Scenario generation for network optimization with uncertain demands^

Bradley Eck*, Francesco Fusco, Nicole Taheri

IBM Research, Damastown Ind. Estate Building 3, Mulhuddart, Dublin 15, Ireland
*bradley.eck@ie.ibm.com; Ph: +353 1 826 9354

^ Please cite as: Eck, B., Fusco, F., and Taheri, N. (2015) Scenario Generation for Network Optimization with Uncertain Demands. World Environmental and Water Resources Congress 2015: pp. 844-852.
doi: 10.1061/9780784479162.079

ABSTRACT

The demand imposed on utility networks is a fundamentally uncertain quantity that varies in both space and time. Utilities have long dealt with this uncertainty by maintaining additional capacity. But, technological advances in measuring and especially transmitting water usage information allow further improvement in operational efficiency. Unlike the monthly or annual values obtained by automatic meter reading, smart water meters measure consumption at sub-daily intervals. This increased resolution, combined with measurements of bulk flows and network pressures, presents an opportunity to characterize the spatio-temporal distribution of demand. Once this representation of demand uncertainty is available, it can be used within optimization models to support decisions in network operations and planning.

Several methods are available for dealing with uncertain parameters in optimization problems including sensitivity analysis, stochastic programming and robust optimization. This work deals with robust optimization, which characterizes uncertainty by scenarios or uncertainty sets. The robust solution to an optimization problem gives the best objective value in the worst case, which is feasible over the range of possible values. The choice of scenarios is a critical step in obtaining meaningful solutions with the right compromise between system performance (value of the objective function) and robustness to variations in uncertain parameters.

This paper examines several techniques for generating scenarios of uncertain demands. Based on the spatial distribution of demands modeled as a multi-variable Gaussian with a mean vector and covariance matrix calculated at each time step, a technique is proposed to generate scenarios of demand which (a) cover a given probability level and (b) adhere to the spatial correlation of demands between nodes. The technique is compared with Monte Carlo sampling using the same co-variance matrix. Particular emphasis is given to the interpretation of scenarios and their effect on the solution to the operational problem of finding valve set points for a water network.

INTRODUCTION

With the growth of smart metering technologies, information on water demand at individual consumption points is becoming more widely available. Opportunities now exist for using such information to improve operational and planning decisions in water systems. A numerical model of the network is often used to support operational and planning decisions. For these purposes, consumption points are represented individually or aggregated over part of the service area. Demand at consumption points varies in both space and time. With smart meters, this variation can be better described and hence used to improve network management.

One way of using data to improve operations is to characterize uncertainty in demands and use this characterization to support decision making. Decision support tools often come in the form of optimization models. Several techniques exist for considering uncertainty in optimization models including stochastic optimization, sensitivity analysis and robust optimization. Each of these methods needs a characterization of the uncertainty. Here we consider a scenario-based robust approach.

Uncertain demands have been included in optimization models by earlier workers using a variety of approaches. The optimization problem most commonly studied is that of network design, though results from this problem are broadly transferrable to other optimization problems on water networks. An early study considering uncertain demands is reported by Lansey et al. (1989) who included uncertain nodal demands in a chance-constrained non-linear program. Tolson et al. (2004) give a first-order reliability method using genetic algorithms as the optimization tool. They observe an efficient approximation of Monte Carlo simulation reliabilities. Kapelan et al. (2005) give a multi-objective approach to minimize design cost whole maximizing robustness. Babayan et al. (2005) constrain the minimum pressure to include a margin of safety. The chance of violating pressure constraints is evaluated for critical nodes by assuming that variations in head are primarily due to variations in pressure at that node. Fu and Kapelan (2011) give an approach based on fuzzy logic. Perelman et al. (2013) proposes a robust approach where demands are constrained by an ellipsoidal uncertainty set.

The present work builds on earlier studies by illustrating how scenario selection influences the result of an optimization problem. Existing techniques for updating nodal demand info and sampling multivariate distributions are also summarized.

METHODS

Estimating mean and covariance of nodal demand. Consider a water network where nodal demand is described by a mean vector d and covariance matrix Ω. We assume a prior estimate of the mean vector and covariance matrix are available from a preliminary analysis. For example, a prior estimate could use rough assumptions about the daily profiles and their standard deviation. By introducing sensor data such as pressure loggers, smart meters, or flow sensors, updated estimates of demand, variance, and co-variance may be obtained through state estimation methods.

The state vector is defined as any minimum set of hydraulic quantities which, at any point in time, are sufficient to fully specify the hydraulic state of the network. For a network with n floating head nodes and m links a combination of m variables can be selected. Examples of state vectors can be the set of m link flows, or the set of n demands plus $(m-n)$ flows through one closure link in each of the $(m-n)$ loops.

The measurements (observations) y are related to the state x by

$$y = h(x) + \epsilon \qquad (1)$$

Where ϵ is a vector of measurement errors. The function $h(x)$ represents the measurement model. The specific form of h(x) depends on the choice of the state vector but in general will depend on the network connectivity, pipe parameters, node elevations, and other parameters. The measurements include both actual field measurements (flow/pressure meters, smart meters) and prior assumptions on the demand. The covariance matrix of the measurement error ϵ is assigned based on the uncertainty of such measurements and priors.

An estimate of the states $\hat{x}$ is found from solving the least squares problem

$$\text{Min J(x)} = [y - h(x)]^T W [y - h(x)] \qquad (2)$$

through the recursion

$$\hat{x}_{k+1} = \hat{x}_k + (H_k^T W H_k)^{-1} H_k^T W [y - h(x_k)] \qquad (3)$$

where W is a weighing matrix typically set to the inverse of the covariance matrix of the measurement error Σ so that the weight given to measurements (or priors) is inversely proportional to their uncertainty. H_k is the Jacobian matrix of the model $h(x)$evaluated at the point x_k. The covariance corresponding to the state-estimate is estimated as

$$\widehat{\Omega} = (H_k^T W H_k)^{-1} \qquad (4)$$

where Σ is the diagonal matrix of measurement variances. From the state estimate $\hat{x}$ a posterior estimate of the demands is given as

$$\hat{d} = g(\hat{x}) \qquad (5)$$

Note that the specific form of the function $g(x)$ also depends on the choice of state variable. For example, if the state variable is the set of link flows, then $g(x)$ is a linear function given by the connectivity matrix, or it is an identity matrix in case the state vector is formed by the nodal demands. In this way, smart meter data and other measurements can be included in an estimate of network demands and their covariance matrix.

Scenario Selection and Sampling. For operational or planning purposes, we would like to generate scenarios of water demand which are consistent with the estimated mean vector and covariance matrix and which enclose a given proportion of the sample space. For multivariate distributions there are an infinite number of scenarios that satisfy these two requirements.

For multivariate Gaussian distributions, a fixed probability level represents a distance from the mean vector. In two dimensions these points form an ellipse, and in higher dimensions they form an ellipsoid. The locus of points on the distribution with the same probability, typically referred to as constant density contour, can be found from (Johnson and Wichern, 2007):

$$x = L(Ru) + \mu \qquad (6)$$

where x are points on the ellipsoid; L is Cholesky factor of the covariance matrix; Ru is a set of points on a spheroid of some radius, and the vector of means is μ. Points on a spheroid of unit radius are u. R is found from the inverse probability of Chi square for the desired confidence level.

$$R = \sqrt{F_k^{-1}(\alpha)} \tag{7}$$

F_k^{-1} is the inverse of the chi-square distribution for k degrees of freedom at cumulative probability level α.

At a chosen cumulative probability there are infinitely many scenarios. Which scenarios are reasonable to select? A single scenario may be selected as the medoid of realizations on the ellipsoid, or based on the total demand of the scenarios, for example picking the high and low total demands. If multiple scenarios are desired these can be chosen as combinations of the above, or by assigning realizations into the desired number of clusters, taking the medoid of each cluster generated by portioning around medoids (Kaufman and Rousseeuw, 1990).

Optimization. Uncertain water demand figures into many operational and planning problems on water networks. We examine an operational problem of valve setting, formulated as a linear program. Uncertain demands are included as discrete scenarios. The chosen objective is to minimize the maximum of the sum of pressures over all nodes in the network over the chosen scenarios. Minimizing the sum of pressures serves as a surrogate for minimizing pressure driven leakage.

$$minimize \ \max_{s} \sum_{i} \mathrm{p_{i,s}} \tag{8}$$

With this objective, the optimization model finds valve settings that minimize the worst-case objective but are feasible for all the scenarios. The optimization is carried out under the constraints of mass conservation around each node

$$\sum_{i} q_{ij,s} + \sum_{k} q_{jk,s} = d_{j,s} \tag{9}$$

where $q_{ij,s}$ is the flow toward node j from node i in demand scenario s and the matrix of nodal demands $d_{j,s}$ contains the demand scenarios. Energy conservation for pipes is modeled as a linear function, where coefficients $a$and b are obtained by linearizing around the flow rate in each pipe for the mean demand case.

$$(p_{i,s} + e_i) - (p_{j,s} + e_j) = a_{ij}q_{ij} + b_{ij} \tag{10}$$

Energy conservation over PRVs is constrained with the assumption that valves have an active status:

$$(p_{i,s} + e_i) - (l_{j,s} + e_j) > 0 \tag{11}$$

In this way the pressure level at the downstream node of the valve, l_j, is a free variable and is used to model the valve setting. Simple bounds are enforced on link flow rates and nodal pressures:

$$q_{\min} < q_{ij} < q_{max} \tag{12}$$

$$0 < p_i < p_{max} \tag{13}$$

Evaluation. Monte Carlo analysis is used to study the effect scenario selection has on the performance of the system. Valve settings obtained from different demand scenarios are simulated using Epanet for randomly sampled cases to understand the distribution of performance statistics. The statistics of primary interest were the frequency and magnitude of constraint violation and the behavior of the objective value.

APPLICATION

As an application example we examine a version of the network first studied by Sterling and Barglia (1984). Pipe properties, node elevations, and mean demands are listed by Eck and Mevissen (2014) except in this paper all reservoirs had a head of 55.5m and PRVs are located as shown in Figure 1. Also shown in Figure 1 is the covariance matrix of nodal demands used in this example. The covariance matrix was derived from a randomly generated matrix of correlation coefficients and a coefficient of variation of 30%.

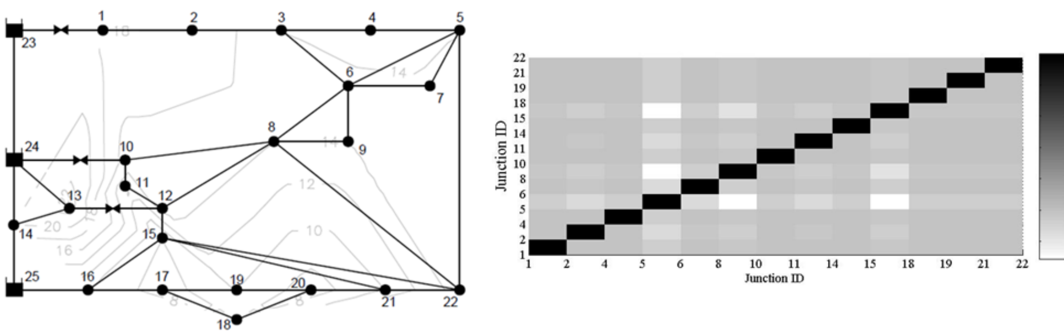

Figure 1 Twenty-five node network shown with elevation contours (left) and covariance matrix for this example (right)

Demand realizations at a cumulative probability of 90% are shown in two different ways. Pair-wise plots of demand for a subset of nodes are shown in Figure 2. Each dot of the same color corresponds to a single demand scenario. For each pair of nodes, the demands fall inside an ellipse.

Another way to visualize realizations of demand in higher dimensions is a cumulative plot (Figure 3). On the horizontal axis nodes are sorted by mean demand. The vertical axis shows the cumulative demand for nodes with a mean demand greater than that of the node indicated by the horizontal axis position. Lines of different colors represent realizations chosen by different criteria. The vertical spread for a node gives a sense of the variance in it’s demand.

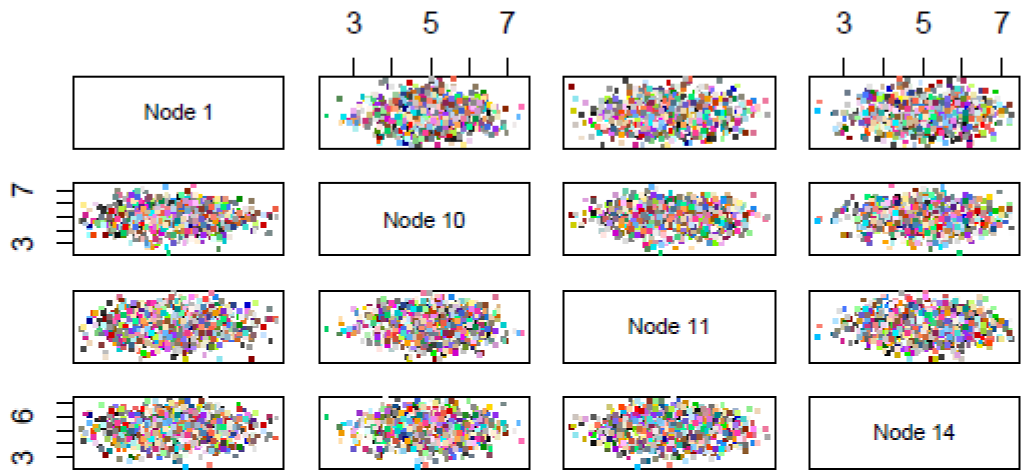

Figure 2 Sample pairwise plot of demand scenarios on the ellipsoid for 90% probability level.

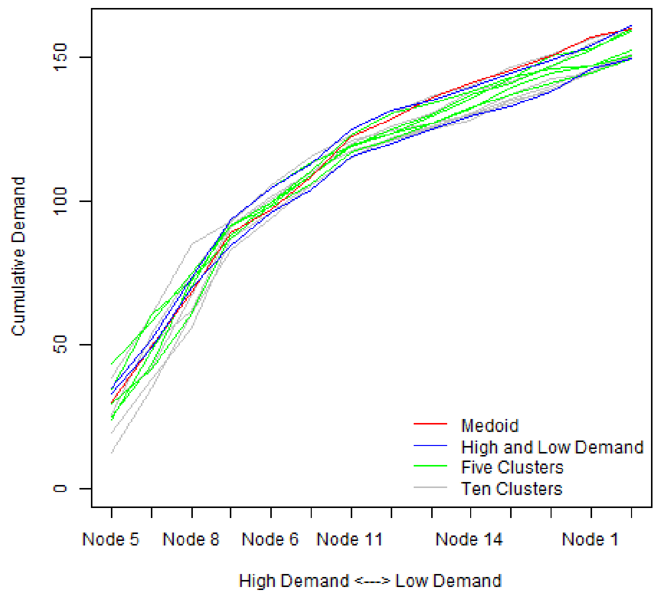

Figure 3 Demand realizations on 90% ellipsoid for SB25

Valve settings were obtained by optimizing over different sets of scenarios for a minimum pressure of 20m (Table 1). Valve settings varied over a small range, but nonetheless some trends were visible. The lowest settings are observed for the minimum demand case, reflecting the need to accomplish more pressure reduction in the PRVs when friction losses are low. As expected, considering more scenarios caused the set points to increase in order to maintain the required minimum pressure for all the cases. The highest settings appear for the case of 1000 scenarios. The same settings were obtained in the case of 25 and 50 scenarios.

Two performance measures were of initial interest. First, how often is the prescribed minimum pressure violated? Second, how does the objective value behave for different settings and scenarios? Monte Carlo simulations of 10,000 realizations were carried out using the optimized valve settings to explore these metrics (Table 1). The mean objective value (sum of pressures) ranged from 666m for the minimum demand case to 671m in the majority of other cases. Although not tabulated, the mean of objective values had a standard deviation of 5.9 in all cases.

Perhaps the most interesting results relate to the frequency of constraint violation. Valve settings obtained from only one demand scenario violated the min pressure constraint in 71-93% of cases. This high occurrence of constraint violation illustrates the challenge of selecting a single scenario on which to base decisions. The optimization model finds settings so that at least one node has the minimum pressure. When the demand distribution changes even slightly constraint violations can occur. A histogram of minimum pressures (Figure 4 gives a typical example) shows that most constraint violations are less than 0.5m but larger violations up to 2.5m do occur. By including more scenarios in the optimization, constraint violation occurs less often. This improved performance comes at a very low cost in terms of the expected objective value. Cases with more than 25 scenarios all had the same objective value to three significant figures, further indicating that most constraint violations were small. Considering more than 25 scenarios showed a diminishing return in terms of constraint violation and objective value.

Constraint violations were surprisingly high when only one scenario was used for optimization. This result is related in part to the linearized model used for this study. Using a Taylor approximation at one point on the head loss curve, the model under-estimates head loss and over-estimates pressures. The resulting settings are therefore too low giving a high occurrence of constraint violations when evaluated by Epanet simulations. Strengths of the approximate linear model include the ability to optimize quickly using commercial solvers so that a large number of scenarios may be considered.

Table 1 Valve settings from optimizing over different demand scenarios and performance statistics from 10,000 Monte-Carlo simulations.

Case	Node 1	Node 10	Node 12	Mean of Objective	Freq. P<Pmin [%]
Min Demand	20.0	21.7	21.6	666	93
Med Demand	20.1	21.8	21.7	668	71
Max Demand	20.0	21.9	21.8	668	87
Min, Med, Max	20.1	21.9	21.7	667	67
5 scenarios	20.1	22.0	21.9	671	45
10 scenarios	20.1	22.0	21.8	670	52
25 scenarios	20.1	22.2	22.0	671	36
50 scenarios	20.1	22.2	22.0	671	36
100 scenarios	20.1	22.2	21.9	671	33
200 scenarios	20.1	22.2	22.0	671	29
500 scenarios	20.1	22.2	21.9	671	32
1000 scenarios	20.1	22.3	22.0	671	27

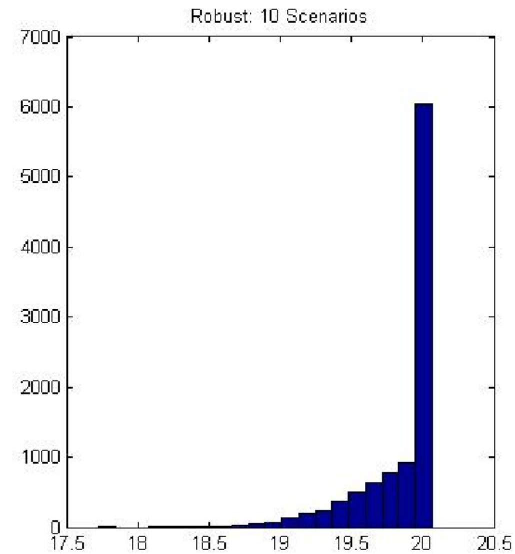

Figure 4 Density of minimum pressure for PRV settings from optimizing over 10 scenarios

CONCLUSION

This paper has examined scenario generation for uncertain demands in water systems. Measurements of demand coming from smart meters or other devices can be used to update the mean vector and covariance matrix of nodal demands via a state estimation technique. Using multivariate Gaussian distributions, scenarios which correspond to a chosen cumulative probability can be generated using the Cholesky decomposition of the covariance matrix. A handful of scenarios can be chosen from the many possible scenarios of the same probability based on extreme values, cluster analysis or other methods.

Incorporating multiple scenarios into the optimization provided a more robust result at very little cost in terms of the objective value. Considering even a small number of scenarios gave a substantial reduction in the occurrence of minimum pressure violations for essentially no change in the objective value. This small effect on the objective indicates that in this case constraint violations were small. Future work will examine nonlinearities in the optimization model and the connection between the cumulative probability of scenarios and the frequency of constraint violation.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318272

REFERENCES

Eck, B. J. and Mevissen, M. (2014) Quadratic approximations for pipe friction. Journal of Hydroinformatics. doi:10.2166/hydro.2014.170

Fu, G., and Z. Kapelan (2011), Fuzzy probabilistic design of water distribution networks, Water Resour. Res., 47,W05538, doi:10.1029/2010WR009739.

Johnson, Richard and Dean W Wichern. (2007). "Applied Multivariate Statistical Analysis". Pearson Education, Inc.
Kapelan, Z. S., D. A. Savic, and G. A. Walters (2005), Multiobjective design of water distribution systems under uncertainty, Water Resour. Res., 41, W11407, doi:10.1029/2004WR003787
Kaufman, L. and Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.
Lansey, K. E., N. Duan, L. W. Mays, and Y. K. Tung (1989), Water distribution system design under uncertainties, J. Water Resoures Planning and Management, 115(5), 630–645, doi:10.1061/(ASCE)0733-9496(1989)115:5(630).
Perelman, L., M. Housh, and A. Ostfeld (2013), Robust optimization for water distribution systems least cost design, Water Resour. Res., 49, 6795–6809, doi:10.1002/wrcr.20539.
Sterling, M. & Bargiela, A. (1984) Leakage reduction by optimized control of valves in water networks. Trans. Inst. Meas. Control 6 (6), 293–298.
Tolson, B., Maier, H., Simpson, A., and Lence, B. (2004). "Genetic Algorithms for Reliability-Based Optimization of Water Distribution Systems." J. Water Resour. Plann. Manage., 130(1), 63–72.