Uncertainty In Water Distribution Network Modeling
Kevin Lansey
Department of Civil Engineering and Engineering Mechanics, University of Arizona

Introduction

Water distribution network modeling requires input from many sources. Much of this data cannot be measured directly and it is rarely known precisely. Input imprecision is propagated to uncertain model predictions. This paper discusses a technique for quantifying the uncertainty of model results based on the variability in model input.

Once the magnitude of the uncertainty is understood, an engineer can use that information in several ways. Based on the prediction uncertainty in present or future conditions, safety factors can be included in design of network improvements. To enhance model predictions, additional field data can be collected which will lead to more precise model parameters. Using the prediction uncertainty, the optimal network conditions and measurement locations can be identified. Finally, different network representations can be evaluated by comparing their predictive uncertainties.

The model input needed to compute the pressure heads and pipe flows are the nodal demands, pump and pipe characteristics, tank dimensions, valve settings, and coefficients describing energy losses in other appurtenances. Some of this data, such as pipe diameter and length, is accurately known and available from construction plans. Other information for pipes, in particular the pipe roughness, varies over time as encrustation occurs. Nodal demands are difficult to predict since they also change over a short time and because they are modeled as lumped at single locations rather than distributed along a pipe. As such, both parameters are not measured exactly but estimated indirectly from other field measurements. This paper focuses on the uncertainties in pipe roughness coefficients and nodal demands.

Pipe network hydraulics

Flow in a pipe network satisfies two basis principles, conservation of mass and conservation of energy.

Conservation of mass states that, for a steady system, the flow into and out of the system must be the same. This relationship holds for the entire network and for individual nodes. A node is included in a network model at (1) a demand location and/or (2) a junction where two or more pipes combine. One mass balance equation is written for each node in the network as:

\[ \Sigma Q_{in} - \Sigma Q_{out} = Q_{demand} \]

where \( Q_{in} \) and \( Q_{out} \) are the flows in pipes entering or exiting the node and \( Q_{demand} \) is the user demand at that location. These demands are uncertain since they are estimated from the local user base that cannot be predicted exactly since they vary nearly continually. In addition, the demand is typically represented as a lumped demand for users near the node.

The second governing equation is a form of conservation of energy that describes the relationship between the energy loss and pipe flow. The most commonly used head loss equation for water networks, the Hazen-Williams equation, will be the only such relationship considered in this paper. In English units, the equation is written as:

\[ H_i - H_j = h_j = \frac{4.73LQ^{1.852}}{D^{4.87}} \]

where \( D \) is the pipe diameter, \( L \) is the pipe length, \( Q \) is the pipe flow, and \( C \) is the Hazen-Williams pipe roughness coefficient. \( h_j \) is the head or energy loss in the pipe. \( H_i \) and \( H_j \) are the energy at nodes at the ends of the pipe measured in dimensions of length. Using this equation, conservation of energy can be written in several ways. Most often, it is written for energy loss around a loop. The two conservation relationships can be used to develop a set of nonlinear equations that can be solved for the pipe flows, \( Q \), and nodal heads, \( H_i \),
which due to uncertainty in input are inherently uncertain.

**Quantifying uncertainty**

If the variability in input parameters, $C$ and $Q_{demand}$, can be quantified several techniques exist to determine the resulting uncertainty in model output. Monte Carlo analysis (MC) is a general enumerative approach for computing the statistics of a model’s output.

Given the input parameters statistics and probability distributions, one iteration of Monte Carlo analysis consists of randomly generating a set of input parameters. The numerical model, in this case the hydraulic relationships, is then solved to determine the model output. In this study, KYPE (Wood, 1981) is used to compute the nodal pressure heads.

To complete a full analysis consisting of a large number of iterations, a set of random input is generated and their corresponding output is determined. The statistics (e.g. mean, variance, and standard deviation) of the model output can then be computed. If a sufficient number of Monte Carlo iterations are completed, the output statistics will converge to their actual values. The probability distribution of the model output can also be examined.

To estimate the predicted head uncertainty, Monte Carlo analysis was applied to several networks and levels of input uncertainty in Araujo (1992). Figure 1 shows one of the systems considered and Table 1 provides the input statistics for pipe roughnesses. Table 2 summarizes the results estimates of the predicted head variances. It was assumed that the roughness coefficients followed a normal probability distribution. As the input uncertainty increased, the output variability increased at nearly the same rate.

**TABLE 1: Pipe roughness statistics for selected pipes**

<table>
<thead>
<tr>
<th>Pipe</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low standard deviation</td>
<td>0.9</td>
<td>1.5</td>
<td>0.4</td>
<td>1.0</td>
<td>1.4</td>
</tr>
<tr>
<td>Med. standard deviation</td>
<td>4.3</td>
<td>7.7</td>
<td>2.0</td>
<td>5.2</td>
<td>7.1</td>
</tr>
<tr>
<td>High standard deviation</td>
<td>9.1</td>
<td>15.6</td>
<td>4.1</td>
<td>10.0</td>
<td>13.2</td>
</tr>
</tbody>
</table>

**TABLE 2: Standard deviations of nodal pressure heads for selected nodes**

<table>
<thead>
<tr>
<th>Node</th>
<th>Low pipe standard deviation MC</th>
<th>FOSM</th>
<th>Medium pipe standard deviation MC</th>
<th>FOSM</th>
<th>High pipe standard deviation MC</th>
<th>FOSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.04</td>
<td>0.04</td>
<td>0.20</td>
<td>0.21</td>
<td>0.40</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>0.41</td>
<td>2.03</td>
<td>2.10</td>
<td>4.00</td>
<td>4.13</td>
</tr>
<tr>
<td>3</td>
<td>0.08</td>
<td>0.08</td>
<td>0.42</td>
<td>0.44</td>
<td>0.82</td>
<td>0.85</td>
</tr>
<tr>
<td>6</td>
<td>0.03</td>
<td>0.03</td>
<td>0.16</td>
<td>0.17</td>
<td>0.32</td>
<td>0.34</td>
</tr>
<tr>
<td>Trace*</td>
<td>0.18</td>
<td>0.19</td>
<td>4.74</td>
<td>5.08</td>
<td>18.44</td>
<td>19.65</td>
</tr>
</tbody>
</table>

*Trace is the sum of the diagonal elements of the covariance matrix, i.e., variances

As shown by Araujo, the distribution of the pressure heads followed a normal distribution until a high input uncertainty level was evaluated. This result allows one to draw conclusions about the confidence level in model
predictions. For example, the 90 percent confidence level that the pressure head at a node is greater than a desired value can be computed. Thus, when designing a system expansion, an engineer can consider confidence levels on pressure heads as well as their expected values.

Monte Carlo analysis requires a large number of model evaluations and long computer times to determine the desired statistics. Alternative approaches that are less computationally intensive have been proposed in the statistics literature. One popular class of methods is known as point estimation approaches with the simplest being the first order second moment approximation (FOSM). To estimate the mean and variance of model output, FOSM requires computing the model output at a single point and determining the derivative of model output to model input (i.e., change of model output due to a change in model input). In the worst case, these derivatives can be computed by finite differences with one additional model evaluation for each input parameter. The variance of the model output, e.g., the nodal pressure heads, can be expressed as a function of the variance of the model parameter, e.g., Hazen-Williams roughness coefficient, as:

$$\text{cov}(H) = \frac{\partial H}{\partial C} \text{cov}(C) \frac{\partial H^T}{\partial C}$$

where \(\text{cov}(H)\) is the covariance matrix of the predicted pressure head and \(\text{cov}(C)\) is the covariance matrix of the roughness coefficient. The derivative terms are the matrix of representing the change in pressure heads due to a change in roughness coefficients. The key result in the \(\text{cov}(H)\) matrix is the diagonal elements that are the variances of the individual pressure heads. The mean value of the output is assumed to occur when the value of the input is evaluated.

FOSM was used to predict model uncertainties of the network in Figure 1 with results again summarized in Table 2. Although FOSM is an approximation technique, the results compare very favorably with Monte Carlo analysis. However, the difference in computations is dramatic. Thus, FOSM was used for the analysis in the following sections. FOSM, however, begins to diverge from Monte Carlo analysis at high levels of uncertainty. This result is due to model nonlinearity at higher uncertainty levels while the FOSM, by using only the first derivatives, neglects this nonlinearity.

Collection of field data to reduce uncertainties

If input uncertainty is very high, a model is likely poorly calibrated and additional field data should be collected to reduce the input uncertainty. Uncertainty analysis can assist in determining where and under what conditions additional data should be taken to improve model calibration. Araujo (1992) developed a methodology for comparing the utility of field data collected from a set of potential demand conditions. The data collection methodology is based on lowering the uncertainty level at one or more nodes through a reduction in one or more pipe roughness coefficient uncertainty.

Roughness uncertainty is dependent upon field measurements. Therefore, a two level gradient was computed. The first level estimates the change in pipe roughness uncertainty given a demand pattern and assumed measurement locations. Extreme demand conditions should be selected, such as simulated fire loads, that stress the network and cause high energy loss. These types of demand conditions can be induced by opening one or more fire hydrants and supply useful information for improving parameter estimates. The uncertainty analysis is completed using the FOSM method assuming that the predicted demand occurs using:

$$\text{cov}(C) = \frac{\partial C}{\partial H_M} \text{cov}(H_M) \frac{\partial C^T}{\partial H_M}$$

where \(H_M\) is the measured pressure heads and \(\text{cov}(H_M)\) is the covariance matrix for the measured heads. The covariance of \(H_M\) varies depending upon the device used for measurement and the accuracy of determining elevations. The gradients are evaluated at the best estimate of the \(C\) values and measured heads.

The second level of the methodology computes the propagation of the expected new pipe roughness uncertainties to nodal pressure head predictions. The second step is identical to the uncertainty predictions described in the previous section. By comparing the reductions in pressure head uncertainties for different demand patterns, the conditions that provide the most
improvement can be identified. This condition should then be induced in the field and the network recalibrated with the new data. If any nodes continue to have unacceptably uncertainty levels, the process can be repeated to determine other measurements that are useful in further reducing nodal prediction uncertainty.

**Pipe network representation**

Uncertainty analysis can also be helpful in identifying how a network should be represented in a mathematical model. In most modeling efforts, limited field data is collected to determine model parameters. Typically, the data is insufficient to estimate parameters for individual pipes. As such to improve the confidence in the calibrated pipe roughness coefficients, pipes are grouped and all pipes within a group are assumed to have the same roughness value. Groupings may be made arbitrarily or using judgement after examining pipe material and installation records. By reducing the number of parameters, confidence in the estimated values increases and the parameter uncertainty or error decreases.

Although the parameter uncertainty decreases, the group roughness coefficients are not individual pipe values and can be considered as averaged or surrogate model parameters. Thus as fewer pipe groups are considered, the mathematical model is less representative of the true system and a model error is introduced. The tradeoff between errors can be seen in Figure 2. As the number of parameters increases the model error decreases (for the reason above) and the parameter error increases since more parameters are being estimated from the same amount of information. An envelope is shown since several pipe groupings can have the same number of parameters but have different errors. For example if all pipes in a pipe group have similar field roughness coefficients, the model error will be smaller than a pipe grouping that the pipe roughnesses are very different. The key to modeling is identifying the tradeoff between the two uncertainties. Mallick and Lansey (1994) developed a technique to compare these errors. The technique is an extension of the work by Yeh and Yoon (1981) for groundwater modeling. For a set of field data and a network model with a defined pipe grouping, the optimal parameters can be identified using one of several approaches (Basnet and Lansey, 1989). Model error is defined as the sum of the squares of the differences between the field measurements and the model predictions after calibration. At the optimal parameter values, parameter error can be computed using FOSM as described in the previous sections.

A number of combinations of pipe groupings are possible for even the smallest pipe network. To compare pipe groupings, model versus parameter error can be plotted for all groupings (Figure 3). Both types of errors are small for points close to the origin. These pipe grouping combinations are the best model representations.

**Conclusions**

Imprecise field measurements result in uncertain model predictions for all types of models including those for water distribution networks. High uncertainties require increased safety factors and result in more costly designs. Thus, an understanding of the magnitude of prediction uncertainties is critical when making planning and operation decisions based on numerical models. Rapid evaluation of prediction uncertainties can be completed for pipe networks using FOSM analysis. Model prediction probability distributions were also determined which allow confidence levels to be estimated.

Because of its speed, FOSM can be used as a block in other network analysis studies. Two applications, data collection and network representation, were described in this paper. In addition, the uncertainty approach can be used in reliability analyses as described by Xu and Goulter (1996).

**References**


**Kevin Lansey** is an Associate Professor in the Department of Civil Engineering and Engineering Mechanics at the University of Arizona. He is an active researcher in applying systems analysis approaches to water resources systems with emphasis on water distribution systems.