Real Time Water Demand Estimation in Water Distribution System

Feng Shang[*], James G. Uber[*], Bart G. van Bloemen Waanders[†], Dominic Boccelli[‡]and Robert Janke[‡]

Abstract

Accurate modeling of chemical transport in water distribution systems depends on accurate knowledge of temporally and spatially variable water demands. Typical network models would include water demands that are allocated from billing or census data, and thus may not be appropriate for specific operational analysis, such as emergency events arising from intentional or accidental contamination. During such an event, water consumption patterns may be significantly different from those assumed when developing the hydraulic model, and may change significantly over short time periods due to the unusual circumstances of the event. To allow accurate hydraulic and water quality prediction in real-time, the water demands should be updated continuously to reflect current conditions. The development of such a real-time water demand calibration method poses many complex issues such as identifiability and uncertainty of the water demand estimates. Given the sparsity of data that are likely to be available in real time, prior statistical information about water demands must be incorporated in the calibration procedure. In this paper, a method and algorithms are proposed for a real time water demand calibration process. A predictor-corrector methodology is proposed to predict statistical hydraulic behavior based on prior estimation of water demands, and then correct this prediction using new, real-time measurements. The problem is solved using the extended Kalman filter, which is a linear algorithm that calculates the estimate of water demands and their uncertainty. As part of the Kalman filter calculation, we calculate direct sensitivities of system hydraulic responses with respect to water demands. Results of numerical experiments illustrate the impacts of statistical demand variability, hydraulic measurement accuracy and sampling design on demand estimation.

1 Introduction

Computer models of water distribution systems, have been widely employed in system design and analysis. Many water utilities have dedicated effort to develop skeletonized or all pipe models for their network systems, with all but a few models and applications being limited to off line analysis. Models are calibrated using data sampled during certain short periods of time (usually several days) and only simulate the hydraulic and water quality conditions under specific water demand and operational scenarios. Despite the wide application of supervisory control and data acquisition (SCADA) systems and the availability of nearly real time data, water distribution system models

---

are not updated frequently. The gap between the efforts of online data collection and real time model updating is significant.

Off line models are not expected to accurately represent the hydraulic or water quality behavior of the system during emergency events, such as a main break, a significant fire, or an intentional or accidental contamination event. In order to appropriately assess the consequences of such events and develop response plans and control strategies, it is critical to update water distribution system models in real time and use the models to simulate the current hydraulic and water quality conditions of the network and changes that will be caused by responsive operations. Daily operation of water utilities can also benefit from real time network models since they provide a powerful mechanism to quickly synthesize the SCADA data and analyze the network system performance. A historical database of models updated for realized operational conditions, and resulting simulations of hydraulics and water quality, would be valuable for analyzing and improving the system operations.

Arguably, the most uncertain and variable dynamic state of a water distribution system is the water demands. The real time calibration of water demand is a challenging task whose difficulty depends on the amount of SCADA information, the aggregation of pipe detail and water demands (skeletonization/parameterization), the statistical characteristics of demand variability, and the accuracy of statistical demand models used to condition the predictor-corrector procedure. In reality, the number of measurements used to estimate demands of a network are limited by cost and these measurements are usually corrupted by a certain level of noise. Therefore, field measurements alone can not provide enough information to reliably estimate system demands that change over time. Even a skeletonized water network can have hundreds of junctions for which the aggregated demands are to be estimated. If the number of measurements is less than than the number of demand values to be estimated, the estimation problem may be ill-posed and there is no unique solution. We must incorporate additional information about the demands to get meaningful estimation results. For the water demand estimation problem, the temporal and spatial correlations of the demands should be considered.

2   Methodology

Predictor-Corrector Approach

The Predictor-corrector (Maybeck, 1979) approach for state estimation is applied to approximate demands. Based on previous information, predictions of the demands and measurements at the next time step are made. After the next set of measurements are taken, the differences between the "measured" and "predicted" measurements are used to "correct" the prediction of the demands.

Figure 1 illustrates the process to estimate water demands in real time. The following notation is defined:

- $d_{k/k}$: estimated demands for time step k after measurements are taken, i.e. posterior demand estimates,

- $d_{k/k-1}$: estimated demands for time step k at time step $k-1$, i.e. prior demand estimates;

- z_k: measurements taken at time step k;

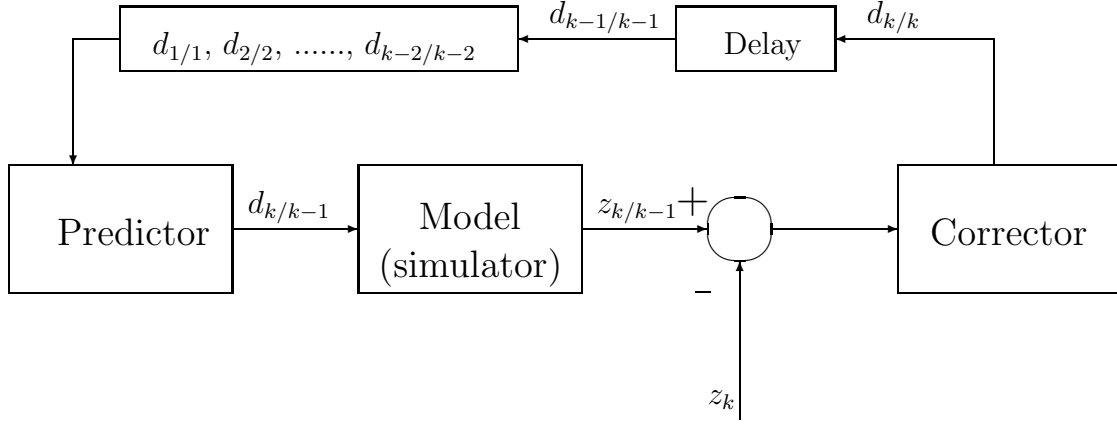- $z_{k/k-1}$: predicted values of z_k one time step ahead;

Figure 1: Predictor-Corrector Approach for Demand Estimation

- k: time step index.

In this paper, a time step interval is taken as one hour and demand estimates get updated every hour, although the methodology is defined for a general time step size. Measurements are limited to junction water heads and pipe flow rates. Demands at time step k are likely to be correlations with demand at multiple steps behind, and it is not appropriate to predict demands at k using demand estimates only one step behind.

In this work, water demands are the only uncertain model input; other model parameters, such as tank levels and pipe roughness, are assumed to be known. Pipe roughnesses are usually calibrated off line and assumed to be constant for a certain period of time. Water levels are assumed to be monitored at every tank and set as the boundary conditions of the system model at each time step. Field practice may require the real time calibration of both water demands and other system parameters that change over times, such as tank water levels, but this paper considers only the real time calibration of water demands.

Water Demand Model

One traditional way to model water demand uses the concept of base demand and pattern value.

$$d_k^i = b^i p_k^i \tag{1}$$

where d_k^i = demand at time step k and junction i; b^i = base demand at junction i; and p_k^i = demand pattern value at time step k and junction i. In such a demand model, base demand is usually set to be the junction average demand. Considering the nature of water demand data, we assume that pattern values can modeled as a periodic time series using a seasonal ARIMA(p,d,q)(P,D,Q)s approach (Box and Jenkins, 1976) which has the general form:

$$(1 - B)^d (1 - B^s)^D p_k^i = \mu^i + \frac{\theta(B)\theta_s(B^s)}{\phi(B)\phi_s(B^s)} \epsilon_k^i \tag{2}$$

where B = backward shift operator and $Bp_k = p_{k-1}$; μ^i = the mean value of the time series $(1 - B)^d (1 - B^s)^D p_k^i$; p = order of autoregressive part; d = order of the differencing; q = order of the moving average part; P = order of the seasonal autoregressive part; D = order of seasonal

differencing; Q = order of the seasonal moving average part; s = length of the seasonal cycle; $\phi(B)$ = the autoregressive operator which is represented as a polynomial in the back shift operator B: $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$; $\theta(B)$ = the moving average operator which is represented as $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$; $\phi_s(B^s)$ = the seasonal autoregressive operator represented as: $\phi_s(B^s) = 1 - \phi_{s1} B^s - \cdots - \phi_{sP} B^{sP}$; $\theta_s(B^s)$ = the seasonal moving average operator represented as: $\theta_s(B^s) = 1 - \theta_{s1} B^s - \cdots - \theta_{sQ} B^{sQ}$; and ϵ^i = a sequence of independent $N(0, \sigma_i^2)$ random variables, i.e. ϵ^i is a white noise.

For water demand data modeling, we have $s = 24$ (1 day) or $s = 168$ (1 week) when the step length is set to be 1 hour. An ARIMA(1,0,1)(0,1,1) model with a 7 day period has been used to described traffic flow (Williams and Hoel, 2003), which also exhibits daily and weekly cycles. In this paper, we assume the following ARIMA(1,0,1)(0,1,1)$_{24}$ model for demand pattern value:

$$(1 - B^{24})p_k^i = \mu^i + \frac{(1 - \gamma B)(1 - \beta B^{24})}{1 - \alpha B}\epsilon_k^i \tag{3}$$

where the model parameters are set to be $\alpha = 0.95$; $\beta = 0.15$; and $\gamma = 0.85$.

Since $\alpha < 1$, the single root of autoregressive polynomial $1 - \alpha B$ lies out of the unit circle, ensuring stationary (Box and Jenkins, 1976) of the 24-hour seasonal difference minus μ^i. We can see that μ^i represents the seasonal trend of pattern value and when it equals zero, there is no trend in pattern values over the 24 hour period and the mean demand is stationary. With pattern values modeled as a time series, the demand model (1) can be written as:

$$d_k^i = d_{k-24}^i + \alpha(d_{k-1}^i - d_{k-25}^i) + b^i(\epsilon_k^i - \beta\epsilon_{k-1}^i - \gamma\epsilon_{k-24}^i + \beta\gamma\epsilon_{k-25}^i) \tag{4}$$

By the white noise assumption, $cov(\epsilon_k^i, \epsilon_l^i) = 0$ for $k \neq l$ and $cov(\epsilon_k^i, \epsilon_l^i) = \sigma_i^2$ for $k = l$. For simplicity, it is assumed in this paper that variances of the white noises ϵ are the same, i.e. $\sigma_1 = \sigma_2 = \cdots = \sigma$.

Simulation can show the seasonal stationarity of the demand data generated using seasonal ARIMA model (4), with model parameters, α, β, and γ, defined above. Demand unit here can be arbitrary. Demands at first 25 time steps (hours) are needed as the initial condition of a time series model and they are sampled from independent Gaussian distributions, of which the mean values are shown in plot (a) of Figure 2. It is noted that the mean at hour 25 is set to be the same as the mean at hour 1 in order to check the stationarity of the time series model: setting mean value at hour 25 different than that at hour 1 will violate the seasonal stationarity. 5000 realizations are generated for a demand time series with duration 240 hours. Errors are sampled from white noise with $\sigma = 5$. Mean demand values over these 5000 realizations are calculated for each of the 240 hours and the results are shown in plot (b) of Figure 2. It can be seen that the mean values at hour $k + 24$ are very close to those at hour k and the time series model is stationary every 24 hours. For example the mean demands at hour 2, 26, 50, 74, 98 ,122 and 218 are 159.8, 159,9, 160.2, 159.9, 159.9, 159.9 and 160.0, respectively.

In a time series model, the random variable ϵ turns out to be the *one step ahead forecast errors* (Box and Jenkins, 1976). In our demand model,

$$b^i\epsilon_k^i = d_k^i - \hat{d}_k^i \tag{5}$$

where $\hat{d}_k$ = best forecast based on demands at previous steps. Under the Gaussian noise assumptions, the best forecast is the conditional mean,

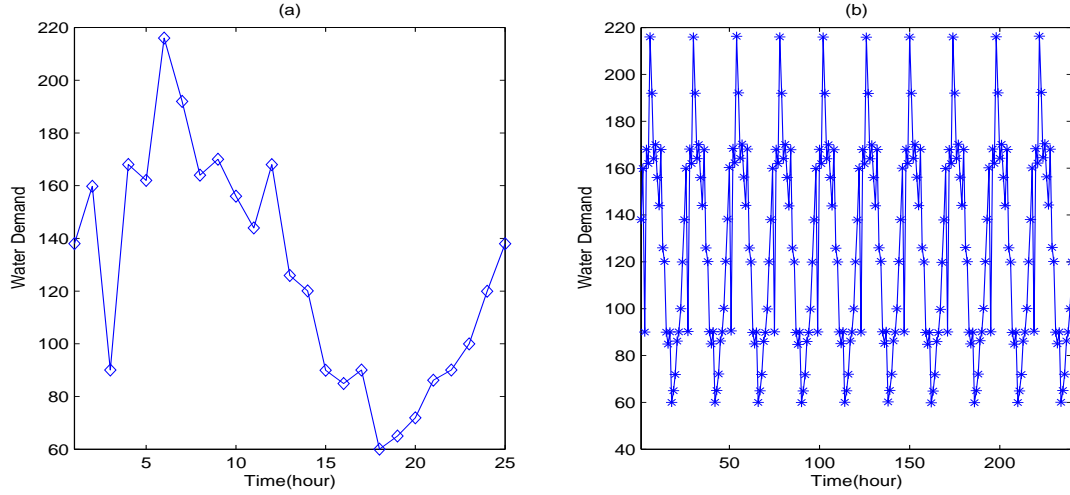$$\hat{d}_k = E[d_k|d_{k-1}, d_{k-2}, \cdots] \tag{6}$$

4

Figure 2: Mean Demands of the first 25 Hours (a) and Sample Mean of the Demands Generated by Time Series Model (b)

Therefore ϵ represents the deviation of demands from their conditional mean values.

The variance and covariance of the demand forecast errors can be calculated,

$$var(d_k^i - \hat{d}_k^i) = (b^i\sigma)^2 \tag{7}$$

$$cov(d_k^i - \hat{d}_k^i, d_k^j - \hat{d}_k^j) = b^i b^j cov(\epsilon_k^i, \epsilon_k^j) \tag{8}$$

If $\epsilon_k^i = \epsilon_k^j$, demand forecast errors at nodes i and j are perfectly correlated because $cov(\epsilon_k^i, \epsilon_k^j) = var(\epsilon_k^i) = var(\epsilon_k^j) = \sigma^2$. Another extreme situation is that demand deviation at nodes i and j are independent, i.e. $cov(\epsilon_k^i, \epsilon_k^j) = 0$. In this paper, ϵ_k^i is modeled as a linear combination of two independent white noises ν_k and ω_k^i with $N(0, \sigma^2)$ distribution.

$$\epsilon_k^i = a_i\nu_k + b_i\omega_k^i \tag{9}$$

The white noise ν_k is shared over all junctions and account for the common factors (global randomness) that affect the demand deviation from the mean, or normal demands, while ω_k^i represents the local randomness and $cov(\omega_k^i, \omega_k^j) = 0$ if $i \neq j$. Since $var(\epsilon_k^i) = var(a_i\nu_k + b_i\omega_k^i) = (a_i^2 + b_i^2)\sigma^2$, $a_i^2 + b_i^2$ must equal 1. The covariance between ϵ_k^i and ϵ_k^j $(i \neq j)$ can be calculated: $cov(\epsilon_k^i, \epsilon_k^j) = cov(a_i\nu_k + b_i\omega_k^i, a_j\nu_k + b_j\omega_k^j) = cov(a_i\nu_k, a_j\nu_k) + cov(a_i\nu_k, b_j\omega_k^j) + cov(b_i\omega_k^i, a_j\nu_k) + cov(b_i\omega_k^i, b_j\omega_k^j) = a_i a_j\sigma^2$.

Spatial correlation between demand forecast errors over junctions now can be controlled using different a and b: demand forecast model errors at node i and j are perfectly correlated when $a_i = a_j = 1$, while they are independent when either a_i or a_j equals zero.

The proposed demand model is a simple multivariate time series model. General multivariate time series such as node demands can be modeled by vector autoregressive moving average models, which have a more complex mathematical structure. At the current stage, this simple model provides a mechanism to study the impact of spatial correlation between demand forecast errors on demand estimation.

5

State Space Demand Model and Kalman Filter

A wide range of time series models can be put into state space form so that the Kalman filter can be applied as one predictor-corrector algorithm (Harvey, 1989). For demand time series model (4), one state space form is:

$$x_k^i = F^i x_{k-1}^i + G^i \epsilon_k^i \tag{10}$$

$$d_k^i = Z^i x_k^i \tag{11}$$

where F^i is a 26×26 matrix,

$$F^i = \begin{bmatrix} \alpha & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} \\ 1 & 0 & 0 & \cdots & 1 & 0 \\ -\alpha & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{12}$$

G^i is a 26×1 vector,

$$G^i = [b^i, b^i\beta, \mathbf{0}, b^i\gamma, b^i\beta\gamma]^T \tag{13}$$

Z^i is a 1×26 vector,

$$Z^i = [1, \mathbf{0}] \tag{14}$$

x_k^i is a 26×1 vector,

$$x_k^i = \begin{bmatrix} d_k^i \\ d_{k-23}^i - \alpha d_{k-24}^i + b^i\beta\epsilon_k^i + b^i\gamma\epsilon_{k-23}^i + b^i\beta\gamma\epsilon_{k-24}^i \\ d_{k-22}^i - \alpha d_{t-23}^i + b^i\gamma\epsilon_{k-22}^i + b^i\beta\gamma\epsilon_{k-23}^i \\ \vdots \\ d_{k-1} - \alpha d_{t-2}^i + b^i\gamma\epsilon_{k-1}^i + b^i\beta\gamma\epsilon_{k-2}^i \\ -\alpha d_{k-1}^i + b^i\gamma\epsilon_k^i + b^i\beta\gamma\epsilon_{k-1}^i \\ b^i\beta\gamma\epsilon_k^i \end{bmatrix} \tag{15}$$

x is usually called the state and F the state transition matrix. For a network, there are multiple junctions for which the demands need to be calibrated, and these demands may be spatially correlated. The total demand vector for all nodes at time step k, d_k, can be obtained form the total state vector for all nodes, x_k, so that

$$x_{k+1} = F x_k + G \epsilon_k \tag{16}$$

$$d_{k+1} = Z x_{k+1} \tag{17}$$

where F, G, and Z are diagonal in terms of individual F^i, G^i and Z^i, respectively.

$$F = \begin{bmatrix} F^1 & & & \\ & F^2 & & \\ & & \ddots & \\ & & & F^N \end{bmatrix} \tag{18}$$

6

$$G = \begin{bmatrix} G^1 & & & \\ & G^2 & & \\ & & \ddots & \\ & & & G^N \end{bmatrix} \tag{19}$$

$$Z = \begin{bmatrix} Z^1 & & & \\ & Z^2 & & \\ & & \ddots & \\ & & & Z^N \end{bmatrix} \tag{20}$$

where d_k is the $N \times 1$ demand vector, x_k is the $26N \times 1$ state vector, and N is the number of junctions of which the demands are to be modeled. As described in the water demand model, $\{\epsilon_k\}$ is a Gaussian white noise vector and

$$E[\epsilon_k \epsilon_l^T] = Q\delta_{kl} \tag{21}$$

where δ_{kl} is the Kronecker delta, which is 1 for $k = l$ and 0 otherwise. The $N \times N$ covariance matrix Q is assumed to be constant over time and $Q = cov(\epsilon_k, \epsilon_k)$. The i row j column element of the Q is the covariance between ϵ_k^i and ϵ_k^j: $Q(i,j) = cov(\epsilon_k^i, \epsilon_k^j)$, which equals σ^2 if $i = j$, and $a_i a_j \sigma^2$ otherwise.

The hydraulic head and flow rate measurements used for calibration, z_k, are a subset of the total hydraulic state vector of all flows and heads at step k, s_k, which is a function of the water demands and measurement errors:

$$z_k = M s_k + v_k \tag{22}$$

$$s_k = h(d_k) \tag{23}$$

where M is a (0-1) observation matrix, h models the nonlinear network hydraulics and $\{v_k\}$ is a gaussian white process like $\{\epsilon_k\}$, with

$$cov(v_k, v_l) = E[v_k v_l^T] = R_k \delta_{kl} \tag{24}$$

The hydraulic state vector s_k can also be written as function of the demand state vector x_k

$$s_k = g(x_k) \tag{25}$$

It is assumed that $\{\epsilon_k\}$ and $\{v_k\}$ are independent. It is also reasonable to assume measurement errors at different locations are independent so that R is a diagonal matrix of which the diagonal values are the variances of measurement errors.

The nonlinear function $g(x_k)$ is expanded by a Taylor series about prior estimates of the state vector, $x_{k/k-1}$, as

$$s_k = g(x_k) \approx g(x_{k/k-1}) + H_k(x_k - x_{k/k-1}) + \cdots \tag{26}$$

$$H_k = \left.\frac{\partial s}{\partial x}\right|_{x=x_{k/k-1}} = \left.\frac{\partial s}{\partial d}\right|_{d=d_{k/k-1}} \frac{\partial d}{\partial x} = \left.\frac{\partial s}{\partial d}\right|_{d=d_{k/k-1}} Z \tag{27}$$

And the sensitivity of measured hydraulic states with respect to water demand states can be calculated:

$$\left.\frac{\partial z}{\partial x}\right|_{x=x_{k/k-1}} = M H_k \tag{28}$$

If it is assumed that the higher order terms in Equation (26) are negligible, and that the white noise ϵ_k and v_k are gaussian and mutually independent, extended Kalman filtering (EKF) provides

7

the maximum likelihood estimate of demand over time. When the gaussian assumptions are not valid, EKF is still the optimal linear algorithm to minimize the sum of estimate variance, i.e. the trace value of the covariance matrix for the estimates (Anderson and Moore, 1979; Gelb, 1974).

The prediction part of EKF algorithm for the the prediction of state vector x is as follows

$$x_{k/k-1} = F x_{k-1/k-1} \tag{29}$$

$$P^x_{k/k-1} = F P^x_{k-1/k-1} F^T + GQG^T \tag{30}$$

The correction part of the EKF is:

$$x_{k/k} = x_{k/k-1} + K_k(z_k - Mg(x_{k/k-1})) = x_{k/k-1} + K_k(z_k - Mg(x_{k/k-1})) \tag{31}$$

$$P^x_{k/k} = (I - K_k H_k) P^x_{k/k-1} \tag{32}$$

The Kalman gain K is defined as:

$$K_k = P^x_{k/k-1}(MH_k)^T (MH_k P_{k/k-1}(MH_k)^T + R_k)^{-1} \tag{33}$$

In the above equations, $x_{k/k-1} =$ prediction or prior estimation of x_k, $x_{k/k} =$ posterior estimation of x_k, $P^x_{k/k-1} =$ prior estimation variance of x_k, and $P^x_{k/k} =$ posterior estimation variance of x_k. Equation (30) shows the demand prediction uncertainties are due to both estimation uncertainties at previous steps and forecast uncertainties of demand time series model. If the uncertainty in demand were zero at time step $k-1$, prediction uncertainties are simply caused by the uncertainties of the demand forecast model. Equation (31) shows the estimate is the linear combination of the prediction and the difference between the "actual" measurements and predicted measurements using predicted demands. With $x_{k/k}$ and $P^x_{k/k}$, it is straightforward to estimate the mean demand and its variance and covariance, according to Equation (17).

3   Sensitivity Analysis

To implement the predictor-corrector EKF algorithm for water demand estimation, we require evaluation of s and its first derivatives (see equations (23), (27), and (33)). This is complicated as the hydraulic network model is a set of nonlinear algebraic equations, and thus $h(d)$ is implicit. Thus the evaluation of $h(d)$ is done numerically, and system sensitivity analysis techniques are required to evaluate its derivatives. The algebraic equations describing the hydraulic network model can be written:

$$f(s, d) = 0 \tag{34}$$

where f is a vector of functions that define the implicit functions $h(d)$. In a water network, f includes flow continuity equations at all junctions and head-loss equations for all the links. Because of the nonlinearity of functions f, iterative Newton-Raphson method is used to get $h(d)$.

The sensitivities of hydraulic modeling results $h(d)$ to demand are obtained by solving sensitivity equations. Based on Equation (34),

$$\frac{\partial f}{\partial d} + \frac{\partial f}{\partial s}\frac{\partial s}{\partial d} = 0 \tag{35}$$

Therefore,

$$\frac{\partial f}{\partial d} + J\frac{\partial s}{\partial d} = 0 \tag{36}$$

8

Where J = Jacobian matrix of function vector f with respect to hydraulic state vector s. At each time step, matrix $\frac{\partial s}{\partial d}$ is solved after solving hydraulic state vector s. The algorithm is efficient since the same Jacobian, J, is calculated and factored during the standard Newton-Raphson method used to solve $f = 0$.

4   Simulation Study

Study Design

The network shown in Figure 3 is used for the simulation study of real time water demand estimation. There are 2 reservoirs (lake and river), 3 tanks and 92 junctions of which 59 have non zero demand. These reservoirs, tanks and junctions are connected through 117 pipes and 2 pumps. There are many factors that will affect the uncertainties of the demand estimates and at this
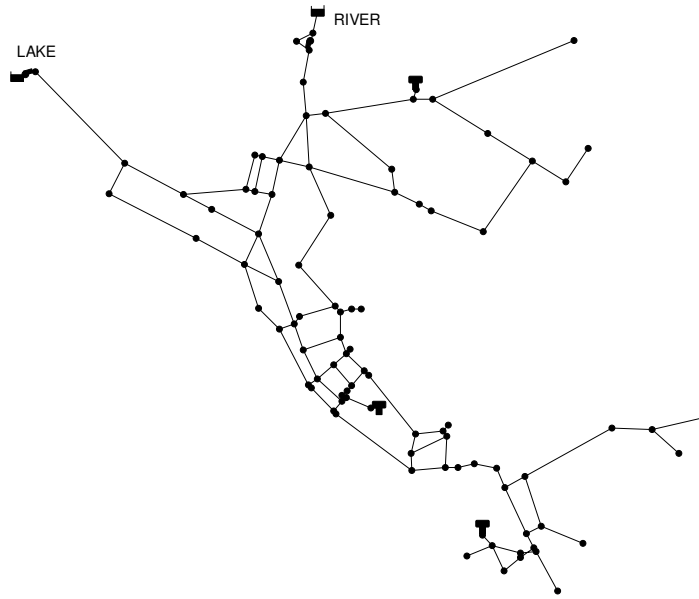

Figure 3: Example Network for Simulation Analysis of Demand Estimation

preliminary stage of research, the following parameters are analyzed:

- Demand forecast errors and spatial correlations of such errors. The variance and covariance of the demand forecast errors are calculated using Equation (8) and (9). For simplicity, it is assumed that $a_i = a$ and $b_i = b$, so the covariances between the demand forecast

errors are controlled with only two parameters: b and σ ($a = \sqrt{1 - b^2}$). A larger value of σ represents more uncertainty of the forecast model, while a larger value of b represents less spatial correlation.

- Measurement uncertainty. It is assumed that water head and flow rate measurement errors are proportional to the true values, and that measurement errors at different locations are independent (R is diagonal). The diagonal elements of R are measurement error variances and calculated using the following formula,

$$R_k(i, i) = c^2 s_k^i \tag{37}$$

where s_k^i is the ith element of hydraulic state vector s at step k and c is a constant which reflect the relative accuracy of the measurements.

- Measurement type and measurement locations. Both flow rate and water head measurements can be used to calibrate water demand. The number of measurement locations and types, demand forecast error characteristics, and measurement accuracy, will affect the water demand estimation results. Sampling network design for reliable demand estimation is not within the scope of this work. Instead, arbitrary sampling network designs will be used in simulation study, which are presumed to be sub-optimal.

- Initial conditions. Initial conditions are needed for any system state estimation problem. For water demand estimation, the demands and demand uncertainties at a 25 hour period are needed to build the initial state estimation vector $x_{25/25}$ and state covariance matrix $P_{25/25}$. The impact of initial conditions is not studied in this paper and it is assumed that initial conditions are known exactly without error, i.e. $P_{25/25} = 0$.

It is assumed that boundary (tank) heads of the network are measured without error. Since such heads are measured, they are not affected by demand uncertainty. Other network model parameters, such as pipe characteristics, are also assumed to be known without error.

Simulation Results

The performance of the estimation method is determined by the ratio of summed demand estimation error (absolute value) to total demand, averaged over time. The performance measure is defined as m_e

$$m_e = \frac{1}{N_t - 25} \sum_{k=26}^{N_t} \left(\frac{\sum_{i=1}^{N_d} |d_{k/k}^i - d_k^i|}{\sum_{i=1}^{N_d} d_k^i} \right) \tag{38}$$

where N_t = total number of simulation steps with step length of 1 hour and N_d = number of junctions of which demands are estimated. $d_{k/k}^i$ and d_k^i are estimated and true demand at junction i and step k, respectively. In this study $N_t = 72$ and $N_d = 59$.

In the first simulation study, in addition to the tank head measurements which are used to set network boundary conditions, water heads are measured at 20 junctions, of which 5 are connected to reservoirs and tanks and the others are selected randomly. Figure 5 shows the estimation performance when $\sigma = 0.05$ for different values of c and b. The larger the value of c, the larger the relative water head measurement uncertainty. Since a larger value of b means prediction uncertainties between different junctions are less correlated, it is not surprising that it is usually the case

10

that large values of b and c lead to poor estimation performance. In the second simulation study, flow rates are measured at 20 pipes, which include the pipes that are connected to the reservoirs and tanks so that total demand over time is measured indirectly. The results for different combinations of b and c are shown in Figure 6. For the same values of b and c, m_e is usually smaller for this design than that obtained in the previous simulation study. From these results alone, we can not draw the conclusion that measuring flow rates leads to more reliable demand estimation than measuring water heads, because the sampling designs are arbitrary and the relative accuracy of water head and flow rate measurements are unknown. Also as a practical matter, measurement costs should be factored in so that the sampling designs are compared on an equal cost basis. In the third simulation study, flow rates at 40 pipes are measured: 20 pipes are selected in the previous study and 20 more are added. Figure 7 shows that estimation errors are reduced after more pipe flow rates are measured. Demand prediction errors (predicted minus true demands) and estimation errors (estimated minus true demands) for one node with base water demand 39 gallon per minute (GPM) are shown for 2 cases: one with $b = 0.0$ and $c = 0.0001$; the other with $b = 1.0$ and $c = 0.1$. The results are shown in Figure 4 and it is obvious that small values of b and c lead to significant improvement of estimation over prediction. Under the same sampling effort, there is no obvious difference between prediction and estimation errors when the values of b and c are large.
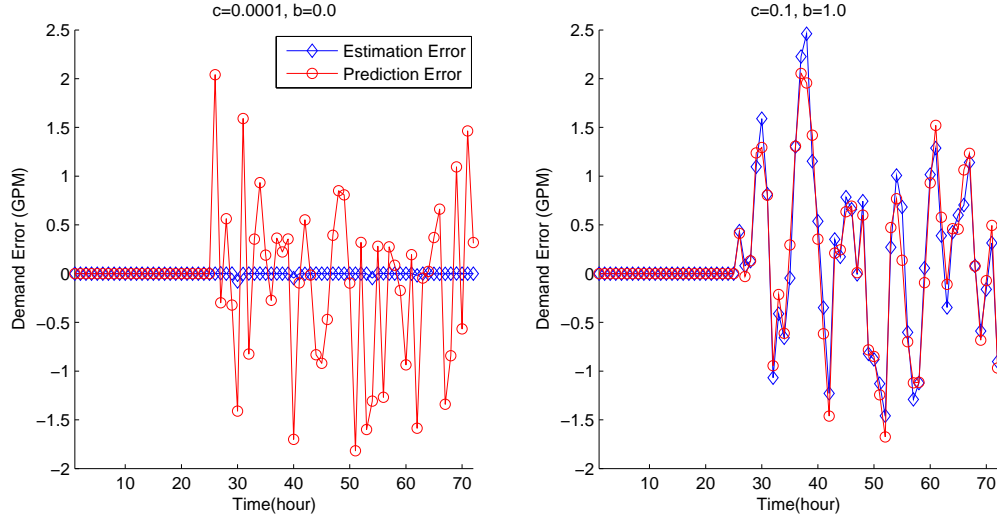

Figure 4: Water Demand Prediction and Estimation Errors for One Node when Flow Rates are Measured at 40 Pipes

In the previous studies, σ is always set to be 0.05 and the error variances of demand forecast model are set to be $(\sigma b^i)^2$. In the last simulation study, a more accurate demand forecast model is used by setting $\sigma = 0.02$. The same 40 pipes as in the third study are measured for flow rates. The results are shown in Figure 8. Comparing the results shown in Figure 7, we can see the estimation errors are significantly reduced when the demand forecast model is more accurate. Intuitively reducing the time step of the time series model may reduce the demand forecast uncertainty: it should be more certain to forecast demand 10 minutes rather than 1 hour later.

It should be noted that the results exhibit randomness since for each sampling design, they are the result of a single simulation realization. In order to get the statistical description of the estimation performance under a given sampling design , larger number of simulations are needed.
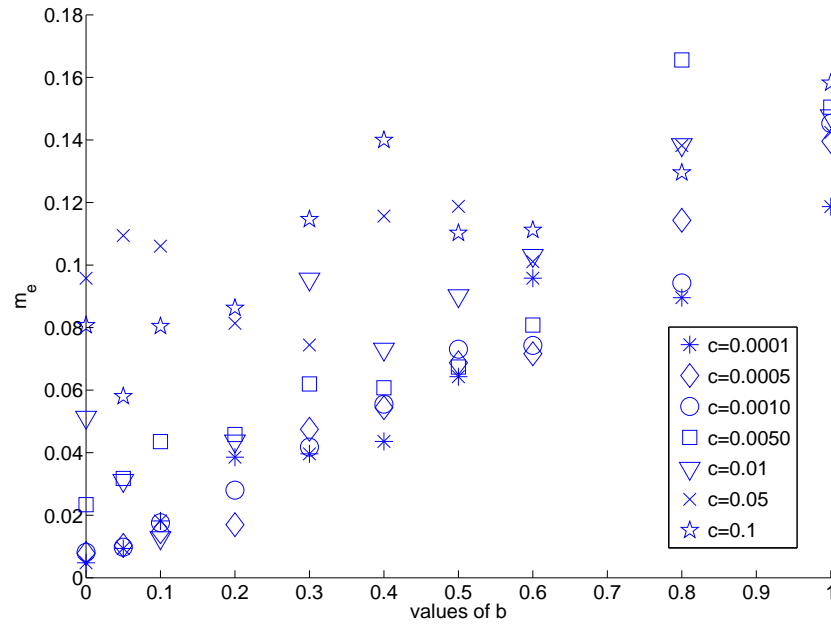
11

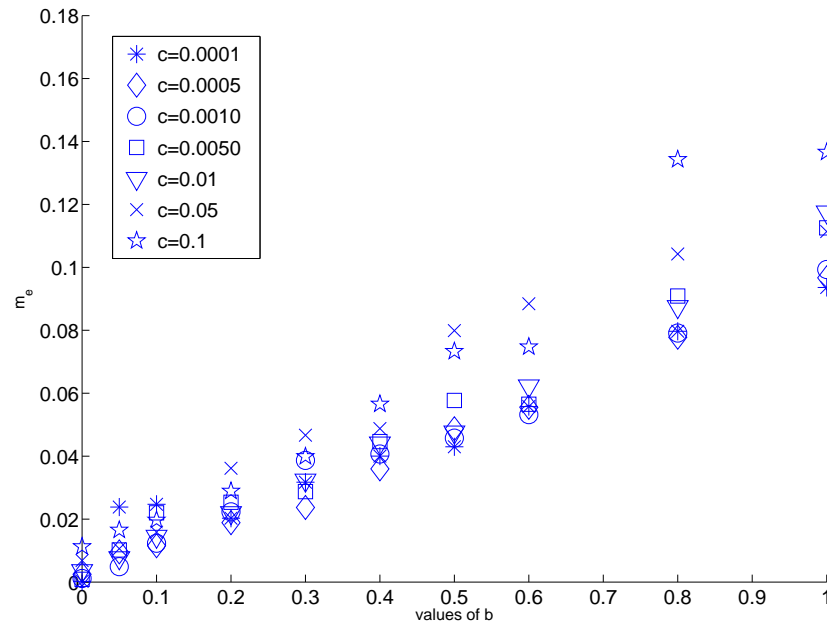Figure 5: Estimation Performance when $\sigma = 0.05$ and Water Heads are Measured at 20 Junctions

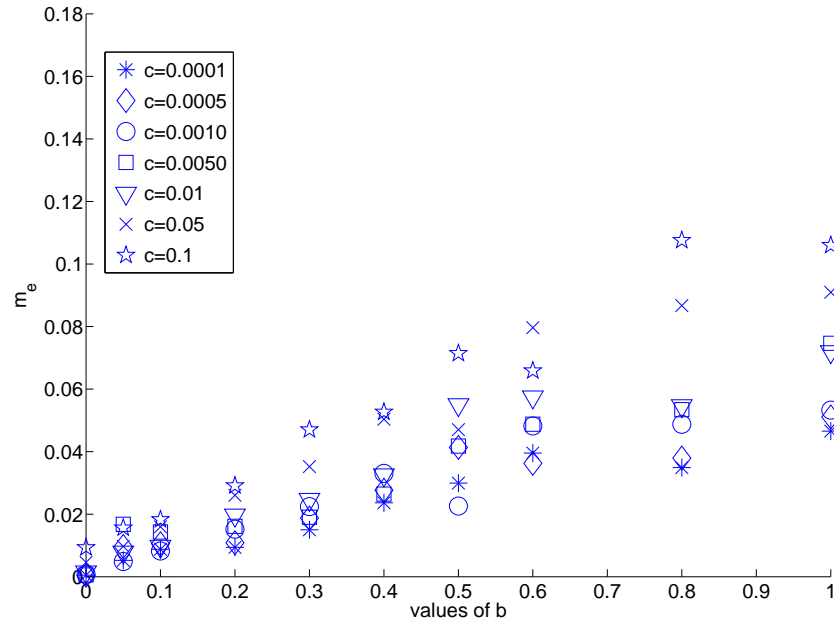Figure 6: Estimation Performance when $\sigma = 0.05$ and Flow Rates are Measured at 20 Pipes

Figure 7: Estimation Performance when $\sigma = 0.05$ and Flow Rates are Measured at 40 Pipes
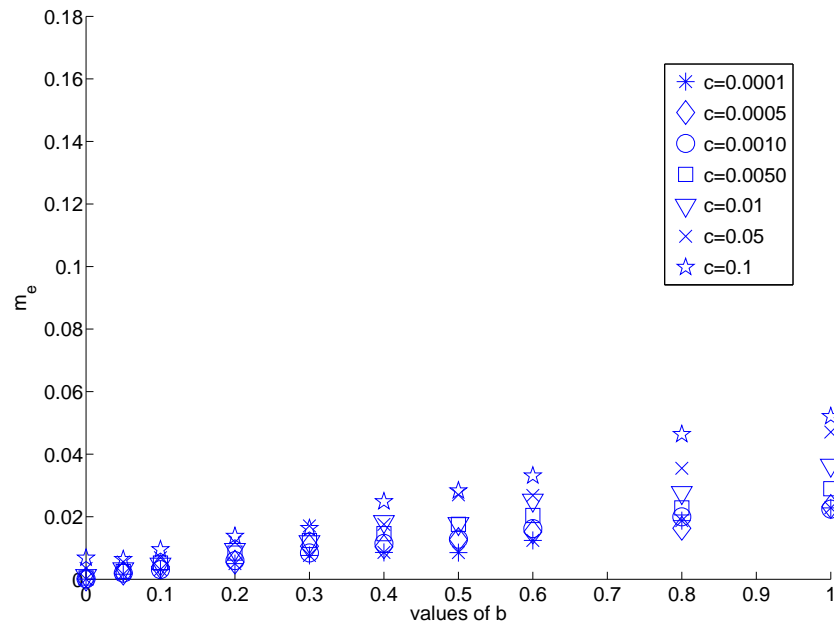

Figure 8: Estimation Performance when $\sigma = 0.02$ and Flow Rates are Measured at 40 Pipes

13

5   Conclusion

A predictor-corrector method is proposed to estimate water demands within distribution systems in real time. A demand time series model is used to predict the water demands based on the estimated demands at previous steps. The predictions are corrected using measured node water heads or pipe flow rates. A simple, hypothetical time series model is used and demand forecast errors are assumed to include both global and local randomness. Local randomness are independent and the more local demand uncertainty are dominated by local randomness, the less spatially correlated are the demands. Extended Kalman filter is used to implement the prediction and correction of water demands. The estimation performance depends on sampling design, measurement uncertainty, demand forecast error and the spatial correlations among the demand forecast errors. The demand forecast model is potentially the most critical factor for good demand estimation. Collecting demand data at multiple locations of the network system with high frequency would help answer many important research questions, such as the impact of time step size on forecast model uncertainty and spatial correlation among the demand forecast errors.

References

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering.* Prentice-Hall.

Box, G. P. and Jenkins, G. M. (1976). *Time Series Analysis: forecasting and control.* Holden-Day.

Gelb, A. (1974). *Applied Optimal Estimation.* The M.I.T. Press.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter.* Cambridge University Press.

Maybeck, P. S. (1979). *Stochastic Modelins, Estimation, and Control*, volume 1. Academic Press.

Williams, B. and Hoel, L. (2003). Modeling and forescating vehicle traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, 129(6):664–672.