elementary effect for the *i*th input is defined to be:

$$d_i(x) = \frac{\left[y\left(x_1,...,x_{i-1},x_i+\Delta,x_{i+1},...,x_k\right) - y(x) \right]}{\Delta} \tag{7}$$

where

y = output
$x = (x_1, x_2,\ldots,x_k)$ set of values selected from set {0, 1/(p-1), 2/(p-1), …, 1},
Δ = a constant value equal to a multiple of 1/(p-1), and
p = the number of input space divisions

The set x is randomly sampled and evaluated until the distribution of elementary effects for input i is found (F_i). The mean and standard deviation of F_i are used to evaluate the relative importance of an input value. The absolute value of d_i may be used in the generation of this distribution for non-monotonic models such as water distribution systems. The mean of F_i is a measure of the input's importance, and the standard deviation indicates non-linearity and interaction effects. These measures are used to refine the model prior to more exhaustive variance-based methods.

Variance-based Sensitivity Analysis

Variance-based sensitivity analysis presumes that variance is a suitable measure of uncertainty in an output. The total variance of the output is attributed to different sources within the set of uncertain input parameters, similar to Analysis of

Variance (ANOVA) techniques in the experimental design literature. The reduction in output variance that occurs if a single input parameter is fixed to its 'true' value is a measure of that parameter's first-order influence. The variance of an output given a single fixed input is,

$$V(Y|P_x = p^t{}_x) \tag{8}$$

where P_x is an uncertain input parameter, Y is the output, $p^t{}_x$ is the true fixed value of the input, V is the variance taken over all input parameters $\neg x$. The true value $p^t{}_x$ is unknown, however, it can be approximated by taking the expected value over P_x:

$$E(V(Y|P_x) \tag{9}$$

The law of total variance states that:

$$V_t = E(V(Y|P_x) + V(E(Y|P_x) \tag{10}$$

where V_t is the total variance of Y over all uncertain parameters. Thus the reduction in the total variance caused by fixing a single input is given by

$$V(E(Y|P_x) = V_t - E(V(Y|P_x) \tag{11}$$

The first order sensitivity index is found by normalizing over the total variance:

$$S_x = \frac{V(E(Y \mid P_x))}{V_y} \tag{12}$$

For water distribution systems, S_x can be determined through Latin Hypercube simulations over all uncertain parameters save the single fixed input. Higher order sensitivity indices represent the interaction effects of multiple parameters. For orthogonal inputs, second order effects are:

$$S_{x1,x2} = \frac{V(E(Y \mid P_{x1}, P_{x2})) - V(E(Y \mid P_{x1})) - V(E(Y \mid P_{x2}))}{V_y} \tag{13}$$

Higher orders are similarly calculated. The total effect ($S_{T,x}$) can be calculated which represents the total contribution of P_x for all orders. A relatively complete picture of a parameter's influence is given by S_x and $S_{T,x}$ and neglecting intermediate orders of interaction (Saltelli, 2004).

Chlorine Demand Simulation

The results of the sensitivity analysis will be used to refine the model and reduce the number of uncertain inputs. The refined model will then be used to estimate the effect of a chlorine demand event on downstream monitoring nodes. Again, monitoring node pairs clustered into similar hydraulic and locational experimental units will be used to characterize the spatial heterogeneity of the model. Random sampling within these units replaces an exhaustive simulation over all nodes.

A chlorine reduction will be simulated at each sampled upstream node. The distribution of chlorine concentration at the downstream node for the chlorine demand scenario will be compared to that of a non-demand scenario at each time step. The statistical difference between these distributions will be estimated. Different magnitudes of chlorine reduction and demand event duration will be simulated in order to determine the severity of event that may be detected. As with the sensitivity analysis, results will be determined for each experimental unit and aggregated over the full system.

CONCLUSIONS AND NEXT STEPS

This paper presents a methodology to determine a water distribution system model response to uncertain inputs for the chlorine fate and transport problem. It also proposes to model the occurrence of chlorine demand events within the context of other sources of uncertainty. Results from this study will help better understand the potential for chlorine monitoring as an indicator of upstream intrusion events. Results from the methodology presented here also may explain the inconsistent influence that uncertain inputs exhibited on chlorine response shown by Lansey et al. (2005). A clustering and sampling approach for characterizing spatial heterogeneity in water distributions is proposed. The next steps for this study will be to implement conservative solute tracer simulations over the set of uncertain inputs to identify the monitoring pair classifications shown in Table 2 and the mean travel time between pairs. These will be used to cluster pairs into the appropriate experimental blocks.

REFERENCES

Babeyan, A.V., Kapelan, Z.S., Savic, D.A., and Walters, G.A.(2004). "Comparison of two approaches for the least cost design of water distribution systems under uncertain demands." *Proceedings of the ASCE World Water and Environmental Resources Congress, June 27-July 1, 2004,* Salt Lake City, Utah, USA.

Babayan, A.V. ,D. A. Savic, and G. A. Walters (2005). "Multiobjective Optimization for the Last-Cost Design of Water Distribution System Under Correlated Uncertain Parameters". *Proceedings of the ASCE World Water and Environmental Resources Congress 2005, Anchorage, Alaska, USA.*

Babayan, A.V. ,Z.S. Kapelan, D.A. Savic, and G.A. Walters (2006). " Comparison of two methods for the stochastic least cost design of water distribution systems". *Engineering Optimization*, 38(3), 281-297.

Barkdoll, B. D., and Didigam, H. (2004). "Effect of user demand on water quality and hydraulics of distribution systems." *Proceedings of the ASCE World Water and Environmental Resources Congress, June 27-July 1, 2004,* Salt Lake City, Utah, USA.

Bao, Y., and Mays,L.W. (1990)."Model for water distribution system reliability." *Journal of Hydraulic Engineering*, ASCE, 116(9),1119-1137.

Branisavljevic, N.,and M. Ivetic (2006). "Fuzzy approach in the uncertainty analysis of the water distribution network of Becej". *Civil Engineering and Environmental Systems*, 23(3),221-236.

Janković-Nišić,B., N. J. D. Graham, Č. Maksimović, and D. Butler, " Cost-effective leakage reduction through district metering". *Proceedings of the Institution of Civil Engineers: Water Management*, 160(3), 181 – 187.

Khanal, N., Steven G. Buchberger, S. G., and McKenna, S. A., (2006). “Distribution System Contamination Events: Exposure, Influence, and Sensitivity.” *Journal of Water Resources Planning and Management*, ASCE, 132(4), 283-292.

Kang,D.S., M.F.K Pasha, and K. Lansey (2007). "Approximate methods for analyzing water quality prediction uncertainty in water distribution systems". *Proceedings of the ASCE World Environmental and Water Resources Congress 2007, Tampa, Florida, USA.*

Lansey, K.E.,Duan, N., Mays,L.W., and Tung, Y.K. (1989)."Water distribution system design under uncertainties." *Journal of Water Resources Planning and Management*, ASCE, 115(5),630-645.

Pasha, M. F. K, and Lansey, K. E. (2005). “Analysis of uncertainty on water distribution hydraulics and water quality.” *Proceedings of the ASCE World Water and Environmental Resources Congress, May 15-19, 2005,* Anchorage, Alaska, USA.

Rossman, L. A. (2002). Epanet User’s Manual, Environmental Protection Agency, EPA, Cincinnati.

Rossman, L. A., Clark, R. M., and Grayman, W. M. (1994). “Modeling chlorine residuals in drinking-water distribution systems.” *Journal of Environmental Engineering*, ASCE, 120(4), 803-820.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. 2004. *Sensitivity analysis in practice: A guide to assessing scientific models*, Wiley, Chichester, England.

Saltelli, A., Chan, K.,Scot, E., 2000. *Sensitivity analysis,*Wiley, Chichester, England.

Wagner, J.M., Shamir, U., and Marks, D.H. (1988a). "Water distribution reliability: analytical methods." *Journal of Water Resources Planning and Management*, ASCE, 114(3),253-275.

Wagner, J.M., Shamir, U., and Marks, D.H. (1988a). "Water distribution reliability: simulation methods." *Journal of Water Resources Planning and Management*, ASCE, 114(3),276-294.

Xu, C., and Goulter,I. (1999). "Reliability-based optimal design of water distribution networks."*Journal of Water Resources Planning and Management,* ASCE, 125(6), 352-362.

ZS Kapelan, DA Savic, GA Walters (2007). " Calibration of Water Distribution Hydraulic Models Using a Bayesian-Type Procedure". *Journal of Hydraulic Engineering,* ASCE, 133(8), 927-936.