I, Xuan Liu, hereby submit this original work as part of the requirements for the degree of Master of Science in Environmental Engineering.

It is entitled:
Real-Time Estimation of Water Network Demands

Student's name: Xuan Liu

This work and its defense approved by:

Committee chair: James Uber, PhD
Committee member: Lili Yeghiazarian, PhD
Committee member: Dominic Boccelli, PhD
Real-Time Estimation of Water Network Demands

A thesis submitted to the
Division of Research and Advanced Studies
of the University of Cincinnati
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in the Department of Civil and Environmental Engineering
of the College of Engineering
August 2011
By

Xuan Liu

B.S. University of Science and Technology of China, 2003

Committee Chair:
Dr. James Uber, Ph.D.
ABSTRACT

Accurate real-time estimation of water demands in water distribution system (WDS) plays an important role in accurate estimation of the system behavior, especially during emergency events. Traditionally, water demands are estimated based on population densities, customer billing records or census data. However, these methods are not appropriate for real-time modeling.

This study describes and tests a Predictor-Corrector methodology for real-time WDS water demand estimation. Current water demands are predicted by a reasonable time-series model. Improved estimation results are generated by correcting the prediction using real-time measurements (i.e., nodal heads or pipe flow rates). Extended Kalman filter algorithm is used to operate the prediction and correction process.

Experiments to test the performance of the algorithm are carried out. The approach is applied to a sample water distribution system comprised of 97 nodes and 119 pipes. The test results demonstrate the impacts of measurement accuracy, sampling design and demand model forecast error on water demand estimation.
Acknowledgments

This thesis would not have been completed without guidance and assistance from many individuals. First of all, my sincere thanks go to the Dr. James Uber, my committee chair for his inspiration, provision, and constructive criticism to my thesis. I am also especially grateful to other committee members, Dr Dominic Boccelli and Dr. Lilit Yeghiazarian, for their valuable advices and suggestions.

My thanks also go to my fellow graduate students for their kindly assistance when I had difficulties during the completion of my thesis.

Last, but not least, I would like to thank my family for the love and affection they have showered upon me over the years.
Table of Contents

CHAPTER 1 INTRODUCTION ...........................................................................................................1
  1.1 Problem Statement ...............................................................................................................1
  1.2 Research Objectives .............................................................................................................3
  1.3 Thesis Organization ..............................................................................................................4

CHAPTER 2 METHODOLOGY ......................................................................................................5
  2.1 Modeling Water Demand .......................................................................................................5
  2.2 Seasonal ARIMA model for water demand patterns .............................................................6
    2.2.1 Time Series Notation .......................................................................................................6
    2.2.2 ARMA Model ..................................................................................................................7
    2.2.3 Seasonal ARIMA Model ..................................................................................................8
  2.3 State Space Model ................................................................................................................9
  2.4 Predictor-Corrector Approach ............................................................................................10
  2.5 Extended Kalman Filtering ................................................................................................11

CHAPTER 3 SIMULATION STUDY .............................................................................................15
  3.1 Assumptions .......................................................................................................................16
  3.2 Synthetic Data Generation ..................................................................................................17
  3.3 Simulation Experiment Design ............................................................................................18
  3.4 Estimation Results ...............................................................................................................19
    3.4.1 Performance Metrics .......................................................................................................19
    3.4.2 Estimation with Adequate and Perfect Measurements ....................................................21
3.4.3 Estimation with Adequate but Imperfect Measurements..........................23
3.4.4 Estimation with Inadequate but Accurate Measurements..........................28
3.4.5 Estimation with Different Level of Model Uncertainty............................33

CHAPTER 4 CONCLUSIONS AND DISCUSSIONS...........................................36

APPENDIX A REFERENCES.................................................................38

APPENDIX B API OF EPANET RTX.......................................................40
List of Tables

Table 3.1 Performance metrics for evaluating forecasting accuracy...............................20
Table 3.2 Pattern Prediction and Estimation results with perfect measurement ...............21
Table 3.3 Pattern Estimation results with imperfect measurements...............................24
Table 3.4 Pattern Estimation results with accurate but inadequate measurements.........29
Table 3.5 Pattern Estimation results with different model uncertainty..........................34
**List of Figures**

Figure 3.1 Application network.................................................................16

Figure 3.2 Time series of Pattern Prediction APE and Estimation APE with Perfect Measurements.................................................................22

Figure 3.3 Statistics of estimated pattern APE with Perfect Measurements.........................23

Figure 3.4 Statistics of estimated pattern APE when $\sigma_f = 0$ and $\sigma_h = 1$................25

Figure 3.5 Statistics of estimated pattern APE when $\sigma_f = 0$ and $\sigma_h = 10$..............26

Figure 3.6 Statistics of estimated pattern APE when $\sigma_f = 1$ and $\sigma_h = 0$..............26

Figure 3.7 Statistics of estimated pattern APE when $\sigma_f = 10$ and $\sigma_h = 0$..............27

Figure 3.8 Statistics of estimated pattern APE when $\sigma_f = 10$ and $\sigma_h = 10$..............27

Figure 3.9 Statistics of estimated pattern APE when $N_f = 1$ and $N_h = 0$.................30

Figure 3.10 Statistics of estimated pattern APE when $N_f = 2$ and $N_h = 0$.................31

Figure 3.11 Statistics of estimated pattern APE when $N_f = 0$ and $N_h = 1$.................31

Figure 3.12 Statistics of estimated pattern APE when $N_f = 0$ and $N_h = 2$.................32

Figure 3.13 Statistics of estimated pattern APE when $N_f = 1$ and $N_h = 1$ ...............32

Figure 3.14 Statistics of estimated pattern APE when $N_f = 2$ and $N_h = 1$.................33

Figure 3.15 Statistics of estimated pattern APE when $\sigma = 0.01$.............................35

Figure 3.16 Statistics of estimated pattern APE when $\sigma = 0.1$.................................35
Chapter 1

Introduction

1.1 Problem Statement

In traditional water demand estimations, population growth and billing record are the primary decision variable. However, demand estimated based on population and customer billing are not proper for real-time optimization of pump and valve settings for efficient power management. Also, during emergency events like pollution events, main break, or significant fire, real-time updated water demand estimations are critical in order to accurately predict the system performance during those events and develop right control strategies.
With the fast growth of computer technology, computer simulation models of water networks have been widely used by water systems operators. Typically, water network models are calibrated off-line, using hydraulic data sampled during certain short period. In this off-line method, water network models are not updated frequently. Water demands and pipe roughness are adjusted until the model outputs and real observations are matched.

There have been a lot of off-line model calibration works. In 1989, Ormsbee and Lansey and Basnet already developed optimization methods to determine the uncertain system elements. In 1998, Artificial Neural Networks were used by Lingireddy and Ormsbee for calibration. As a recent trend, a lot of recent researches are focused on calibration methods using Genetic Algorithm approach following methods.

However, off-line calibration results can only represent the system hydraulics during the short period of the sampling procedure, and they could not represent the system conditions for the full range of operational conditions, especially during emergency events, such as a big fire accident. As water demands is the most uncertain and variable dynamic state of a water distribution network, the problem is more serious.
Nowadays, Supervisory Control and Data Acquisition (SCADA) system are widely installed by water systems operators. With the real-time hydraulic measurements provided by SCADA system, more accurate water demand estimation results could be reached.

In 2006, Shang and Uber developed EPANET-RTX (real-time extension for EPANET) toolkit. Extended Kalman Filter (EKF) algorithm was introduced in EPANET-RTX toolkit to estimate water demands in real-time. Given the assumption that the water network model is seasonal seasonal ARIMA (autoregressive integrated moving average) model, the water demands is predicted based on the estimated demands at previous steps. Then the real-time measurements of nodal water heads and pipe flow rates are used to correct the predicted water demands. The EPANET-RTX toolkit provided extension possibility for real-time water demand estimation on complex water network systems.

1.2 Research Objectives

This study aims to test the real-time water demands estimation techniques of the existing EPANET-RTX toolkit. It includes the following tasks:

1. Generate synthetic true water demands and true measurements for calibration.

2. Get a set of simulation results by using EPANET-RTX toolkit;

3. Results analysis
1.3 Thesis Organization

This study contains chapters that describe efforts to meet the objectives. Chapter 1 introduces the outline of the research. The review of previous studies on water demand model and water demand estimation is also included in Chapter 1. Chapter 2 describes the methodology which was used in the research. Chapter 3 presents the simulation experiment for testing the strategy. Analytical results and comparisons are also included in Chapter 3. Chapter 4 is a final discussion and scope for future work.
Chapter 2

Methodology

2.1 Modeling Water Demand

Traditionally, water demand at a specific water consumption node can be calculated as the product of base demand and pattern value.

\[ D_t = D_{base} \times P_t \]  \hspace{1cm} (2.1)

where,

- \( D_t \) is the actual nodal demand at time step \( t \);
- \( D_{base} \) is the base demand;
- \( P_t \) is the demand pattern value at time step \( t \).
The base demand of each consumption node is a deterministic value that usually set to be the average demand for water by the main category of consumer at the junction (Rossman 2000), calculated from meter readings and customer billing records.

2.2 Seasonal ARIMA model for water demand patterns

Since water demand data has seasonal trend, water demand pattern can be modeled as seasonal ARIMA (autoregressive integrated moving average) model (Box and Jenkins, 1976).

2.2.1 Time Series Notation

In order to understand seasonal ARIMA process, understanding of difference operator and backshift operator is prerequisite.

1. Difference operator $\nabla$:

Using this symbol, the first of an lagged time series $\{X_t\}$ can be defined as:

$$\nabla X_t = X_t - X_{t-1}$$

(2.2)

Higher-order seasonal differencing can be specified by using an integer superscript greater than one. For example, the second difference would be defined as:

$$\nabla^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$$

(2.3)
Differencing can also be applied at a seasonal lag. In this case, a subscript is employed to specify the length of the seasonal cycle. For example, for a series with a seasonal cycle of 12 intervals, the first seasonal difference would be defined as

$$\nabla_{12}X_t = X_t - X_{t-12}$$  \hspace{1cm} (2.4)

2. Backshift operator B:

In ARIMA model expressions it is more common to see the backward shift operator B used to define the required differencing.

$$BX_t = X_{t-1}$$  \hspace{1cm} (2.5)

$$B^jX_t = X_{t-j}$$  \hspace{1cm} (2.6)

Therefore, $\nabla = 1 - B$.

The first and second differences can be written as

$$\nabla X_t = (1 - B)X_t = X_{t-1}$$  \hspace{1cm} (2.7)

$$\nabla^2 X_t = (1 - B)^2X_t = (1 - 2B + B^2)X_t = X_t - 2X_{t-1} + X_{t-2}$$  \hspace{1cm} (2.8)

In general seasonal ARIMA expressions, seasonal differencing can denoted by the expression $(1 - B^s)^dX_t$, with $s$ denoting the length of the seasonal cycle and $d$ denoting the order of seasonal differencing.

### 2.2.2 ARMA Model

ARMA(p,q) time series model (Box and Jenkins, 1976) is defined as:

$$Y_t = c + \phi_1Y_{t-1} + \cdots + \phi_pY_{t-p} + \theta_1\epsilon_t + \cdots + \theta_q\epsilon_{t-q}$$  \hspace{1cm} (2.9)

Where,
\( c \) is a constant

\( p = \) autoregressive polynomial order

\( q = \) moving average polynomial order

\( \phi_i = \) autoregressive coefficients, \( i = 1, \ldots, p \)

\( \theta_j = \) moving average coefficients, \( j = 1, \ldots, q \)

\( \epsilon_t = \) a sequence of \( N(0, \sigma^2) \) random variables, \( \text{cov}(\epsilon_t, \epsilon_k) = 0 \) when \( t \neq k \)

### 2.2.3 Seasonal ARIMA Model

Seasonal ARIMA (p,d,q)(P,D,Q)s model is a specific form of ARMA(p,q) model. A time series \( \{X_t\} \) is a seasonal ARIMA (p,d,q)(P,D,Q)s process if the differenced series \( Y_t = (1 - B)^d (1 - B^s) X_t \) is a stationary ARMA process defined by the expression below:

\[
\phi_p(B)\Phi_p(B^s)Y_t = \mu + \theta_q(B)\Theta_q(B^s)\epsilon_t \tag{2.10}
\]

Where, \( B = \) backshift operator

\( \mu = \) the mean value of the time series\( \{Y_t\} \)

\( \phi_p(B) = 1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p \)

\( \theta_q(B) = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q \)

\( \Phi_p(B^s) = 1 - \Phi_1B^s - \Phi_2B^{2s} - \cdots - \Phi_pB^{ps} \)

\( \Theta_q(B^s) = 1 - \Theta_1B^s - \Theta_2B^{2s} - \cdots - \Theta_qB^{qs} \)

\( \epsilon_t = \) a sequence of \( N(0, \sigma^2) \) random variables, \( \text{cov}(\epsilon_t, \epsilon_k) = 0 \) when \( t \neq k \)
The parameters $p$ and $P$ represent the nonseasonal and seasonal autoregressive polynomial order, respectively; the parameters $q$ and $Q$ represent the nonseasonal and seasonal moving average polynomial order, respectively; the parameter $d$ represents the order of normal differencing, and the parameter $D$ represents the order of seasonal differencing; and $s$ is the length of the seasonal cycle.

In practice, to build a proper seasonal ARIMA model for water demand, the orders $p$, $d$, $q$, $P$, $D$, $Q$, and $s$ can be identified by analyzing the known water demand data. After the model order is fixed, the parameters $\phi_i, \varphi_i, \Phi_i, \Theta_i, \sigma^2$ can be estimated by Regularized Least Squares and Gauss-Newton Method.

### 2.3 State Space Model

Seasonal ARIMA models can be readily expressed in state space form, thereby allowing adaptive Kalman filtering techniques to be employed to provide a self-tuning forecast model.

One state space format of Equation (2.9) can be written as:

$$Y_t = Z X_t$$ \hspace{1cm} (2.11)

$$X_t = A X_{t-1} + cB + G\epsilon_t$$ \hspace{1cm} (2.12)

Where,

$$Z = [1, 0, \cdots, 0]$$ \hspace{1cm} (2.13)

$$B = [1, 0, \cdots, 0]^T$$ \hspace{1cm} (2.14)
A is the state transition matrix,

\[
A = \begin{bmatrix}
\phi_1 & 1 & 0 & \ldots & 0 \\
\phi_2 & 0 & 1 & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & 0 \\
\vdots & 0 & \cdots & 0 & 1 \\
\phi_m & 0 & \cdots & \cdots & 0
\end{bmatrix}
\]  \hspace{1cm} (2.15)

G is the white noise matrix,

\[
G = \begin{bmatrix}
\theta_1 \\
\theta_2 \\
\vdots \\
\theta_m
\end{bmatrix}
\]  \hspace{1cm} (2.16)

X is the state vector, the size of the state vector X is m and \( m = \max(p, q) \);

\[
\phi_k = 0 \text{ when } m \geq k > p \\
\theta_k = 0 \text{ when } m \geq k > q
\]

The conversion between \( Y_t \) and elements in the state vector \( X_t \) is:

\[
X_t[k] = \begin{cases} 
Y_t & k = 1 \\
\sum_{i=k}^{m} \phi_i Y_{t-i+k-1} + \sum_{i=k}^{m} \theta_i \epsilon_{t-i+k} & m \geq k > 1
\end{cases}
\]  \hspace{1cm} (2.17)

Once the seasonal ARIMA model was put into state space form, Kalman Filter can be applied as one predictor-corrector algorithm.

### 2.4 Predictor-Corrector Approach

The Predictor-corrector (Maybeck, 1979) approach is a state estimation algorithm which applied to water demand estimation in RTX toolkit. First, predictions of the water demands at the next time step are made based on the water demand estimation at previous time step. Also, predicted field measurements such as flow rates and nodal hydraulic heads can be calculated with the predicted water demands.
Field measurements are assumed to be either junction hydraulic heads or pipe flow rates.

\[ s_t = f(d_t) \]  \hspace{1cm} (2.18)

Where, \( s_t \) = vector of all flows and heads at step \( t \)
\( d_t \) = vector of estimated demands at time step \( t \)
\( f(d_t) \) is the implicit hydraulic function between \( s_t \) and \( d_t \).

Actually we cannot get all flows and heads measurements, therefore, the measurements used for calibration are subset of all flows and heads.

\[ m_t = M s_t + \nu_t \]  \hspace{1cm} (2.19)

Where, \( m_t \) = vector of hydraulic head and flow rate measurements used for calibration at step \( t \);
\( \nu_t \) = vector of demands and measurement errors at step \( t \).

After the measurements are taken at the next step, the differences between the measured and predicted field measurements are used to correct the prediction of the demands.

2.5 Extended Kalman Filtering

As a well known method of state estimation in linear systems, Kalman Filter (KF) is described and developed by Kalman (1960).
All model calculations and measurements contain uncertainty. In order to reduce the estimated uncertainty, the Kalman Filter first predicts a calculated value based on the system model. It then updates the predicted value by computing a weighted average of predicted value and measured value. Bigger weight is given to the value with less uncertainty. The result of Kalman Filter method is a new estimate of the true value, which lies in between predicted and measured value, and tends to be closer to the true value because the weighted average has a better estimated accuracy than either of the values that went into the weighted average.

The Kalman filter model assumes the true state at time k is connected to the state at (k – 1) as:

\[
x_k = F_k x_{k-1} + w_k
\]  

(2.20)

Where, \( F_k \) is the state transition model which is applied to the previous state \( x_{k-1} \).

\( w_k \) is the process noise which is assumed to be drawn from a zero mean normal distribution with covariance \( Q_k \), i.e., \( w_k \sim N(0, Q_k) \).

Another important component in KF method is the measurement equation, which transforms the state vector, \( x_k \), into a measurement vector, \( z_k \), as below:

\[
z_k = H_k x_k + v_k
\]  

(2.21)

Where, \( H_k \) is the observation model which maps the true state space into the observed space; \( v_k \) is the observation noise which is assumed to be zero mean Gaussian white
noise with covariance $R_k$, i.e., $v_k \sim N(0, R_k)$.

The initial state and the noise vectors at each step $\{x_k, w_1, ..., w_k, v_1, ..., v_k\}$ are all assumed to be mutually independent.

The state of the filter is represented by two variables:

1. $\bar{x}_{k|k}$, the a posteriori state estimate at time $k$ given observations up to and including at time $k$;
2. $P_{k|k}$, the a posteriori error covariance matrix (a measure of the estimated accuracy of the state estimate).

The prediction part of KF algorithm for the prediction of state vector $x$ is as follows:

$$\bar{x}_{k|k-1} = F_k \bar{x}_{k-1|k-1} + B_k w_k$$  \hspace{1cm} (2.22)

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k$$  \hspace{1cm} (2.23)

The correction part of the KF is:

$$\bar{x}_{k|k} = \bar{x}_{k|k-1} + K_k \left(z_k - H_k \bar{x}_{k|k-1}\right)$$  \hspace{1cm} (2.24)

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}$$  \hspace{1cm} (2.25)

The Kalman gain $K$ is defined as:

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1}$$  \hspace{1cm} (2.26)

The above calculations were executed in sequence at every time step.

However, the KF method is limited to linearity and Gaussian. The transition matrix for water demand may not be linear when using seasonal ARIMAL model. For the nonlinear systems, Extended Kalman Filter will be used.
Extended Kalman Filter (EKF) (Harvey, 1989) is the modification of Kalman Filter for nonlinear space-states. EKF model uses a first order linear approximation to simplify the procedures, and that is suboptimal solution. It is a predictor-corrector algorithm providing the maximum likelihood estimate of demand over time.

Eqs. (6) and (8) will be rewritten arbitrarily as below:

\[
\bar{x}_{t/t} = \bar{A}\bar{x}_{t-1/t-1} + \bar{B}\bar{c} \tag{2.27}
\]

\[
P_{t/t-1}^\bar{x} = \bar{A}P_{t-1/t-1}^\bar{x}A^T + \bar{GQG}^T \tag{2.28}
\]

\[
\bar{x}_{t/t} = \bar{x}_{t/t-1} + K_t \left( z_t - M\bar{g} \left( \bar{x}_{t/t-1} \right) \right) \tag{2.29}
\]

\[
P_{t/t}^\bar{x} = (I - K_t M\bar{h}) P_{t/t-1}^\bar{x} \tag{2.30}
\]

\[
K_t = P_{t/t-1}^\bar{x} (M\bar{h}_t) (M\bar{h}_t P_{t/t-1}^\bar{x} (M\bar{h}_t)^T + R_t)^{-1} \tag{2.31}
\]
Chapter 3

Simulation Study

The water demand estimation studies were applied to the network shown in Figure 3.1. The system consists of 1 lake, 1 river, 3 tanks and 92 junctions. Since lake, river and tanks were all considered as network nodes, there are totally 97 nodes for demand estimation. These nodes are connected by 117 pipes and 2 pumps.

The original EPANET .net file was included in EPANET2.0 software package, which can be downloaded at http://www.epa.gov/nrmrl/wswrd/dw/epanet.html.
3.1 Assumptions

Some assumptions are made for this study:

1. Water demands are assumed to be the only uncertain hydraulic model input;

2. Pipe roughness coefficients are assumed to be known and certain during the simulation process;

3. Tank levels are assumed to be measured without error and set as the boundary conditions of the system model at every time steps;

4. Both nodal head and pipe flow rate can be used as measurements for demand calibration. Measurements errors are assumed to be independent and normally
distributed with zero mean.

5. There is only one pattern model in the water distribution network.

6. Initial conditions of the pattern are assumed to be known exactly without error.

3.2 Synthetic Data Generation

Actual nodal demand in a given time period is considered as the product of base demand and pattern value. Base demands are contained in the network input file for use. In this study, seasonal ARIMA(0,0,1)(0,1,1)24 model is used as water demand pattern model.

Assume each simulation process will last 10 days, and the time step is one hour. Thus, the total number of time steps is 240.

To assess the performance of proposed algorithms, true demand data sets and true measurement data sets are synthetically generated by following steps:

(1) First, Seasonal ARIMA time series is calculated as water demand pattern;

(2) Assign the demand pattern to each node of the network; Thus, the true nodal demand is the product of base demand and pattern value;

(3) Run EPANET2.0 using generated true demands to get true node head and true pipe flows as measurements.
3.3 Simulation Experiment Design

There are a lot of factors which may affect the uncertainties of the demand estimation. In this study, measurement accuracy, number of measurement, measurement type and model accuracy are analyzed.

(1) Perform demand estimation with perfect measurement (all node heads and all pipe flows are known and accurate).

(2) Add white noises for each measurement to make measurement data sets containing errors. Assuming measurements errors are normally distributed with a mean of zero and standard deviation $\sigma_f$ (for flow rate measurements) or $\sigma_h$ (for head measurements). Perform simulations with different values of $\sigma_f$ and $\sigma_h$, so to test the effect of measurement errors.

(3) Consider the estimation under insufficient measurements. Set the number of measurement $N_s$ as 1, 2, 6, and 20. For each $N_s$, run a large number of simulations. Observe the statistic result, to figure out whether the number of measurements and the sampling location of measurements will affect the simulation significant or not.

(4) The variance of white noises process in SARIMA model represents model forecast error. Do simulation with different values of white noise variance, to test the effect of model accuracy.
3.4 Simulation Results

3.4.1 Performance Metrics

The forecasting accuracy can be evaluated by several performance metrics (Table 3.1):

1. Mean Absolute Error (MAE): MAE is a frequently-used measure of the differences between forecasted and true values; it ranges from 0 to infinity, with 0 corresponding to ideal fit:

\[
MAE = \frac{1}{N} \sum_{t=1}^{N} |f_t - a_t|
\]  

(3.1)

Where, \(f_t\) is the forecasted value at time \(t\);

\(a_t\) is the real value at time \(t\);

\(N\) is the total number of time steps;

2. Mean Absolute Percentage Error (MAPE): MAPE is similar to MAE, but it cannot possibly be computed when true value is zero; it ranges from 0 to infinity, with 0 corresponding to ideal fit:

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{f_t - a_t}{a_t} \right|
\]  

(3.2)

Where, \(f_t\) is the forecasted value at time \(t\);

\(a_t\) is the real value at time \(t\);

\(N\) is the total number of time steps;
3. Root Mean Squared Error (RMSE): RMSE is often used as an estimate of standard deviation of forecasted values; it ranges from 0 to infinity, with 0 corresponding to ideal fit:

\[ RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (f_t - a_t)^2} \]  

(3.3)

Where, \( f_t \) is the forecasted value at time \( t \);
\( a_t \) is the real value at time \( t \);
\( N \) is the total number of time steps;

4. Root Relative Squared Error (RRSE): RRSE value is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values; it ranges from 0 to infinity, with 0 corresponding to ideal fit:

\[ RRSE = \sqrt{\frac{\sum_{t=1}^{N} (f_t - a_t)^2}{\sum_{t=1}^{N} (a_t - a_v)^2}} \]  

(3.4)

Where, \( f_t \) is the forecasted value at time \( t \);
\( a_t \) is the real value at time \( t \);
\( a_v \) is the mean of the real data = \( \frac{1}{N} \sum_{t=1}^{N} a_t \)
\( N \) is the total number of time steps;

<table>
<thead>
<tr>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
<th>RRSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{1}{N} \sum_{t=1}^{N}</td>
<td>f_t - a_t</td>
<td>)</td>
<td>( \frac{1}{N} \sum_{t=1}^{N} \frac{</td>
</tr>
</tbody>
</table>

Table 3.1: Performance metrics for evaluating forecasting accuracy
### 3.4.2 Estimation with Adequate and Perfect Measurements

The water demand pattern model is set as Seasonal ARIMA(0,0,1)(0,1,1)_{24} model. The white noise process of demand pattern model has zero mean and 0.05 standard deviation. In addition to the tank head measurements which are used as network boundary conditions, water heads are measured at 92 junctions, and flow rates are measured at 117 pipes. Perfect measurements means all node heads and all pipe flow rates are available and accurate without error.

**Single Simulation Realization:**

Figure 3.2 displays the time series of demand pattern prediction and estimation errors with perfect measurement. The performance metrics of this estimation scenario were summarized in Table 3.2.

It can be seen that estimation results were improved significantly compared to prediction results. The mean absolute error between estimated result and true demand value is close to zero, indicating the estimation fits well with true data.

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
<th>RRSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>0.0512</td>
<td>0.0953</td>
<td>0.0619</td>
<td>0.3459</td>
</tr>
<tr>
<td>Estimation</td>
<td>8.8426e-007</td>
<td>1.6640e-006</td>
<td>1.4954e-006</td>
<td>8.3498e-006</td>
</tr>
</tbody>
</table>

Table 3.2: Pattern Prediction and Estimation results with perfect measurement
Multi-simulation Realization:

In order to get the statistical description of the estimation performance with perfect measurement, 500 simulations were realized. The mean, the 5-percentile and 95-percentile values of estimated APE over time were shown in Figure 3.3. The means of estimated APE are around 2e-6 while the peak of error is also below 2e-5, indicating the accuracy of estimation result.
3.4.3 Estimation with Adequate but Imperfect Measurements

Perfect measurements are impractical. We must consider the impact of measurement uncertainty. Therefore, white noises were added to each measurement to make measurement data sets containing errors. Assuming measurements errors are normally distributed with a mean of zero and standard deviation $\sigma_f$ (for flow rate measurements) or $\sigma_h$ (for head measurements). The larger the value of $\sigma_f$ and $\sigma_h$, the larger the measurement uncertainty is. Simulations were performed with different values of $\sigma_f$ (gallon per minute) and $\sigma_h$ (foot).
Single Simulation Realization:

The performance metrics are tabulated in Table 3.3 for different levels of measurement uncertainty.

<table>
<thead>
<tr>
<th>$\sigma_f$ (GPM)</th>
<th>$\sigma_h$ (foot)</th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
<th>RRSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.0023</td>
<td>0.0047</td>
<td>0.0038</td>
<td>0.0038</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0025</td>
<td>0.0047</td>
<td>0.0046</td>
<td>0.0255</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>0.0030</td>
<td>0.0064</td>
<td>0.0051</td>
<td>0.0276</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0.0032</td>
<td>0.0072</td>
<td>0.0055</td>
<td>0.0276</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.0063</td>
<td>0.0109</td>
<td>0.0077</td>
<td>0.0440</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>0.0105</td>
<td>0.0185</td>
<td>0.0128</td>
<td>0.0696</td>
</tr>
<tr>
<td>0</td>
<td>0.01</td>
<td>6.2963e-007</td>
<td>1.0844e-006</td>
<td>1.0274e-006</td>
<td>5.8959e-006</td>
</tr>
<tr>
<td>0</td>
<td>0.1</td>
<td>6.9907e-007</td>
<td>1.1591e-006</td>
<td>1.1446e-006</td>
<td>6.7223e-006</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>7.6852e-007</td>
<td>1.3194e-006</td>
<td>1.1547e-006</td>
<td>6.2514e-006</td>
</tr>
<tr>
<td>0</td>
<td>10</td>
<td>7.2685e-007</td>
<td>1.2139e-006</td>
<td>1.1076e-006</td>
<td>6.0612e-006</td>
</tr>
</tbody>
</table>

Table 3.3: Pattern Estimation results with imperfect measurements

The uncertainty in either flow rate measurements or nodal head measurements significantly reduces the demand estimation accuracy. Note that when flow measurements are accurate ($\sigma_f = 0$), both accurate and less accurate head measurements are able to provide good estimation performance (errors are close to 0). However, even when head measurements are accurate, small variance of flow
measurements would lead to much higher errors. We may draw the conclusion that water demand is more sensitive to variations in pipe flow measurements.

**Multi-simulation Realization:**

In order to get the statistical description of the estimation performance with imperfect measurement, 5 scenarios for different levels of measurement uncertainty were set. For each scenario, 500 simulations were done to observe the range of errors. The Absolute Percentage Errors (APE) is set as statistic sample. The mean, the 5-percentile and 95-percentile values of estimated APE over time are plotted and shown in Figure 3.4 to Figure 3.8.

![APE statistics](image)

*Figure 3.4: Statistics of estimated pattern APE when $\sigma_f = 0$ and $\sigma_h = 1$*
Figure 3.5: Statistics of estimated pattern APE when $\sigma_f = 0$ and $\sigma_h = 10$

Figure 3.6: Statistics of estimated pattern APE when $\sigma_f = 1$ and $\sigma_h = 0$
Figure 3.7: Statistics of estimated pattern APE when \( \sigma_f = 10 \) and \( \sigma_h = 0 \)

Figure 3.8: Statistics of estimated pattern APE when \( \sigma_f = 10 \) and \( \sigma_h = 10 \)
From the plots above, we can confirm the conclusion that water demand is more sensitive to variations in pipe flow measurements.

### 3.4.4 Estimation with Inadequate but Accurate Measurements

From the above discussion, it can be seen that when measurements is adequate, estimation results are reliable even with measurement errors. However, in practice, few measurements are usually available. Thus, it is necessary to test the algorithm with insufficient measurements. In this case, assume measurements are accurate without error.

**Single Simulation Realization:**

The estimation results for different combination of $N_f$ and $N_h$ are shown in Table 3.4. Here, $N_f$ is the number of pipe flow measurements, and $N_h$ is the number of nodal head measurements. The locations of measurements are selected randomly.

The estimation performance is significantly improved when more flow measurements are added. The demand is less sensitive to number of head measurements because more head measurements did not lead to distinct improvement of estimation accuracy when flow measurements number is certain.
<table>
<thead>
<tr>
<th>$N_f$</th>
<th>$N_h$</th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
<th>RRSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>4.1667e-008</td>
<td>6.4487e-008</td>
<td>2.0412e-007</td>
<td>1.1287e-006</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1.3833e-005</td>
<td>2.3821e-005</td>
<td>2.2222e-005</td>
<td>1.2250e-004</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1.8333e-006</td>
<td>3.0902e-006</td>
<td>2.8932e-006</td>
<td>1.6145e-005</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>1.1204e-006</td>
<td>1.8617e-006</td>
<td>1.8534e-006</td>
<td>1.0605e-005</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.0027</td>
<td>0.0053</td>
<td>0.0051</td>
<td>0.0254</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0.0027</td>
<td>0.0049</td>
<td>0.0047</td>
<td>0.0283</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
<td>0.0033</td>
<td>0.0057</td>
<td>0.0063</td>
<td>0.0390</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>0.0021</td>
<td>0.0037</td>
<td>0.0040</td>
<td>0.0220</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>6.9588e-005</td>
<td>1.1763e-004</td>
<td>1.2166e-004</td>
<td>6.7216e-004</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4.2144e-005</td>
<td>7.0316e-005</td>
<td>7.7660e-005</td>
<td>4.0039e-004</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>4.7685e-007</td>
<td>8.4303e-007</td>
<td>9.0010e-007</td>
<td>4.5445e-006</td>
</tr>
</tbody>
</table>

Table 3.4: Pattern Estimation results with accurate but inadequate measurements

**Multi-simulation Realization:**

Within a WDS, there are numerous locations where nodal head and pipe flow can be measured for the purpose of water demand calibration. Where to take measurements is an issue of sampling design, which may affect the simulation results significantly. Therefore, multiple simulations are needed to test the effect of sampling design.
For each scenario, 500 simulations were done to observe the range of errors. The Absolute Percentage Errors (APE) of each scenario is set as statistic sample. The max, the 5-percentile and 95-percentile values of estimated APE for different combination of $N_f$ and $N_h$ over time are plotted in Figure 3.9 to Figure 3.14.

![Figure 3.9: Statistics of estimated pattern APE when $N_f = 1$ and $N_h = 0$](image_url)
Figure 3.10: Statistics of estimated pattern APE when $N_f = 2$ and $N_h = 0$

Figure 3.11: Statistics of estimated pattern APE when $N_f = 0$ and $N_h = 1$
Figure 3.12: Statistics of estimated pattern APE when $N_f = 0$ and $N_h = 2$

Figure 3.13: Statistics of estimated pattern APE when $N_f = 1$ and $N_h = 1$
Figure 3.14: Statistics of estimated pattern APE when $N_f = 2$ and $N_h = 1$

It can be observed that the flow measurement provides significantly better estimations for the demand. When few measurements are available, measuring flow rates leads to more reliable estimation.

3.4.5 Estimation with Different Level of Model Uncertainty

The white noises process variance $\sigma$ in SARIMA model represents model forecast error. In the previous simulations, $\sigma$ is always set as 0.05. Changing the value of $\sigma$ will rise or reduce the model accuracy, which will lead to lower or higher estimation errors.
**Single Simulation Realization:**

Assuming the measurements are perfect, estimation results with different values of $\sigma$ are shown in Table 3.5. It is not surprising that more accurate demand model will reduce the demand estimation inaccuracy.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
<th>RRSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.2963e-007</td>
<td>2.1632e-007</td>
<td>3.6004e-007</td>
<td>2.1315e-006</td>
</tr>
<tr>
<td>0.05</td>
<td>8.8426e-007</td>
<td>1.6640e-006</td>
<td>1.4954e-006</td>
<td>8.3498e-006</td>
</tr>
<tr>
<td>0.1</td>
<td>5.1481e-006</td>
<td>1.0819e-005</td>
<td>8.7934e-006</td>
<td>4.8238e-005</td>
</tr>
</tbody>
</table>

Table 3.5: Pattern Estimation results with different model uncertainty

**Multi-simulation Realization:**

500 simulations were realized for $\sigma = 0.01$ and $\sigma = 0.1$. The mean, the 5-percentile and 95-percentile values of estimated APE over time were shown in Figure 3.15 and Figure 3.16. It can be seen that even when the water demand model is relatively inaccurate, the Kalman filtering algorithm can still achieve good estimation performance.
Figure 3.15: Statistics of estimated pattern APE when $\sigma = 0.01$

Figure 3.16: Statistics of estimated pattern APE when $\sigma = 0.1$
Chapter 4

Conclusions and Discussions

The objective is to test the EKF Algorithm which was applied in EPANET-RTX toolkit. This study consists of two parts: synthetic data generation part and simulation case study part.

For this purpose, a seasonal ARIMA time series model is utilized to simulate water demands in a sample distribution network. The “true” time series of water demand thus generated, which is then used to generate synthetic pipe flow rates and hydraulic nodal heads as true measurements for calibration. With these synthetic “true” data, the EKF Algorithm was applied to several case studies. The results show that the estimation performance depends on measurement uncertainty and demand model forecast error.
Due to the limitation of data sources, there are some constraints in this study. Future work may include the tasks below:

1. Multi-pattern condition and Pattern categorization with error

2. Develop a sampling design approach to select measurement locations that are more likely to produce robust estimates of water demands.

3. Test under real water network with SCADA data.
Appendix A: References


Appendix B: API OF EPANET RTX
(Feng Shang and James Uber 2007)

This appendix describes the application programming interface (API) of EPANET RTX which is used in the simulation study.

The EPANET-RTX toolkit is a library of functions that programmers can use to access the proposed water demand estimation algorithm. The description of each toolkit function using C/C++ syntax to represent argument variables and return types is as follows.

```c
int ENRTXopen(char *fname, int numpmodel, long step)
```

1. *Description*

   Opens EPANET model input file to get network information.

2. *Arguments*

   *fname*: EPANET input file name.

   *numpmodel*: number of patterns for demand modeling.

   *step*: time step for pattern modeling, in seconds.

3. *Returns*

   Returns an error code, or 0 if no error.
int ENRTXsetdemandmodel(int nindex, int pindex, double basedemand)

1. Description

Specifies demand model for individual water consumption node.

2. Arguments

nindex: EPANET model node index.

pindex: index of the pattern to model demand at node with index nindex.

basedemand: base demand to model demand at node with index nindex.

3. Returns

Returns an error code, or 0 if no error.

int ENRTXsetpmodel(int pindex, double c, int p, double * ar, int q, double *ma)

1. Description

Sets water demand pattern model.

2. Arguments

pindex: index of the pattern model.

c: constant in water demand pattern model as equation 2.

p: autoregressive order.

ar: vector of autoregressive coefficients, ar[i] = \phi_i in equation 2.

q: moving average order.

ma: vector of moving average coefficients, ma[i] = \mu_i in equation 2.

3. Returns

Returns an error code, or 0 if no error.
int ENRTXsetpnoise(int i, int j, double var)

1. Description

Sets covariance value for pattern model noise

2. Arguments

i: index of pattern model.

j: index of pattern model.

var: covariance value between white noises of pattern models $i$ and $j$, $\text{var} = Q_{ij}$ (see equation 21).

3. Returns

Returns an error code, or 0 if no error.

int ENRTXinitpmodel(int pindex, double *initp)

1. Description

Initializes demand pattern model

2. Arguments

pindex: index of pattern model.

initp: vector of initial pattern values for pattern $\text{pindex}$. Assuming $t$ is the starting time step for real time demand modeling, $\text{initp}[i] = \text{pattern value at time step } t-i$ ($i = 1, 2, ..., p$ and $p$ is the model autoregressive order).

3. Returns

Returns an error code, or 0 if no error.
int ENRTXinit()

1. Description

Converts the input pattern models into states space model and prepares for water demand real time estimation.

2. Arguments

3. Returns

Returns an error code, or 0 if no error.

int ENRTXpredict()

1. Description

Predicts the water demand pattern values for the next time step based on the current estimation or given initial values.

2. Arguments

3. Returns

Returns an error code, or 0 if no error.

4. Note

At the beginning of the real time modeling, the prediction is based on the initial values, assuming there is no uncertainty in the given initial values.

int ENRTXinputmeasurement(int type, int index, double val, double var)

1. Description

Provides measured or simulated hydraulic measurement.
2. Arguments

type: measurement type.

index: node or pipe index.

val: hydraulic head or flow rate value

var: variance of the measurement

3. Measurement type codes

ENRTX_HEAD 1 Hydraulic head at node

ENRTX_FLOW 2 Flow rate at pipe

4. Returns

Returns an error code, or 0 if no error.

int ENRTXcorrect()

1. Description

Corrects predicted water demand pattern values after hydraulic measurements input.

2. Arguments

3. Returns

Returns an error code, or 0 if no error.

int ENRTXgetpatvalue(int pindex, double *pvalue, double *pvar)

1. Description

Obtains predicted or corrected demand pattern value.
2. **Arguments**

   pindex: pattern index.

   pvalue: demand pattern value, predicted or corrected depending on whether
   the function is called after ENRTXpredict() or ENRTXcorrect().

   pvar: estimated variance of estimated demand pattern value.

3. **Returns**

   Returns an error code, or 0 if no error.

4. **Note**

   At the beginning of the real time modeling, the prediction is based on the
   initial values, assuming there is no uncertainty in the given initial values.

```c
int ENRTXgetdemand(int nindex, int type, double *value)
```

1. **Description**

   Obtains water demand value

2. **Arguments**

   nindex: node index.

   type: demand type.

   value: demand value.

3. **Demand type codes**

   ENRTX_PREDICT 1 Predicted water demand

   ENRTX_CORRECT 2 Corrected water demand

   ENRTX_BASE 3 nodal base demand
4. Returns

Returns an error code, or 0 if no error.

int ENRTXgetdemandvar(int nindex, double *dvar)

1. Description

Obtains estimated variance of the predicted or corrected demand value.

2. Arguments

nindex: node index.

dvar: estimated variance of estimated variance of the predicted or corrected demand, depending on whether the function is called after ENRTXpredict() or ENRTXcorrect().

3. Returns

Returns an error code, or 0 if no error.

int ENRTXclose()

1. Description

Closes the EPANET RTX and release the memory allocated.

2. Arguments

3. Returns

Returns an error code, or 0 if no error.
int ENRTXsetdemand(int nindex, double demand)

1. **Description**

   Sets nodal demand for simulation purpose.

2. **Arguments**

   nindex: node index.

   demand: demand at current time step.

3. **Returns**

   Returns an error code, or 0 if no error.

int ENRTXsimulate(long st, long *at)

1. **Description**

   Runs hydraulic simulation with demands set through ENRTXsetdemand().

2. **Arguments**

   st: expected simulation time in seconds.

   at: actual simulation time in seconds.

3. **Returns**

   Returns an error code, or 0 if no error.

int ENRTXgethead(int nindex, double *head)

1. **Description**

   Obtains simulated nodal head value.

2. **Arguments**
nindex: node index.

head: hydraulic head value.

3. Returns

Returns an error code, or 0 if no error.

int ENRTXgetflow(int pindex, double *flow)

1. Description

Obtains simulated flow rate at pipe.

2. Arguments

nindex: pipe index.

head: flow rate value.

3. Returns

Returns an error code, or 0 if no error.