Real-Time Water Quality Modeling with Ensemble Kalman Filter for State and Parameter Estimation in Water Distribution Networks

Rajakumar, Anjana G.; Mohan Kumar, M. S.; Amrutur, Bharadwaj; Kapelan, Zoran

DOI
10.1061/(ASCE)WR.1943-5452.0001118

Publication date
2019

Document Version
Accepted author manuscript

Published in
Journal of Water Resources Planning and Management

REAL-TIME WATER QUALITY MODELLING WITH ENSEMBLE KALMAN FILTER FOR STATE AND PARAMETER ESTIMATION IN WATER DISTRIBUTION NETWORKS

G R Anjana [1], M S Mohan Kumar[2*], Bharadwaj Amrutur[3] and Zoran Kapelan[4,5]

[1]Research Scholar, Department of Civil Engineering, Indian Institute of Science, Bangalore, India.

E-mail: anjanagr@iisc.ac.in

[2]PhD, Professor, Department of Civil Engineering, ICWaR, RBCCPS and IFCWS, Indian Institute of Science, Bangalore, India.

E-mail: msmk@iisc.ac.in

[3]PhD, Professor, Robert Bosch Center for Cyber-Physical Systems (RBCCPS) and Department of Electrical Communications Engineering, Indian Institute of Science, Bangalore, India.

E-mail: amrutur@iisc.ac.in

[4]PhD, Delft university of Technology, Faculty of Civil Engineering and Geosciences, Department of Water Management, Delft, The Netherlands.

E-mail: Z.Kapelan@tudelft.nl

[5]University of Exeter, College of Engineering, Mathematics and Physical Sciences, Exeter, United Kingdom.

E-mail: Z.Kapelan@exeter.ac.uk

*Corresponding author: M.S. Mohan Kumar (msmk@iisc.ac.in)

Abstract

This study presents a novel approach for real time water quality state (chlorine concentration) and reaction parameter estimation in Water Distribution Systems (WDS) using Ensemble Kalman Filter (EnKF) based data assimilation techniques. Two different types of EnKF based methods are used in this study: (a) Non-Iterative Restart-EnKF (NIR-EnKF) and Iterative Restart-EnKF (IR-EnKF). Use of these data assimilation frameworks for addressing key uncertainties in water quality models such as (i) uncertainty in the source or initial concentration of chlorine and (ii) uncertainty in wall reaction parameter, is studied. The effect of ensemble size, number and location of measurement nodes, measurement error and noise are also studied extensively in this work. The performance of the methodology proposed is tested on two different water networks: (i) Brushy Plains Network; (ii) and a big, city-wide WDS, Bangalore inflow network. The results of the simulation study show that, both NIR-EnKF and IR-EnKF methods are appropriate for dealing with uncertainty in source chlorine concentration, whereas IR-EnKF method performs better than NIR-EnKF method in case of reaction parameter uncertainty.

Introduction

Advancement in engineering and nanotechnology has resulted in the development of several sensors that can log online water quality data such as residual chlorine, pH, electrical conductivity, dissolved oxygen etc. in Water Distribution Systems (WDS) (Suresh et al. 2014). Installation of these online sensors help in safeguarding the WDS against accidental and intentional contamination (Hall et al.2007). But, deployment of these sensors at all the nodes of a WDS is not feasible, considering the cost that will be incurred in doing so. Hence, sensors are placed only at a few strategic locations in the WDS (Aral et al. 2009; Ostfeld et al. 2008; Hart and Murray2010, Simone et al. 2016). In such systems with limited sampling locations, data assimilation creates the best estimate of the system state at the non-measurement nodes. In this work, the main objective is to assimilate the real time chlorine concentration data from these sensors, in a water quality model, for estimating the water quality state and parameters of the WDS in real time.

In this work, water quality state refers to the chlorine concentration at all the nodes of the WDS. Traditional methods for nodal chlorine concentration estimation involve a well calibrated water quality simulation model. In literature, numerous alternate methods for state and parameter estimation methods are available. A few of them are, inverse modelling (Clark et al. 1993; Biswas et al.1993; Rossman et al. 1994; Munavalli and Kumar 2004 &2005), time series analysis (Rodriguez and Serodes 1998; Polycarpou et al. 2002; Bowden et al. 2006; Gibbs et al. 2006) and soft computational methods such as neural networks, genetic algorithm, machine learning etc. (Rodriguez and Serodes1998; Baxter et al. 1999, 2001; Serodes et al. 2001; Milot et al. 2002; Maier et al. 2004; Gibbs et al. 2006; Bowden et al. 2006; May et al. 2008; D D'Souza and Kumar 2010; Soyupak et al. 2011). Water quality models are sensitive to uncertainties in parameters like reaction coefficient, initial concentration, hydraulic model errors, structural errors, demand uncertainties etc. (May et al.

2008). Data assimilation techniques were found to perform better than inverse modelling approaches, even in presence various of system uncertainties. (Liu et al. 2012). Also, unlike most of the conventional methods, data assimilation methods are also capable of incorporating real time sensor data for estimating the system state and parameters, thereby, making it an efficient tool for real time modelling of dynamic systems (Hendricks Franssen and Kinzelbach 2008).

Application of data assimilation span across numerous scientific disciplines such as electrical systems (Beides and Heydt 1991; Doucet et al. 2001; Blood and Krough 2008), oceanic sciences (Park and Kaneko 2000; Carton and Giese 2008), meteorological/atmospheric sciences (van Loon et al. 2000; Kalnay 2003), groundwater (Dre'court et al. 2006; Hendricks Franssen et al. 2008) , gas and petroleum engineering (Benkherouf and Allidina 1988, Emara-Shabaik et al. 2002, Liu et al.2005), surface water quality (Pastres et al. 2003).

In WDS, existing applications of data assimilation techniques are mainly focused on hydraulic state estimation and event detection (Kang and Lansey 2009; Ye and Fenner 2010 & 2013; Jung and Lansey 2014, Okeya et al. 2014). It was observed that, most of the techniques used for hydraulic state estimation involves a linear data assimilation technique, such as - Kalman Filter or Extended Kalman Filters etc. (Hutton et al. 2014) . Owing to the high non-linearily of the water quality models, these linear data assimilation models cannot be applied directly for water quality state estimation in WDS. Hence, in this study, Monte Carlo based Ensemble Kalman Filter (EnKF) (Burgers et al. 1998) was used for chlorine data assimilation . This method is applied to WDS under two different uncertainities : (i) Source concentration (C_0) uncertainty , (ii) wall decay parameter uncertainty (k_w). Two different variants of EnKF (non-iterative and iterative EnKF) were formulated in this study and these methods were altered to deal with the problem of model variable initialization at intermediate time steps by implementing the Restart technique (Geir et al. 2003). These methods were tested on two WDS: (i) Brushy plains network (Rossman et al. 1994) (ii) and a big, city-wide WDS, Bangalore inflow network (Manohar and Mohan Kumar 2013). In this study, it is assumed that the hydraulic model of the WDS is fully calibrated and hence, the uncertainties related to pipe roughness coefficient and systems demands are not considered.

The main objectives of this study is to compare the two variants of EnKF for application in water quality state estimation under system parameter uncertainties. Different scenarios are tested for assessing the applicability of these data assimilation methods. These scenarios studied are : Scenario (i) : The source concentration value (C_0),is considered uncertain; Scenario ii: The reaction parameter value (k_w), is considered uncertain. For each of these scenarios, the following sub-scenarios are also studied in this work : Sub-scenario (a). the number of realizations (n) are varied ; Sub-scenario (b). The number (m) and location of sensors are varied, Sub-scenario (c), measurement error and measurement noise is considered, in order to understand the sensitivity of data assimilation model.

Methodology

The over-all methodology adopted in this study has two parts: (i) Hydraulic and water quality simulation and (ii) water quality data assimilation model.

Water Quality Prediction Model

In this study, the water quality simulation (i.e. prediction) model consists of hydraulic and chlorine reaction and transport components, modelled using EPANET (Rossman 2000). A mass balance equation based one directional advection- dominated transport and reaction kinetics is used for chlorine concentration modelling in WDS. The partial differential equation governing chlorine transport in a pipe is:

$$\frac{\partial C_i(x,t)}{\partial t} + v_i \frac{\partial C_i(x,t)}{\partial x} - R[C_i(x,t)] = 0 \quad (1)$$

where, $C_i\ (x,t)$ is the chlorine concentration at any point x within link i, at time t. v_i is the mean flow velocity of the water*;* and $R[C_i(x, t)]$ is the reaction- rate expression. In this study, a first order wall and first order bulk reaction model is being used:

$$R[C_i(x,t)] = -k_b C_i(x,t) - \frac{k_w k_f}{r_h(k_w+k_f)} C_i(x,t) \qquad (2)$$

where, k_b is the first order decay rate constant in the bulk flow (1/day), k_w is the wall decay parameter (m/day), k_f is the mass-transfer coefficient (m/day) and r_h is the hydraulic radius of pipe (one half the pipe radius).

More details about water quality modelling in WDS is available in the literature (Biswas et al. 1993; Clark et al.1993 & 1995; Hallam et al. 2002; Grayman et al. 1988; Munavalli and Kumar 2005; Rossman et al. 1994; Vasconcelos et al. 1997).

Accurate modelling of chlorine concentrations in a WDS needs accurate understanding of decay mechanisms in the bulk water and on the pipe walls. Uncertainty analysis of water quality models have established the wall decay coefficient as the most sensitive parameter for water quality model output (Pasha and Lansey 2010). The wall decay coefficient in a WDS depends on the diameter of the pipe, flow in the pipe, concentration of chlorine, pipe service age etc. (Al-Jasser 2007; Fisher et al. 2017), whereas the bulk decay parameter mainly depends on the source water properties, and it seldom varies unless there is change in the source water quality. Along with decay parameters for chlorine in WDS, the water quality model output is sensitive to the source concentration value as well. The source chlorine concentration (C_0) is usually monitored in WDS, but in case of measurement errors or sensor failure, the estimate of chlorine concentration across the system might vary and will lead to under or overdosing of the disinfectant. Hence, in this study, we are dealing with two different system uncertainties in water quality model development: (i) uncertainty in the input data or in this case, source concentration of chlorine (C_0) and (ii) uncertainty in the wall reaction parameter for chlorine reaction in pipelines (k_w). In this study, the uncertainties related to the hydraulic model such as demand uncertainty, pipe roughness coefficient etc. are not considered, since accounting for these uncertainties make the problem more complex, and the EnKF based data assimilation methodologies adopted in this study cannot be directly applied to deal with these uncertainties.

Data assimilation for water quality state and parameter estimation

Data assimilation involves estimating the state of a particular system based on the predictions and observations leading up to the present time. EnKF (Evensen 1994; Burgers et al. 1998; Evensen 2003), is a Monte Carlo implementation of the Bayesian update problem. EnKF is a special case of Kalman Filter (Kalman 1960) which uses ensembles or stochastic realization (with different parameter and initial condition values) for approximating the states of the system. EnKF based data assimilation consists of two steps: (i) Prediction step and (ii) Update step. In the prediction step, a forward simulation model is used to predict the system state as in equation (3):

$$x_{t+1}^{i-} = f\left(x_t^i, u_t^i, \theta, t\right) + \omega_t \, , i = 1, \ldots., n \qquad (3)$$

According to equation (3), the x_{t+1}^{i-} is the i^{th} ensemble member forecast at time $t+1$, and x_t^i is the i^{th} updated ensemble member at time t. Here, f is the forward simulation model (in this case, EPANET water quality model of the system). ω_t is the process noise (assumed to be zero in this study),θ is the system parameters, u_t are the forcing data or the system inputs . Ensembles of the forcing data (u_t^i) are created by adding noise ε_t^i, sampled from a distribution of mean zero and variance, Σ_t^u , to the input data u_t.

$$u_t^i = u_t + \varepsilon_t^i \, , \; \varepsilon_t^i \sim N(0, \Sigma_t^u) \qquad (4)$$

The parameters in this study are: pipe roughness coefficient (C), hourly demand multiplier (d_m), initial concentration of chlorine (C_0), chlorine reaction parameters. Among the above listed parameters, C and d_m are assumed to be known, hence can be classified as system input u. Additional inputs required for predicting the system states are the network boundary conditions (tanks initial level, reservoir head etc.) and base demand values at the nodes. .

From x_{t+1}^{i-}, the predicted states of the system, $\hat{y}_{t+1}^{i}$, the predicted measurements are computed as

$$\hat{y}_{t+1}^{i} = h\left(x_{t+1}^{i-}, \theta\right) \tag{5}$$

where h shows the relationship between the system states, parameters and the observations/measurements.

y_{t+1} is the field observation at the $t+1^{th}$ time step, for which ensembles are generated by adding a noise, λ_{t+1}^{i}.

$$y_{t+1}^{i} = y_{t+1} + \lambda_{t+1}^{i}, \lambda_{t+1}^{i} \sim N\left(0, \Sigma_{t+1}^{y}\right) \tag{6}$$

The forecasted states ensembles (equation 3) are updated using a linear correction equation according to the standard Kalman filter (equation 7):

$$x_{t+1}^{i} = x_{t+1}^{i-} + K_{t+1}\left(y_{t+1}^{i} - \hat{y}_{t+1}^{i}\right) \tag{7}$$

Here, K_{t+1} is the Kalman gain matrix which is estimated from the covariance matrices as shown in equation 8 (Moradkhani et al.,2005):

$$K_{t+1} = \Sigma_{t+1}^{xy}\left[\Sigma_{t+1}^{yy} + \Sigma_{t+1}^{y}\right]^{-1} \tag{8}$$

where, $\Sigma_{t+1}^{xy}$is the forecast cross covariance of a priori state estimate x_{t+1}^{i-} and prediction $\hat{y}_{t+1}^{i}$, and Σ_{t+1}^{yy} is the forecast error covariance of prediction $\hat{y}_{t+1}^{i}$.In equation (7), the term $K_{t+1}\left(Y_t - \hat{Y}_t\right)$, is the perturbation vector (Hendricks Franssen and Kinzelbach 2008). In equation (3), one of the key assumption is that the system parameter θ is deterministic. In scenarios, where the parameters θ are unknown or uncertain, non-iterative or iterative EnKF methods need to be used. These methods enable estimation of the uncertain system parameter along with states using the real time observations from the field. In this work, the system state and the model parameters (C_0 and k_w) are updated using two ensemble-based data assimilation methodologies: (i) Non-Iterative Restart EnKF (NIR-EnKF), and (ii) Iterative Restart EnKF (IR-EnKF).

In non-iterative- EnKF method, the parameter and the system states are combined to form an augmented state vector, which enable simultaneous estimation of states and parameters (Naevdal et al. 2003; Hendriks Franssen 2008). Whereas, in an iterative-EnKF, first the parameters are updated using the current system measurements, and the updated parameters are used to predict and update the system states for the same time step (Moradkhani et al. 2005). In both non-iterative and iterative EnKF methods, after updating the system parameter, the forward simulation model (equation 3) is restarted from *t*: 0. This technique of starting the simulation from t :0 is called Restart EnKF (Gu and Oliver 2007, Hendricks Franssen and Kinzelbach 2008, Song et al., 2014). In this study, Restart procedure was implemented to reduce the error in EnKF model output due to parameter and system initialization during intermediate water quality time steps.

Non-Iterative Restart EnKF

As mentioned earlier, in this approach, the states and the parameters are updated jointly. If there are N states and M parameters, the augmented state vector will be of size (N+M,1). Forecasted ensembles of system parameters are created by adding a noise ζ_t^i with covariance Σ_t^0 to the updated parameter value of the previous timestep.

$$\theta_{t+1}^{i-} = \theta_t^i + \zeta_t^i\,, \;\; \zeta_t^i \sim N(0, \Sigma_t^0) \tag{9}$$

These forecasted parameter ensembles θ_{t+1}^{i-}, are updated (equation 10) simultaneously with the forecasted states x_{t+1}^{i-} (equation 7)

$$\boldsymbol{\theta}_{t+1}^{i} = \boldsymbol{\theta}_{t+1}^{i-} + \boldsymbol{K}_{t+1}^{\boldsymbol{\theta}}\left(\boldsymbol{y}_{t+1}^{i} - \widehat{\boldsymbol{y}}_{t+1}^{i}\right) \tag{10}$$

Here, $\boldsymbol{K}_{t+1}^{\boldsymbol{\theta}}$, is the Kalman gain for updating the model parameter.

$$\boldsymbol{K}_{t+1}^{\boldsymbol{\theta}} = \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\theta y}}\left[\boldsymbol{\Sigma}_{t+1}^{\boldsymbol{yy}} + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{y}}\right]^{-\mathbf{1}} \tag{11}$$

Here $\Sigma_{t+1}^{\theta y}$ is the cross covariance of the predicted parameter and measurement ensembles. Rest of the terms are same as that of EnKF. The state vector is updated using equation (8) and the corresponding Kalman gain is calculated as in equation (9). In this method, after each time step (after updating the states and parameter ensembles), the simulation is restarted from *t*:0 [i.e. Equation 3 is run from *t*: 0 for this algorithm, making it a NIR-EnKF].

Iterative Restart EnKF

IR-EnKF involves sequential forecast and update of parameters, followed by forecast and update of system states for a particular time period. In this method, the updated parameters (calculated using equation (10)), θ_{t+1}^{i}, are used to forecast the system states for the same time step (*t+1*) (Equation 12).

$$x_{t+1}^{i-} = f\left(x_t^i, u_t^i, \theta_{t+1}^i, t\right) \tag{12}$$

θ_{t+1}^{i} is the updated parameters for the time step *t+1*. The a priori water quality state of the system x_{t+1}^{i-}, is updated using the Kalman gain for state correction (Equation 8).

This two-step approach is supposed to limit the problems associated with the linearization of the relation between parameters and the observations.

Filter Inbreeding

During data assimilation, if the number of realizations are small, there exists an error due to sampling, and it will be reflected in the error covariance matrix. When there are insufficient realizations to span the model state space, the estimated error covariance will degrade after each time step, and this process is known as Filter inbreeding (Houtekamer and Mitchell1998; Lorenc2003). Whitaker and Hamill (2002) had suggested that the perturbations introduced in observations can also result in filter inbreeding.

Different methods are available in the literature for mitigating filter inbreeding effects (Anderson and Anderson 1999; Hamill et al. 2000; Anderson 2007). In this study, a mitigation approach based on a damping factor α is used to analyse the effect of measurement errors and measurement noise on the data assimilation model output. The value of α varies between 0 and 1(Hendricks Franssen and Kinzelbach 2008), and the state update equation is modified as follows:

$$x_{t+1}^{i} = x_{t+1}^{i-} + \alpha K_{t+1}^{x}\left(y_{t+1}^{i} - \hat{y}_{t+1}^{i}\right) \tag{13}$$

The data assimilation algorithms were implemented using the EPANET Toolkit in MATLAB.

Case Studies

The EnKF based data assimilation methodologies developed for water quality state estimation in WDS is tested and validated in two WDS: (i) Brushy plains WDS and (ii) Bangalore inflow network. This section provides details on the two networks used in this study.

Case Study 1: Brushy Plains WDS

This network has been used in various studies related to water quality and WDS hydraulics (Rossman et al.1994; Boccelli et al. 1998; Nilsson et al. 2005; May et al. 2008; Clark 2015).

Details of this WDS are available in Rossman et al. (1994), in which chlorine concentration data from 8 sampling nodes across the network can be found. The estimated bulk reaction coefficient value for this WDS was found to be -0.55 /day, and the wall reaction coefficient value was found to be in the range of -0.45 to -0.15 m/day. The source concentration of chlorine is maintained at 1.1-1.16 mg/L, injected at a constant rate at the pumping station. Fig.1 shows the schematic of Brushy Plains WDS. Eight nodes were selected as measurement nodes for this network (in accordance with earlier research carried out on this network, Rossman et al. (1994)). Those measurement nodes are: 3,6,10,11,19, 25, 28 and 34. Synthetic chlorine measurements were generated every 15 minutes for the total duration of simulation (16 hours). The hydraulic time step of the simulation was about 60 minutes.

Data assimilation was carried out for scenario (i) and scenario (ii). Sub-scenarios (a), (b) and (c) were also studied for this case study. For scenario (i) and scenario (ii), the initial ensembles of parameters (C_0 and k_w) were sampled from a normal distribution, respectively. Both NIR-EnKF and IR-EnKF were tested for their application under (i) uncertainty in C_0 value and (ii) uncertainty in k_w value, for this WDS.

For sub-scenario (a), various sizes of stochastic realizations (n) ranging from 20 -100 were generated for studying the variation in model accuracy with ensemble sizes. Sub-scenario (b) is simulated by reducing the number of measurement nodes (m). The number of measurement nodes (m) in the system are varied from 4 to 8 nodes, there by varying the measurement density in the system from 22 to 11percent. Two different sets of measurements are studied, each with 4 data sets. Measurement set A consists of data from nodes 3,6, 10 and 11, concentrated near to the pumping station, and measurement set B consisting of data from nodes 19, 25, 28 and 34, concentrated near to the tank. Varying the measurement locations and the measurement density in the WDS gives an idea of its effect on data assimilation.

The model performance in the presence of measurement errors and Gaussian noise for *n*: 20 is also studied in detail, using a damping factor α.(scenario (c)). In the sub-scenarios (a) and (b), the measurements used were assumed to be perfect, i.e. without any systematic errors or random noise. In order to replicate field measurements, the simulated measurement values were corrupted to generate noisy measurements and bad measurements. Hence, in this sub-scenario, two types of measurement ambiguities were considered: (i) systematic error, where a fixed value of 0.2 mg/L is added to a few of the measurements nodes (nodes 19, 25, 27 and 33); (ii) random noise, where a Gaussian noise of mean zero and standard deviation 0.05 mg/L is added to readings from all the measurement nodes. Presence of noise or error in the measurements usually induces filter inbreeding during data assimilation. Different values of damping factor α was used to mitigate the effects of these observational errors.

Case Study 2: Bangalore Inflow Network

The second case study is carried out as a verification problem, to validate the algorithm and to establish its applicability on a large WDS for a big city. The Bangalore water supply network is maintained and operated by Bangalore Water Supply and Sewerage Board (BWSSB) and was established by Karnataka Govt. during different time periods: Stage I of the system was established in year the 1974, Stage 2 was established in year the 1983, Stage 3 (year 1993) and Stage 4 Phase 1 (year 2002). Stage 1 of this network supplies about 140 MLD of water, Stage 2 supplies another 140 MLD, followed by 315 MLD by Stage 3 and 315 MLD by Stage 4 Phase 1, all of it amounting to a total of 910 MLD of water for Bangalore city. Since the system was established in different stages, zoning of pipes are carried out for Hazen William C value and wall decay parameter k_w. Further details of this network are available in Manohar and Kumar (2013). The hydraulic model of the WDS used was calibrated using field values.

A schematic of Bangalore inflow WDS is given in Fig. 2. In this network, the pipes are grouped into 4 different class: pipes 1-41, 42-69, 70-137 and 138-180 and the k_w values are -1 (Stage I), -0.75 (Stage II), -0.5 (Stage III) and -0.25 m/day (Stage IV Phase I). The first order bulk reaction coefficient is taken as 2.0 day^{-1}, and a constant chlorine concentration of 0.75 mg/L is assumed to be injected from all the four sources (Munavalli and Kumar, 2003 & 2005). The consumer demands are loaded on the GLRs and are assumed to vary temporally based on a bi-modal demand pattern (peak factor: 1.6, and 1.2).

A total of 60 measurement nodes are assumed to be present in this network. The chlorine measurements were generated once every 15 minutes for a total duration of 16 hours. The hydraulic time step is about 60 minutes. As in the case study 1, two different scenarios are tested for this network: scenario (i) uncertainty in source concentration (C_0) and scenario (ii) uncertainty in wall decay coefficients (k_{w1}, k_{w2}, k_{w3} and k_{w4}) for all the pipe groups. In the previous case study, the global wall reaction coefficient is considered (k_w value same for all the pipes in the WDS), where as in this study, a zoned wall reaction coefficient is considered. Complexity of this WDS is much higher than the previous case study owing to its size and multi-source supply. For this case study, the conclusions drawn from the previous case study are used to reduce the computational complexity, and to validate the developed algorithms.

Performance criteria

Two different performance measures are used in this study to assess the data assimilation accuracy: (i) Average Absolute Error (AAE) and (ii) Average Ensemble Standard Deviation (AESD) (Hendricks Franssen and Kinzelback 2008):

$$AAE = \frac{1}{M*T}\sum_{i=1}^{M}\sum_{t=1}^{T}\left|\bar{x}_{i,t} - y_{i,t}\right|, \ i: 1,2 \dots\dots M \qquad (14)$$

$$AESD = \frac{1}{M*T}\sum_{i=1}^{M}\sum_{t=1}^{T}\sqrt{\frac{\sum_{j=1}^{n}\left(x_{i,j,t} - \bar{x}_{i,t}\right)^2}{n}} \qquad (15)$$

where, x is the simulated chlorine concentration for each realizations, y is the measured chlorine concentration at the node, $\bar{x}$ indicate the ensemble average value, T is the total time of simulation, M is the number of non-measurement nodes in the WDS and n indicates the number of stochastic realizations (number of ensembles) [j : 1,2..., n]. Here, AAE and AESD indicate the overall performance of the EnKF based data assimilation techniques for the entire time of simulation, T, for the WDS.

Visual comparison based on simulated and measured values of free chlorine at different measurement nodes in the WDS are also carried out to assess the model performance. Mean Average Percentage Error (MAPE) for the entire duration of simulation is also calculated to assess the WDS performance under different scenarios.

Results and Discussions

In this section, the results obtained for each case study and the corresponding scenarios are presented and discussed in detail.

Case Study 1: Brushy Plains WDS

Scenario (i) and scenario (ii) were tested for this case study along with sub-scenarios (a), (b) and (c) .The results of this study is presented in the following sections.

Scenario (i): Uncertainty in source chlorine concentrations (C_0)

The main observations of this study are summarised below:

Comparison of NIR-EnKF and IR-EnKF: Fig.3 shows the variation of MAPE for the WDS for the duration of simulation. It can be deduced from Fig.3, that both NIR-EnKF and IR-EnKF, reduced the prediction error to 10 % by the end of simulation (IR-EnKF reduced the MAPE to 5% by the end of simulation). The AAE values estimated at all the nodes in the WDS for the duration of simulation ranged from 0-0.19mg/L. For this scenario, the difference between NIR-EnKF and IR-EnKF is negligible. It is observed that, IR-EnKF is slightly more accurate than NIR-EnKF, whereas IR-EnKF takes more computational time than NIR-EnKF

Sub-scenario (a): Simulations are carried for different values of *n*, and it is observed that, as the number of stochastic realizations (*n*) increased, the model output accuracy increased, but for *n* values greater than 20, change in the AAE values are negligible (Table 1). Filter inbreeding was not observed in any of these simulated results, even for *n*=20. As the *n* value was increased from 20 to 100, the estimated AESD values increased for each node for the duration of simulation. The AESD values are higher than AAE values, for most of the nodes. This indicates adequate spread of the updated state ensemble. Similar results were observed in data assimilation studies in the groundwater domain. (Hendriks Franssen and Kinzelbach, 2008). In Hendriks Franssen and Kinzelbach (2008), it was observed that AESD in the estimated log-transmissivity increased with the number of realizations.

Sub-scenario (b): This sub-scenario was simulated for *n*=20. In this study, it was found that the location and number of measurements points were essential for reducing the AAE for the assimilated quality states in WDS (see Table 1). Fig.4 shows that, measurement set A is able to assimilate the water quality measurements for the entire WDS, and it is better than measurement set B, as set B gives higher values of MAPE (around 30-55%higher) at certain time steps. Among measurement sets A and B, measurement set A is able to estimate chlorine concentration at almost all the nodes with substantial accuracy. It might be due to the fact that, set A is very close to the pump station which is a boundary condition for Brushy Plains WDS, and it is the chlorine source as well.

Sub-scenario (c): Table 2. illustrates the effect of damping factor on the model output, in presence of measurement error and measurement noise. Under C_0 uncertainty, α :1 could handle the measurement errors during data assimilation at all the nodes in the WDS, for the duration of simulation (Fig. 5(a)),but the AAE for this sub-scenario is higher than the scenario when no measurement error was present (Table 1).

For mitigating the effects of measurement noise in the system, clearly α: 1 is better than all other values of α (see Fig.5(b)) . α : 0.1 and 0.01 have better model output at a few time steps (Fig. 5(b) and Table 2). Hence, it can be concluded that for a given WDS, the effect of measurement noise and measurement error on model performance is negligible and *n:* 20 is adequate to simulate the system state at all time periods, without covariance degradation. The quality of the state estimates were found to be affected by measurement noise and errors, but α : 1 provides a better estimate of the states compared to other values of the damping factor. Lower values of α gives better results during certain time-steps because, at these time-steps, the impact of spurious numerical co-variances on the updating of states is reduced(i.e. the value perturbation vector($\boldsymbol{K}(\boldsymbol{Y}_t - \widehat{\boldsymbol{Y}}_t)$)is reduced at these time-steps (Hendricks Franssen and Kinzelbach, 2008).

Scenario (ii): Uncertainty in wall reaction coefficient, k_w

In this scenario, the wall reaction coefficient is used as the uncertain input to the water quality data assimilation model. The initial/source chlorine concentration is considered known (1.1-1.16 mg/L). NIR- EnKF and IR-EnKF methods are compared for chlorine concentration estimation by assimilating the field measurements under uncertainty in the k_w value, for different sub-scenarios.

Comparison of NIR-EnKF and IR-EnKF: Table 3 summarizes the AAE and AESD for the WDS, for the duration of study, for different scenarios. Also, the MAPE for the system reduced to < 5% for IR-EnKF at the end of simulation (see Fig.6). It can be observed that IR-EnKF is better than NIR-EnKF when dealing with uncertainty in the wall reaction coefficient during data assimilation. Due to the nonlinear relationship between the parameter and the observations, iterative filters are more appropriate for state estimation in WDS under reaction parameter uncertainty.

Sub-scenario (a): The effect of the number of realizations on the model output was similar to scenario (i). When the number of ensembles was increased from 20-100, AAE values were found to reduce, but the reduction in AAE is not substantial for n>20 (Table 3).

Sub-scenario (b) : Fig.7 shows the MAPE values of the estimated chlorine concentration for the WDS under k_w uncertainty, for measurement set A and measurement set B. It is clear from Fig.7 that, for every time steps, set A performs better than set B. The overall performance of the data assimilation technique reduces with reduction in the number of measurement nodes.

Sub-scenario (c) : It was found that model performance was unaffected by measurement error, though α: 1 and α: 0.1 had similar response at all nodes, at all time-steps (Table 2.). When measurement noise was introduced, it was found that, α: 0.1, performed better than α:1 for most of the time-steps (see Fig.8(b)), but the improvement in model performance was not substantial (the change in MAPE was about 1-2%). Hence, it can be deduced that, noise or error induced degeneration of the covariance matrix was not much in this WDS for n: 20.

Based on the results from sub-scenario (c) (for both scenario (i) and scenario (ii)), it is observed that measurement noise and measurement error is not creating large variations in the perturbation vector (when compared with the case when no measurement error or noise is considered) [perturbation vector : $\boldsymbol{K}(\boldsymbol{Y_t} - \boldsymbol{\hat{Y}_t})$]. But, it should be noted that , measurement error and measurement noise reduced the accuracy of the data assimilation model (Table 1, Table 2 and Table 3).

Estimated Parameter Values

Table 4 show the computed mean values for the parameters, C_0 and k_w at the end of the simulation period. Mean values were computed for the simulation where *n:* 20. It is clear from the results that, data assimilation technique based on EnKF can be used for dynamic state estimation and parameter estimation (C_0 and k_w) in WDS under various uncertainty and measurement location scenarios. The values obtained using data assimilation techniques were found to be comparable to parameters estimated using inverse modelling methodologies (Munavalli and Kumar, 2005).

Case Study 2: Bangalore Inflow System

In this case study, data from 60 measurement nodes (30 network junctions and 30 tanks) were assimilated with the network water quality model. The number and location of these 60 measurement nodes were chosen heuristically for an optimal concentration estimation across the WDS. Initially, 10 nodes were assigned across the network at random, such that they are uniformly distributed across the network. Data assimilation was carried out (for scenario (i)) , and based on the error in estimation of nodal chlorine concentration, nodes with higher error values were added to the measurement node set. The measurement nodes were added such that, no two measurements nodes were adjacent. Similar procedure was carried out for deciding the measurements tanks as well. The locations chosen include 30 tanks and 30 junctions spread across the network. Fig.9 shows the variation of AAE with *m* value for this study. It was found that, as *m* value increased, the error in estimation reduced, but the reduction in error was not substantial after certain *m* value. For in-depth understanding of the sensitivity of the number and location of measurements nodes on the data assimilation model accuracy, a detailed analysis need to be carried out. A detailed sensitivity analysis is beyond scope of this paper, and will be carried out in future works.

In this analysis, the tank measurements were used to assimilate the chlorine concentration values at the tanks and junction measurements were used to assimilate the chlorine concentration data at the junctions in the network, and the tank and junction states are updated simultaneously. The conclusions drawn from the previous case study was utilized here, as this case study is considered as a validation problem for water quality data assimilation application in large scale WDS. Scenario (i) and scenario (ii) are considered for this case study. The number of stochastic realizations, n is 50 , for this case study, since it was observed that the AESD and AAE values do not change significantly for values of n > 50. No measurement errors are considered in this WDS. In this case study, the performance indicators (AAE and AESD) are slightly modified, since these values are calculated for each node, and are not averaged over all non-measurement nodes (i.e. in equation (14) and (15), averaging over M is not considered).

Scenario (i): Uncertainty in source concentration (C_0)

NIR-EnKF is used for state estimation in this scenario. Fig.10 show the AAE (mg/L) for non-measurement nodes and tanks in the WDS. It can be observed that NIR-EnKF is able to estimate the chlorine concentration estimate of the network with an AAE accuracy of about 0.005 - 0.2 mg/L. It is observed that the AAE values at nodes upstream and downstream of valves and pumps were generally higher (AAE > 0.2 mg/L) compared to the error estimates at other nodes. This is due to the hydraulic modelling constraint associated with the forward simulation model. In the forward simulation model adopted (EPANET) in this work, valves and pumps are modelled as network links without length, i.e. the nodes upstream and downstream of these links are hypothetical. Due to of this constraint, the variation in flow velocity across the valves and pumps, generates an estimate of chlorine concentration, which is higher than the actual value. All the remaining nodes have AAE value below 0.18 mg/L, and about 75% of the nodes have AAE value below 0.12 mg/L (Fig.10). AAE for chlorine estimates at the tanks of this network were found to be below 0.2 mg/L for all the non-measurement tanks (Fig.10). This high level of accuracy might be due to a high measurement density with respect to tanks in the network. In Fig. 10, AAE values are presented only at the non-measurement nodes in the figure; Measurement nodes, and the nodes upstream and downstream of valves and pumps are not shown in the figure.

Scenario (ii): Uncertainty in wall decay coefficient (k_w)

IR-EnKF was used to estimate the water quality state under uncertainty in reaction coefficient for case study 2. The k_w parameters were zoned in the network according to the pipe age (dependent on the phase of development of the WDS). In this case study, IR-EnKF is able to estimate the chlorine concentration at the tanks and nodes of this network with an accuracy of $\leq$0.2mg/L. Fig.11 shows the AAE for all the non-measurement nodes and tanks in the WDS (AAE values are not reported at the measurement nodes, and the nodes upstream and downstream of valves and pumps in the figure). It was observed that the number of nodes with AAE ~ 0.2 mg/L is greater than the previous scenario. Frequent flow reversal occurs in many pipes in this WDS, which along with disparity in k_w value across the system contributed to a higher value of AAE in many nodes. As many as 36 nodes in the system have AAE values almost equal to 0.2 mg/L. More than 75% of the nodes in this system have AAE value below 0.18 mg/L and it was observed that the tank estimates for chlorine concentration are good and all the tanks have AAE < 0.2 mg/L.

Estimated Parameter Values

The parameter values estimated at the end of the simulation are given in Table 5. The ensemble mean value of C_0 was calculated to be 0.7534 mg/L. Mean value for k_{w2} and k_{w3} were: -0.7784 and -0.504 m/d respectively. The k_{w1} value for this case study was estimated to be lesser than the actual value, whereas, k_{w4} value was estimated to be higher than the actual value. Frequent flow reversal happens in pipes in group 1 (k_{w1}) and group 4 (k_{w4}), and grouping of pipes solely based on the service age, are the reasons for this disparity between actual and estimated k_{w1} and k_{w4} values. The estimated values are compared with the steady state-inverse modelling study carried out by Munavalli and Mohan Kumar (2003) on an earlier version of the network, which had only Stage 1, 2 and 3. From these results it is concluded that the data assimilation method is able to achieve the same level of accuracy as that of inverse modelling.

Summary and Conclusions

This work introduces a novel method for estimating chlorine concentration across a WDS in real time using data assimilation techniques. Two variants of the EnKF are studied and applied on two WDS. The major conclusions drawn from this study are stated in this section.

In this study, it was found that, the uncertainty in the source concentration can be dealt by both NIR-EnKF and IR-EnKF. However, the computational time required for NIR-EnKF method is lesser than IR-EnKF based data assimilation method.

It was found that, the non-linear relationship between the parameters and the measurements cannot be addressed with non-iterative data assimilation methods, hence IR-EnKF was more accurate than NIR-EnKF for data assimilation in presence of k_w uncertainty. For both the case studies, the data assimilation approach was able to accurately estimate the dynamic state and parameter of the system under different input parameter uncertainties- C_0 uncertainty and k_w uncertainty.

The NIR-EnKF and IR-EnKF based data assimilation technique were able to reach the good output accuracy across Brushy plains network, for state estimation under uncertainty in C_0 and k_w. Since case study 2 was developed in stages, , the pipes in the WDS were grouped based on wall reaction coefficients, to estimate accurate values of the chlorine concentrations across the system. The results of this case study illustrate the capability of EnKF based assimilation methods to deal with system uncertainties irrespective of the size of the network. The limited sensitivity analysis carried out in this study showed the variation of model accuracy with the number and location of measurement nodes. For an in-depth understanding of the sensitivity of the number and location of measurements nodes on the data assimilation model, a detailed sensitivity analysis need to be carried out.

With regard to the field application of this method, the model output will be influenced by uncertainties in the hydraulic model of the system. Uncertainties related to the hydraulic model induces additional non-linearity, in the forward simulation model, hence, the output of the proposed data assimilation methods could become sub-optimal. Also, response of the data assimilation methods when the water quality reaction equation is of different order is not considered in this study. The data assimilation models will be sensitive to the order of water quality reactions, hence uncertainty in the order of reaction equation will also reduce the model accuracy. The results obtained in this paper could certainly be improved if these system constraints are also considered.

References

Al-Jasser, A. O. (2007). "Chlorine decay in drinking-water transmission and distribution systems: Pipe service age effect." *Water Res.*, *41*(2), 387-396.

Anderson, J. L. (2007). "An adaptive covariance inflation error correction algorithm for ensemble filters." *Tellus A*, *59*(2), 210-224.

Anderson, J. L., and Anderson, S. L. (1999). "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts." *Monthly Weather Review*, *127*(12), 2741-2758.

Aral, M. M., Guan, J., and Maslia, M. L. (2009). “Optimal design of sensor placement in water distribution networks.” *J. of Water Resour.Plann. and Manage.*, *136*(1), 5-18.

Baxter, C. W., Stanley, S. J., and Zhang, Q. (1999). “Development of a full-scale artificial neural network model for the removal of natural organic matter by enhanced coagulation.” *J. Water Supply: Res. and Tech.-AQUA*, *48*(4), 129-136.

Baxter, C. W., Zhang, Q., Stanley, S. J., Shariff, R., Tupas, R. R., and Stark, H. L. (2001). “Drinking water quality and treatment: the use of artificial neural networks.” *Canadian J. of Civil Engg.*, *28*(S1), 26-35.

Beides, H. M., and Heydt, G. T. (1991). “Dynamic state estimation of power system harmonics using Kalman filter methodology.” *IEEE Transactions on Power Delivery*, *6*(4), 1663-1670.

Benkherouf, A., and Allidina, A. Y. (1988, March). “Leak detection and location in gas pipelines.” In *IEE Proceedings D-Control Theory and Applications* (Vol. 135, No. 2, pp. 142-148). IET.

Biswas, P., Lu, C., and Clark, R. M. (1993). “A model for chlorine concentration decay in pipes.” *Water Res.*, *27*(12), 1715-1724.

Blood, E. A., Krogh, B. H., and Ilic, M. D. (2008, July). “Electric power system static state estimation through Kalman filtering and load forecasting.” In *Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE* (pp. 1-6). IEEE.

Boccelli, D. L., Tryby, M. E., Uber, J. G., Rossman, L. A., Zierolf, M. L., and Polycarpou, M. M. (1998). “Optimal scheduling of booster disinfection in water distribution systems.” *J. of Water Resour.Plann. and Manage.*, *124*(2), 99-111.

Bowden, G. J., Nixon, J. B., Dandy, G. C., Maier, H. R., and Holmes, M. (2006). “Forecasting chlorine residuals in a water distribution system using a general regression neural network.” *Mathematical and Computer Model.*, *44*(5-6), 469-484.

Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly Weather Rev.*, *126*(6), 1719-1724.

Carton, J. A., and Giese, B. S. (2008). A reanalysis of ocean climate using Simple Ocean Data Assimilation (SODA).” *Monthly Weather Rev.*, *136*(8), 2999-3017.

Clark, R. M. (2015). “The USEPA's distribution system water quality modelling program: a historical perspective.” *Water and Envi. J.*, *29*(3), 320-330.

Clark, R. M., Grayman, W. M., Males, R. M., and Hess, A. F. (1993). “Modeling contaminant propagation in drinking-water distribution systems.” *J. of Envi.Engg.*, *119*(2), 349-364.

Clark, R. M., Rossman, L. A., and Wymer, L. J. (1995). “Modeling distribution system water quality: Regulatory implications.” *J. of Water Resour.Plann. and Manage.*, *121*(6), 423-428.

D D'Souza, C., and Kumar, M.S.M. (2010). “Comparison of ANN models for predicting water quality in distribution systems.” *American Water Works Association. J.*, *102*(7), 92.

Doucet, A., Gordon, N. J., and Krishnamurthy, V. (2001). “Particle filters for state estimation of jump Markov linear systems.” *IEEE Transactions on signal processing*, *49*(3), 613-624.

Drécourt, J. P., Madsen, H., and Rosbjerg, D. (2006). “Calibration framework for a Kalman filter applied to a groundwater model.” *Adv. in Water Resour.*, *29*(5), 719-734.

Emara-Shabaik, H. E., Khulief, Y. A., and Hussaini, I. (2002). “A non-linear multiple-model state estimation scheme for pipeline leak detection and isolation.” *Proceedings of the*

Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, *216*(6), 497-512.

Evensen, G. (1994). “Sequential data assimilation with a nonlinear quasi- geostrophic model using Monte Carlo methods to forecast error statistics.” *J. of Geophysical Res.: Oceans*, *99*(C5), 10143-10162.

Evensen, G. (2003). “The ensemble Kalman filter: Theoretical formulation and practical implementation.” *Ocean Dynamics*, *53*(4), 343-367.

Fisher, I., Kastl, G., and Sathasivan, A. (2017). “New model of chlorine-wall reaction for simulating chlorine concentration in drinking water distribution systems.” *Water Res.*, *125,* 427-437.

Geir, N., Johnsen, L. M., Aanonsen, S. I., and Vefring, E. H. (2003, January). “Reservoir monitoring and continuous model updating using ensemble Kalman filter.” In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Gibbs, M. S., Morgan, N., Maier, H. R., Dandy, G. C., Nixon, J. B., and Holmes, M. (2006). “Investigation into the relationship between chlorine decay and water distribution parameters using data driven methods.” *Mathematical and Computer Model.*, *44*(5-6), 485-498.

Grayman, W. M., Clark, R. M., and Males, R. M. (1988). “Modeling distribution-system water quality; dynamic approach.” *J. of Water Resour.Plann. and Manage.*, *114*(3), 295-312.

Gu, Y., and Oliver, D. S. (2007). “An iterative ensemble Kalman filter for multiphase fluid flow data assimilation.” *SPE J.*, *12*(04), 438-446.

Hall, J., Zaffiro, A. D., Marx, R. B., Kefauver, P. C., Krishnan, E. R., Haught, R. C., and Herrmann, J. G. (2007). “On-line water quality parameters as indicators of distribution system contamination.” *J.American Water Works Assoc.*, *99*(1), 66-77.

Hallam, N. B., West, J. R., Forster, C. F., Powell, J. C., and Spencer, I. (2002). “The decay of chlorine associated with the pipe wall in water distribution systems.” *Water Res.*, *36*(14), 3479-3488.

Hamill, T. M., Mullen, S. L., Snyder, C., Baumhefner, D. P., and Toth, Z. (2000). “Ensemble forecasting in the short to medium range: Report from a workshop.” *Bulletin of the American Meteorological Society*, *81*(11), 2653-2664.

Hart, W. E., and Murray, R. (2010). “Review of sensor placement strategies for contamination warning systems in drinking water distribution systems.” *J. of Water Resour.Plann. and Manage.*, *136*(6), 611-619.

Hendricks Franssen, H. J., and Kinzelbach, W. (2008). “Real- time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem”. *Water Resour. Res.*, *44*(9).

Houtekamer, P. L., and Mitchell, H. L. (1998). “Data assimilation using an ensemble Kalman filter technique.” *Monthly Weather Review*, *126*(3), 796-811.

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L., andSavić, D. A. (2012). “Dealing with uncertainty in water distribution system models: A framework for real-time modeling and data assimilation.” *J. of Water Resour.Plann. and Manage.*, *140*(2), 169-183.

Jung, D., and Lansey, K. (2014). “Water distribution system burst detection using a nonlinear Kalman filter.” *J. of Water Resour.Plann. and Manage.*, *141*(5), 04014070.

Kalman, R. E. (1960). “A new approach to linear filtering and prediction problems.” *J. of Basic Engg.*, *82*(1), 35-45.

Kalnay, E. (2003). “*Atmospheric modeling, data assimilation and predictability”*. Cambridge university press.

Kang, D., and Lansey, K. (2009). "Real-time demand estimation and confidence limit analysis for water distribution systems." *J.ofHyd.Engg.*, *135*(10), 825-837.

Larson, T. E. (1966). "Deterioration of water quality in distribution systems. " *J-American Water Works Assoc.*, *58*(10), 1307-1316.

Liu, M., Zang, S., and Zhou, D. (2005). "Fast leak detection and location of gas pipelines based on an adaptive particle filter." *International Journal of Applied Mathematics and Computer Science*, *15*(4), 541.

Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D-J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., and Restrepo, P. (2012). "Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities."*Hyd. and Earth Sys. Sci.*, 16, 3863.

Lorenc, A. C. (2003). "The potential of the ensemble Kalman filter for NWP- a comparison with 4D- Var." *Quarterly Journal of the Royal Meteorological Society*, *129*(595), 3183-3203.

Maier, H. R., Morgan, N., and Chow, C. W. (2004). "Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters." *Envi.Model.& Software*, *19*(5), 485-494.

Manohar, U., and Mohan Kumar, M. S.M. (2013). Modeling equitable distribution of water: Dynamic inversion-based controller approach. *J. of Water Resour.Plann. and Manage.*, *140*(5), 607-619.

May, R. J., Dandy, G. C., Maier, H. R., and Nixon, J. B. (2008). "Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems." *Envi.Model.& Software*, *23*(10-11), 1289-1299.

Milot, J., Rodriguez, M. J., and Sérodes, J. B. (2002). “Contribution of neural networks for modeling trihalomethanes occurrence in drinking water.” *J. of Water Resour.Plann. and Manage.*, *128*(5), 370-376.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). “Dual state–parameter estimation of hydrological models using ensemble Kalman filter.” *Adv. in Water Resour.*, *28*(2), 135-147.

Munavalli, G.R. (2002), *Simulation and parameter estimation of water quality in water distribution system*, PhD Thesis, Indian Institute of Science, Bangalore, India.

Munavalli, G. R., and Kumar, M. S. M. (2003). "Water quality parameter estimation in steady-state distribution system". *J. of Water Resour. Plann. and Manage.*, 129(2), 124-134.

Munavalli, G. R., and Kumar, M. S. M. (2004). “Modified Lagrangian method for modeling water quality in distribution systems.” *Water Res.*, *38*(13), 2973-2988.

Munavalli, G. R., and Kumar, M. M. (2005). “Water quality parameter estimation in a distribution system under dynamic state.” *Water Res.*, *38*(13), 2973-2988.

Naevdal, G., L. V. Johnsen, S. I. Aanonsen, and E. H. Vefring (2003). “Reservoir monitoring and continuous model updating using ensembleKalman filter.”*Soc. Petrol.*, 84372.

Nilsson, K. A., Buchberger, S. G., and Clark, R. M. (2005). Simulating exposures to deliberate intrusions into water distribution systems. *J. of Water Resour. Plann. and Manage.*, *131*(3), 228-236.

Okeya, I., Kapelan, Z., Hutton, C., and Naga, D. (2014). “Online modelling of water distribution system using data assimilation.” *Procedia Engineering*, *70*, 1261-1270.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., and di Pierro, F. (2008). "The battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms." *J. of Water Resour.Plann. and Manage.*, *134*(6), 556-568.

Park, J. H., and Kaneko, A. (2000). "Assimilation of coastal acoustic tomography data into a barotropic ocean model." *Geophysical Research Letters*, *27*(20), 3373-3376.

Pasha, M. F., and Lansey, K. (2010). "Effect of Parameter Uncertainty on Water Quality in Distribution Systems – Case study," *J. of Hydroinformatics*, 12(1),doi: 10.2166/hydro.2010.053.

Pastres, R., Ciavatta, S., andSolidoro, C. (2003). "The Extended Kalman Filter (EKF) as a tool for the assimilation of high frequency water quality data.: *Ecological modelling*, *170*(2-3), 227-235.

Polycarpou, M. M., Uber, J. G., Wang, Z., Shang, F., and Brdys, M. (2002). "Feedback control of water quality." *IEEE Control Systems*, *22*(3), 68-87.

Rodriguez, M. J., and Sérodes, J. B. (1998). "Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems." *Envi. Model.& Software*, *14*(1), 93-102.

Rossman, L. A. (2000). *EPANET 2: User's manual*. Cincinnati, OH, US-EPA.

Rossman, L. A., Boulos, P. F., and Altman, T. (1993). "Discrete volume-element method for network water-quality models." *J. of Water Resour.Plann. and Manage.*, *119*(5), 505-517.

Rossman, L. A., Clark, R. M., and Grayman, W. M. (1994). "Modeling chlorine residuals in drinking-water distribution systems." *J. of Envi.Engg.*, *120*(4), 803-820.

Sérodes, J. B., Rodriguez, M. J., and Ponton, A. (2001). “Chlorcast©: a methodology for developing decision-making tools for chlorine disinfection control.” *Env.Model.& Software*, *16*(1), 53-62.

Simone, A., Giustolisi, O., and Laucelli, D. B. (2016). “A proposal of optimal sampling design using a modularity strategy.” *Water Resour. Res.*, *52*(8), 6171-6185.

Song, X., Shi, L., Ye, M., Yang, J., and Navon, I. M. (2014). “Numerical comparison of iterative ensemble Kalman filters for unsaturated flow inverse modeling.” *Vadose Zone Journal*, *13*(2).

Soyupak, S., Kilic, H., Karadirek, I. E., and Muhammetoglu, H. (2011). “On the usage of artificial neural networks in chlorine control applications for water distribution networks with high quality water.” *J. of Water Supply: Res. and Tech.-Aqua*, *60*(1), 51-60.

Suresh, M., Manohar, U.,Anjana, G.R , Stoleru, R., and Kumar, M. S. M.(2014, October). “A cyber-physical system for continuous monitoring of Water Distribution Systems.” In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2014 IEEE 10th International Conference on* (pp. 570-577). IEEE.

van Loon, M., Builtjes, P. J., and Segers, A. J. (2000). “Data assimilation of ozone in the atmospheric transport chemistry model LOTOS.” *Envi.Model.& Software*, *15*(6-7), 603-609.

Vasconcelos, J. J., Rossman, L. A., Grayman, W. M., Boulos, P. F., and Clark, R. M. (1997). Kinetics of chlorine decay.” *J.American Water Works Assoc.*, *89*(7), 54.

Whitaker, J. S., and Hamill, T. M. (2002). “Ensemble data assimilation without perturbed observations.” *Monthly Weather Rev.*, *130*(7), 1913-1924.

Ye, G., and Fenner, R. A. (2010). “Kalman filtering of hydraulic measurements for burst detection in water distribution systems.” *J of Pipeline Sys. Engg. and Prac.*, *2*(1), 14-22.

Ye, G., and Fenner, R. A. (2013). “Weighted least squares with expectation-maximization algorithm for burst detection in UK water distribution systems.” *J. of Water Resour.Plann.and Manage.*, *140*(4), 417-424.

<u>FIGURE CAPTIONS</u>:

Fig. 1. Schematic of Brushy plains WDS - Case study 1

Fig. 2. Schematic of Bangalore Inflow System - Case study 2 (with Tanks and Junction ID)

Fig. 3. MAPE for the case study 1 under C_0 uncertainty [*n*: 20, *m*: 8]

Fig. 4: MAPE for the case study 1 under C_0 uncertainty for *m*: 4 [*n*:20, using NIR-EnKF]

Fig. 5. MAPE for case study 1 under C_0 uncertainty (a) different α for measurement error[0.2 mg/L]; (b) different α for measurement noise [Gaussian]; [*n* :20 and *m*: 8, using NIR-EnKF]

Fig. 6. MAPE for case study 1 under k_w uncertainty [*n*: 20, *m*: 8]

Fig. 7. MAPE for case study 1 under k_w uncertainty, for *m* :4 [*n*: 20, using IR-EnKF]

Fig. 8. MAPE for case study 1 under k_w uncertainty (a) different α for measurement error [0.2 mg/L]; (b) different α for measurement noise [Gaussian]; [*n* :20 and *m*: 8, using IR-EnKF]

Fig.9. AAE plots for different *m* values (case study 2 - under C_0 uncertainty, using NIR-EnKF)

Fig. 10. AAE (mg/L) for non-measurement nodes and tanks in case study 2, using NIR-EnKF [*n*=50,*m*=60]

Fig. 11. AAE (mg/L) for non-measurement nodes and tanks in case study 2, using IR-EnKF [*n*=50,*m*=60]

Tables

Table 1. Performance of EnKF algorithms for different scenarios where C_0 is uncertain

No. of Measurements	Method	No. of Realizations, *n*	AAE (mg/L)	AESD(mg/L)
8	NIR-EnKF	20	0.071	0.159
	NIR-EnKF	50	0.064	0.167
	NIR-EnKF	100	0.06	0.17
8	IR-EnKF	20	0.065	0.124
	IR-EnKF	50	0.058	0.126
	IR-EnKF	100	0.058	0.131
4-A	NIR-EnKF	20	0.078	0.173
	NIR-EnKF	50	0.076	0.183
	NIR-EnKF	100	0.069	0.181
4-B	NIR-EnKF	20	0.124	0.185
	NIR-EnKF	50	0.112	0.192
	NIR-EnKF	100	0.111	0.193

Table 2. Average Absolute Error (AAE (mg/L)) calculated for different scenarios and α values for dealing with measurement ambiguity during data assimilation

Parameter	α Value	Measurement error (AAE in mg/L)	Measurement noise (AAE in mg/L)
	1	0.083	0.239
	0.1	0.358	0.375
	0.01	0.225	0.2
C_0	0.001	0.566	0.334
	1	0.063	0.064
	0.1	0.07	0.058
	0.01	0.138	0.126
k_w	0.001	0.078	0.146

Table 3. Performance of EnKF algorithms for different scenarios where $k_w$is uncertain

No. of Measurements	Method	No. of Realizations, *n*	AAE (mg/L)	AESD(mg/L)
8	NIR-EnKF	20	0.175	0.075
	NIR-EnKF	50	0.157	0.08
	NIR-EnKF	100	0.157	0.08
8	IR-EnKF	20	0.056	0.075
	IR-EnKF	50	0.055	0.079
	IR-EnKF	100	0.054	0.08
4-A	IR-EnKF	20	0.063	0.081
	IR-EnKF	50	0.06	0.082
	IR-EnKF	100	0.06	0.085
4-B	IR-EnKF	20	0.08	0.101
	IR-EnKF	50	0.71	0.105
	IR-EnKF	100	0.69	0.107

Table 4. Computed mean value of the parameters at the end of simulation for case study 1 (n:20) for different scenarios (Method inside the bracket is the method used for theparameter estimation)

Parameter	True Value	NIR-EnKF	IR-EnKF	Measurement set A (Method)	Measurement set B (Method)	Measurement Error (Method)	Measurement Noise (Method)	Literature (Inverse Modelling)
C_0 (mg/L)	1.15	1.124	1.087	1.147 (NIR-EnKF)	1.297 (NIR-EnKF)	1.169 (NIR-EnKF)	1.131 (NIR-EnKF)	Not Available
k_w (m/d)	-0.3	-1.538	-0.272	-0.211 (IR-EnKF)	-0.347 (IR-EnKF)	-0.228 (IR-EnKF)	-0.276 (IR-EnKF)	-0.365*

***Munavalli and Kumar, 2005**

Table 5. Computed mean value of the parameters at the end of simulation for case study 2 (n:50) for different scenarios

Parameter	Unit	Observed Value	Computed Ensemble Mean	Inverse Modelling
C_0	mg/L	0.75	0.753	0.71[#]
k_{w1}	m/d	-1	-0.842	-1.1066*
k_{w2}	m/d	-0.75	-0.766	-0.7993*
k_{w3}	m/d	-0.5	-0.503	-0.4924*
k_{w4}	m/d	-0.25	-0.327	Not Available

***Munavalli and Kumar, 2003**
[#] Munavalli, 2002

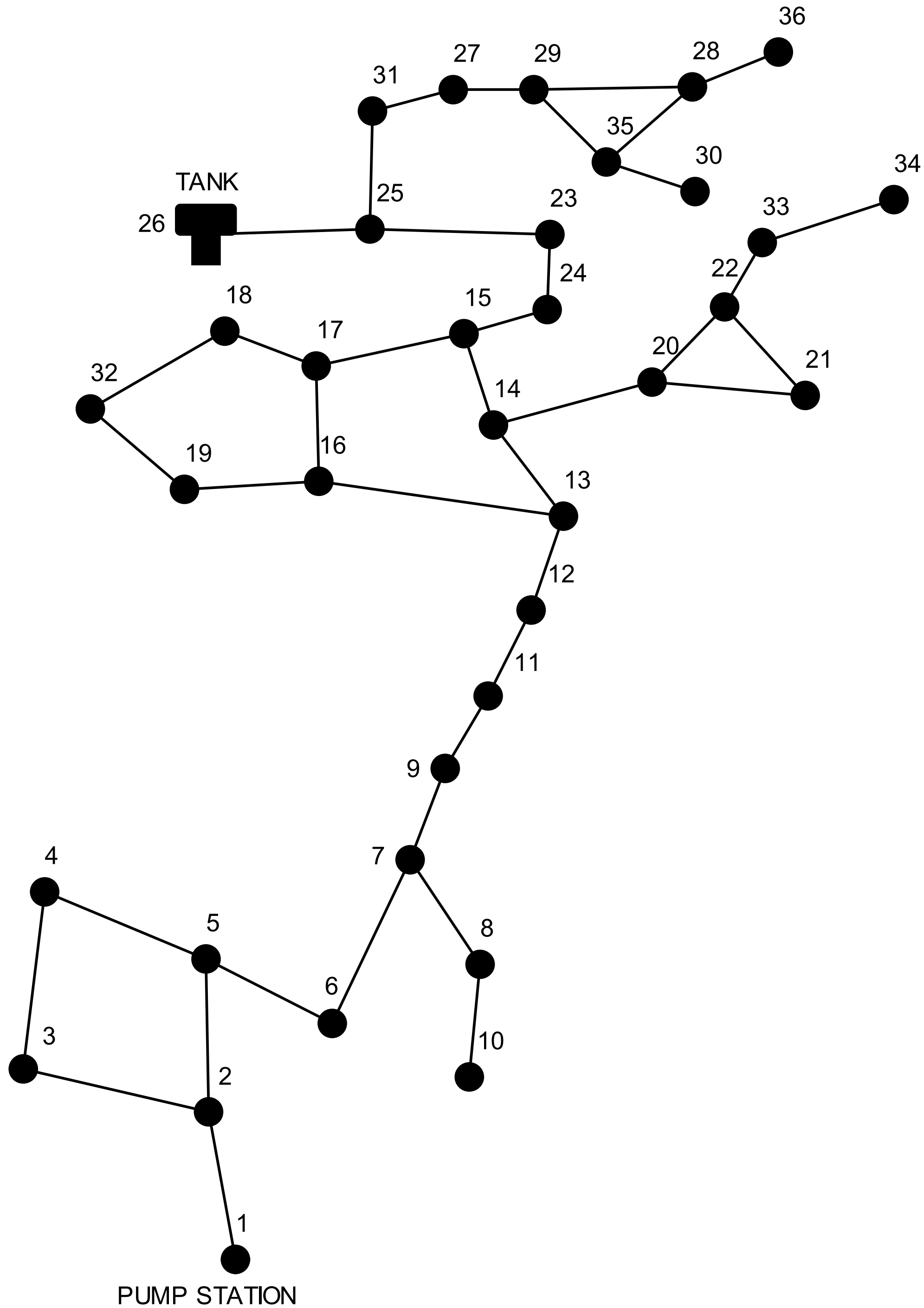

36
31
27
29
28
35
30
34
TANK
25
23
33
26
24
22
18
15
17
20
21
32
14
19
16
13
12
11
9
7
4
5
8
6
3
10
2
1
PUMP STATION

Figure 2

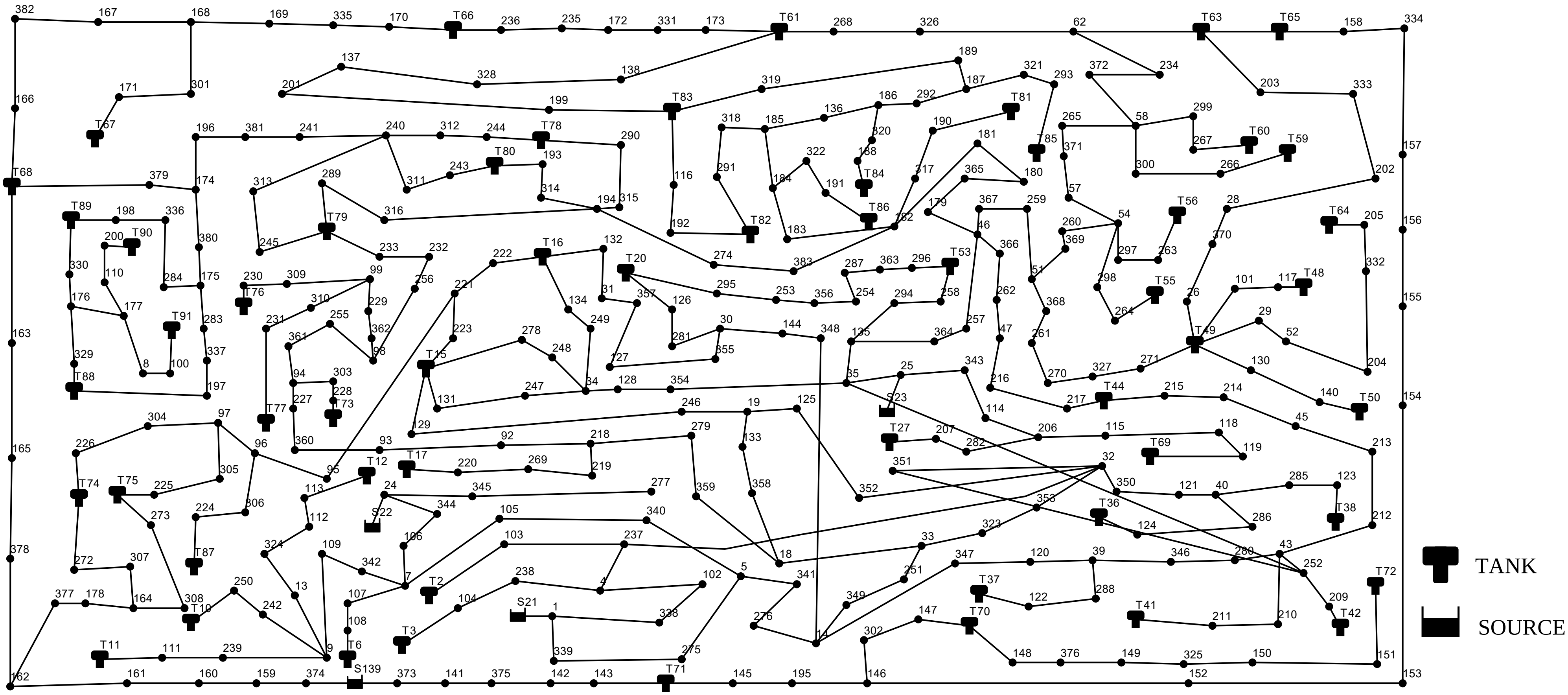
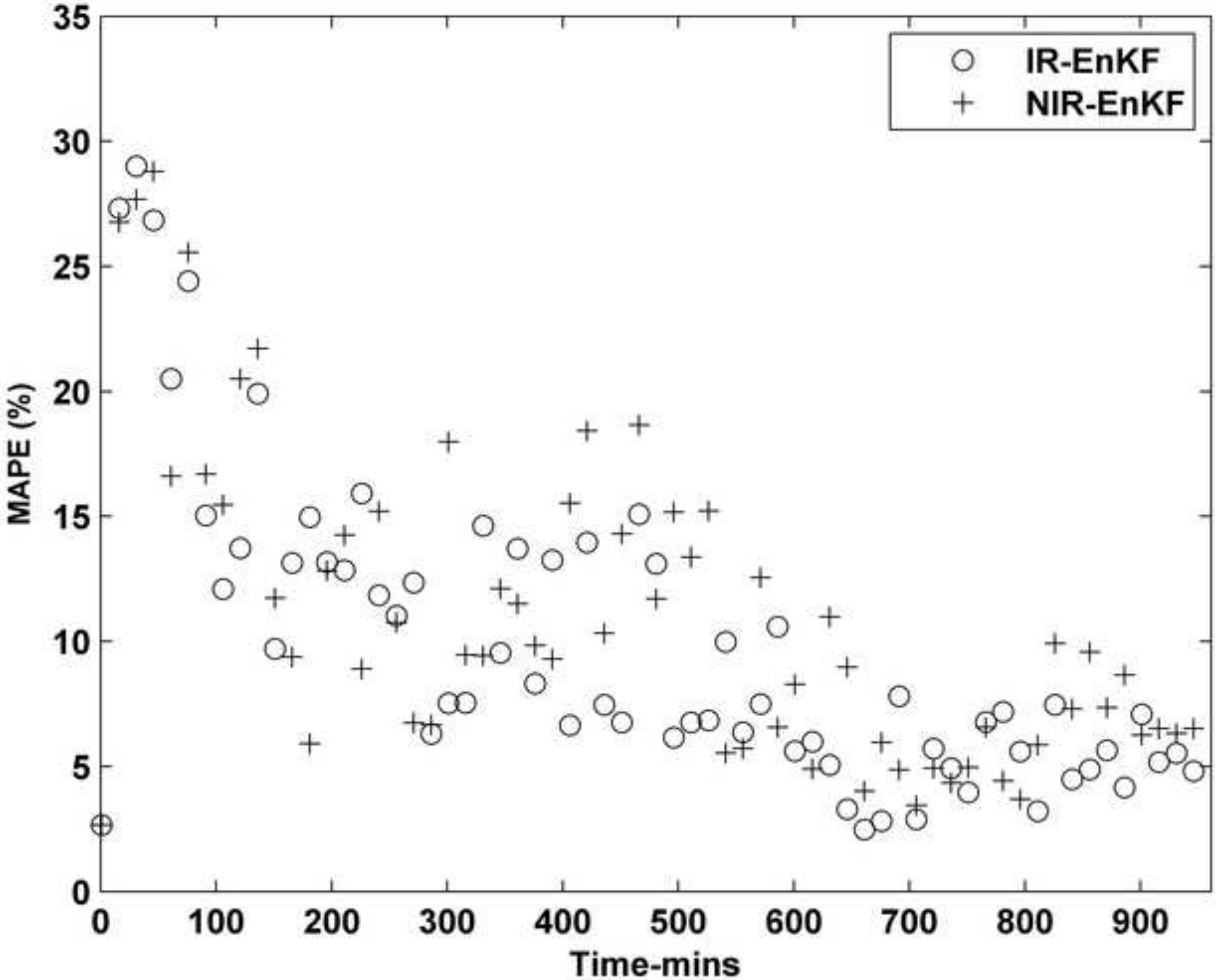


Figure 3



IR-EnKF
NIR-EnKF
MAPE (%)
Time-mins
0
5
10
15
20
25
30
35
0
100
200
300
400
500
600
700
800
900

Figure 4

Click here to access/download;Figure;Figure4.tif

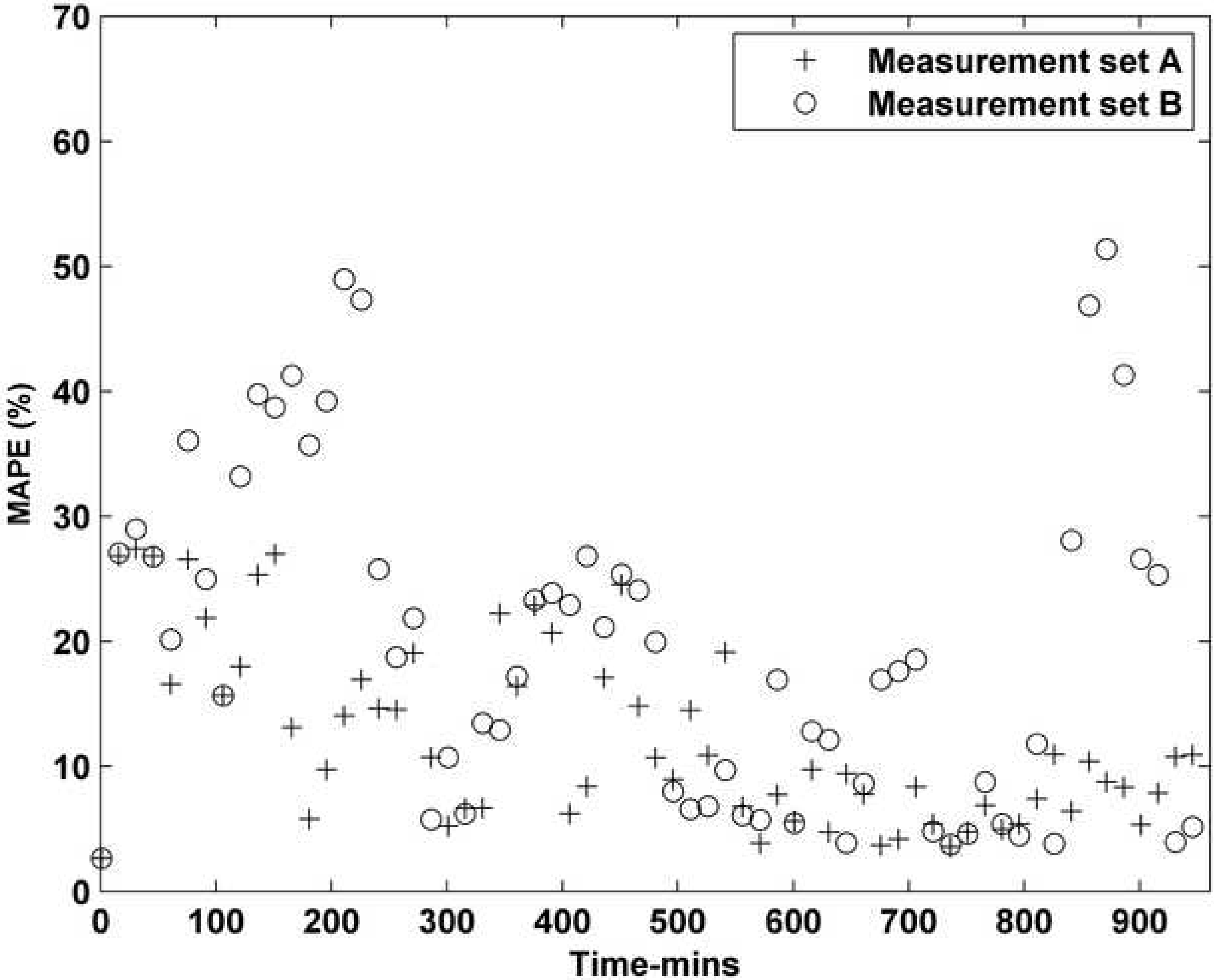
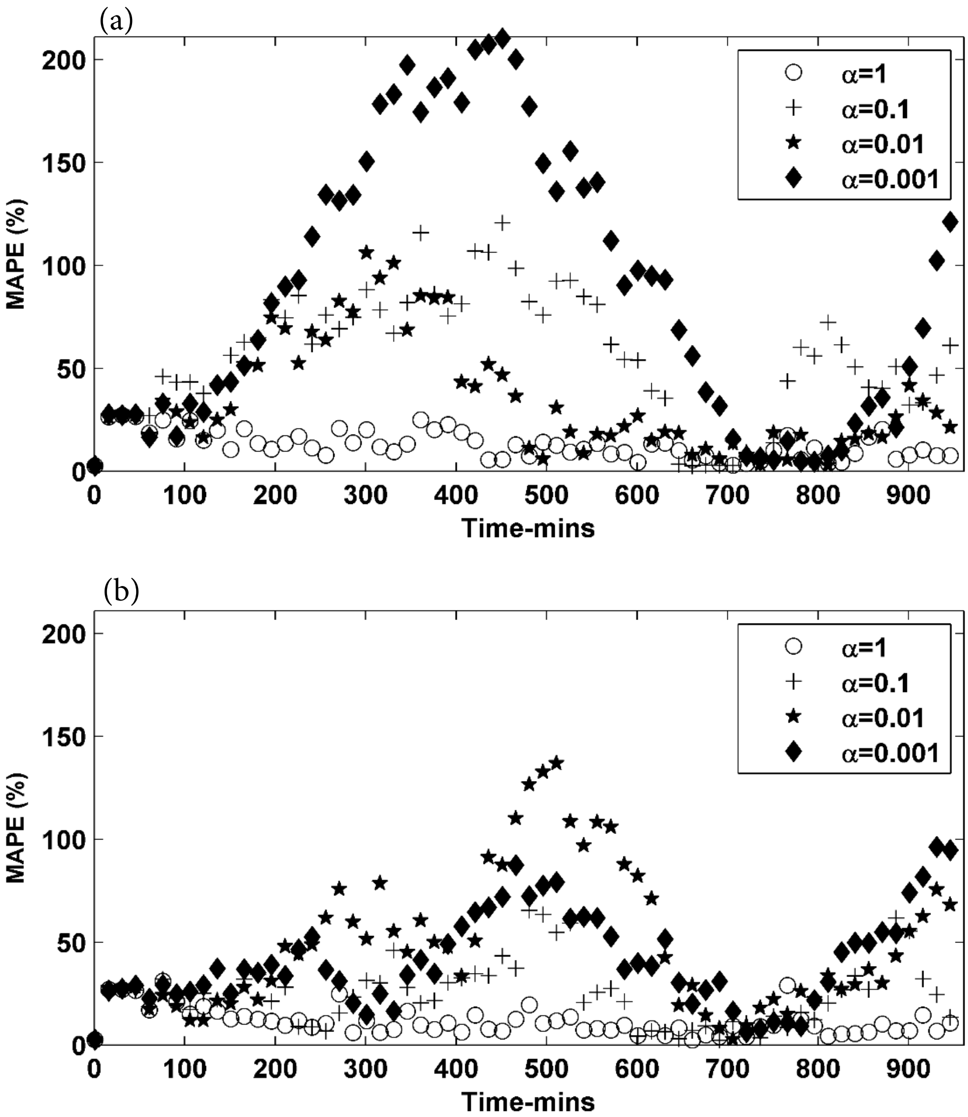


Figure 5



(a)
MAPE (%)
Time-mins
0
50
100
150
200
0
100
200
300
400
500
600
700
800
900
α=1
α=0.1
α=0.01
α=0.001
(b)
MAPE (%)
Time-mins
0
50
100
150
200
0
100
200
300
400
500
600
700
800
900
α=1
α=0.1
α=0.01
α=0.001

Figure 6

Click here to access/download;Figure;Figure6.tif

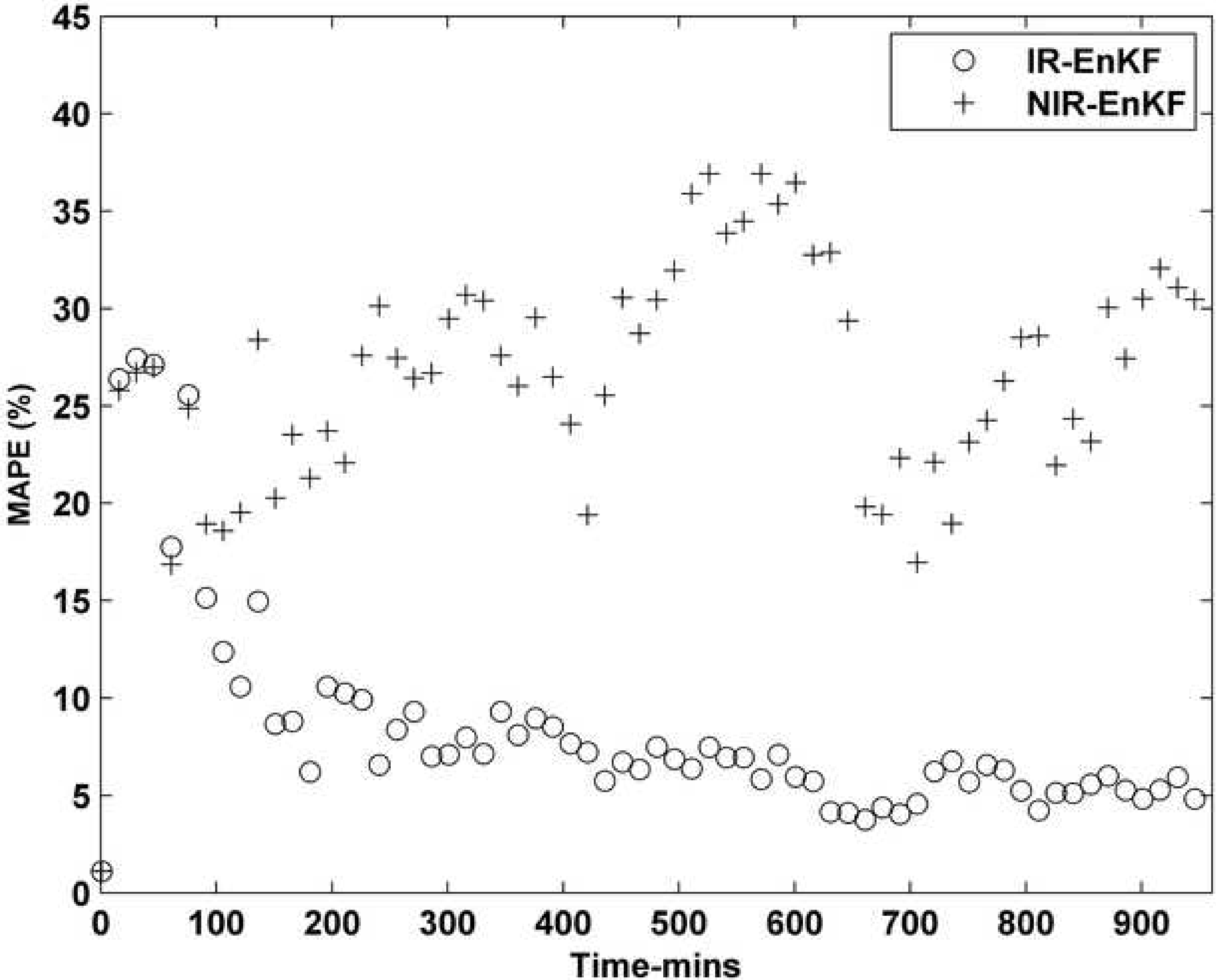


Figure 7

Click here to access/download;Figure;Figure7.tif

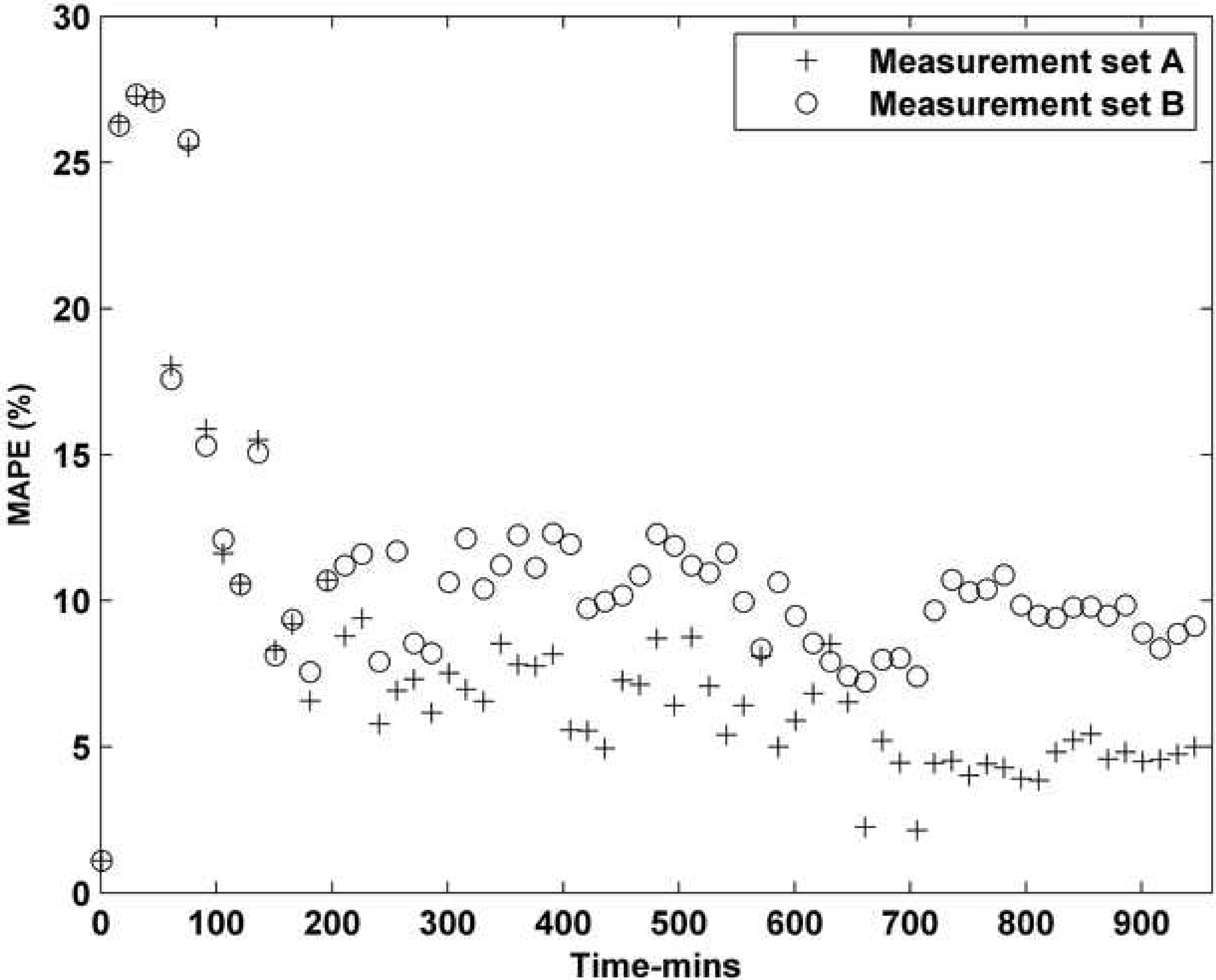


Figure 8

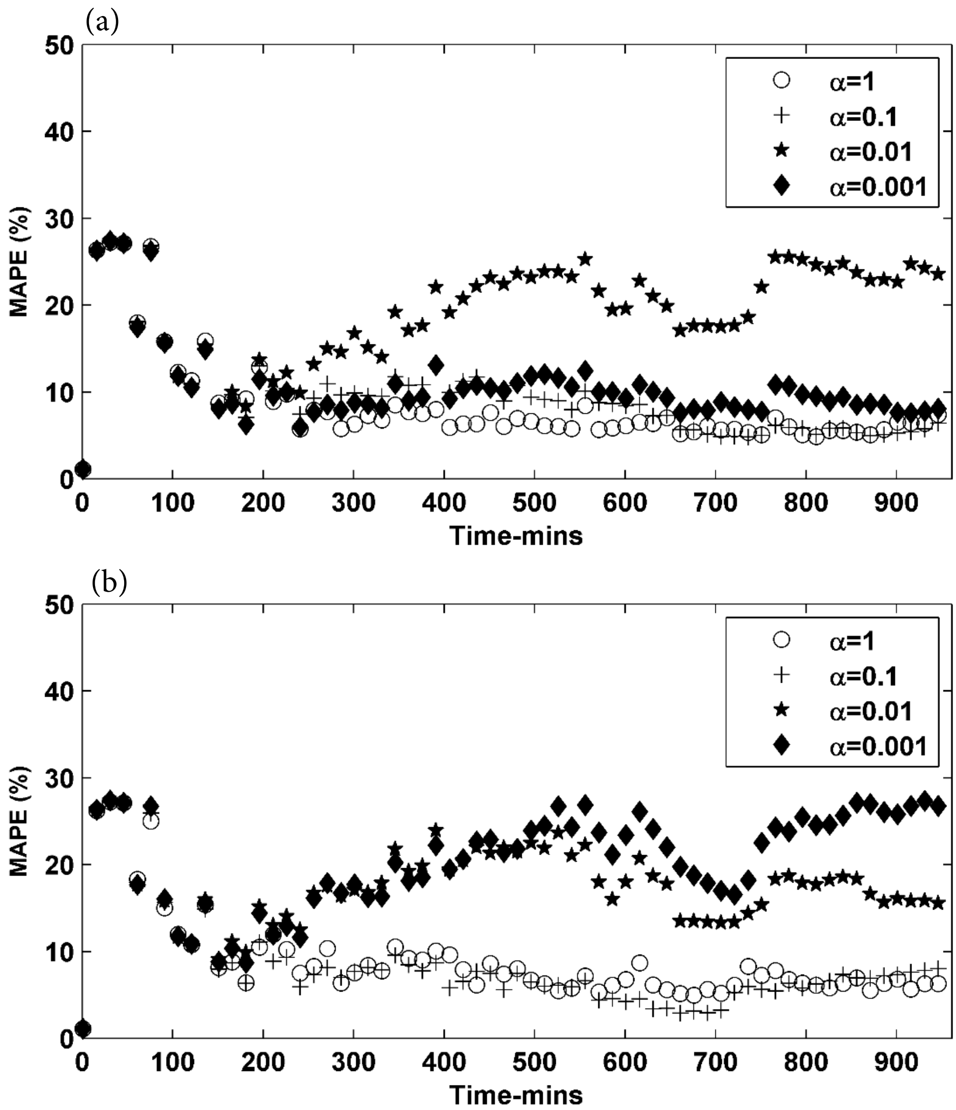
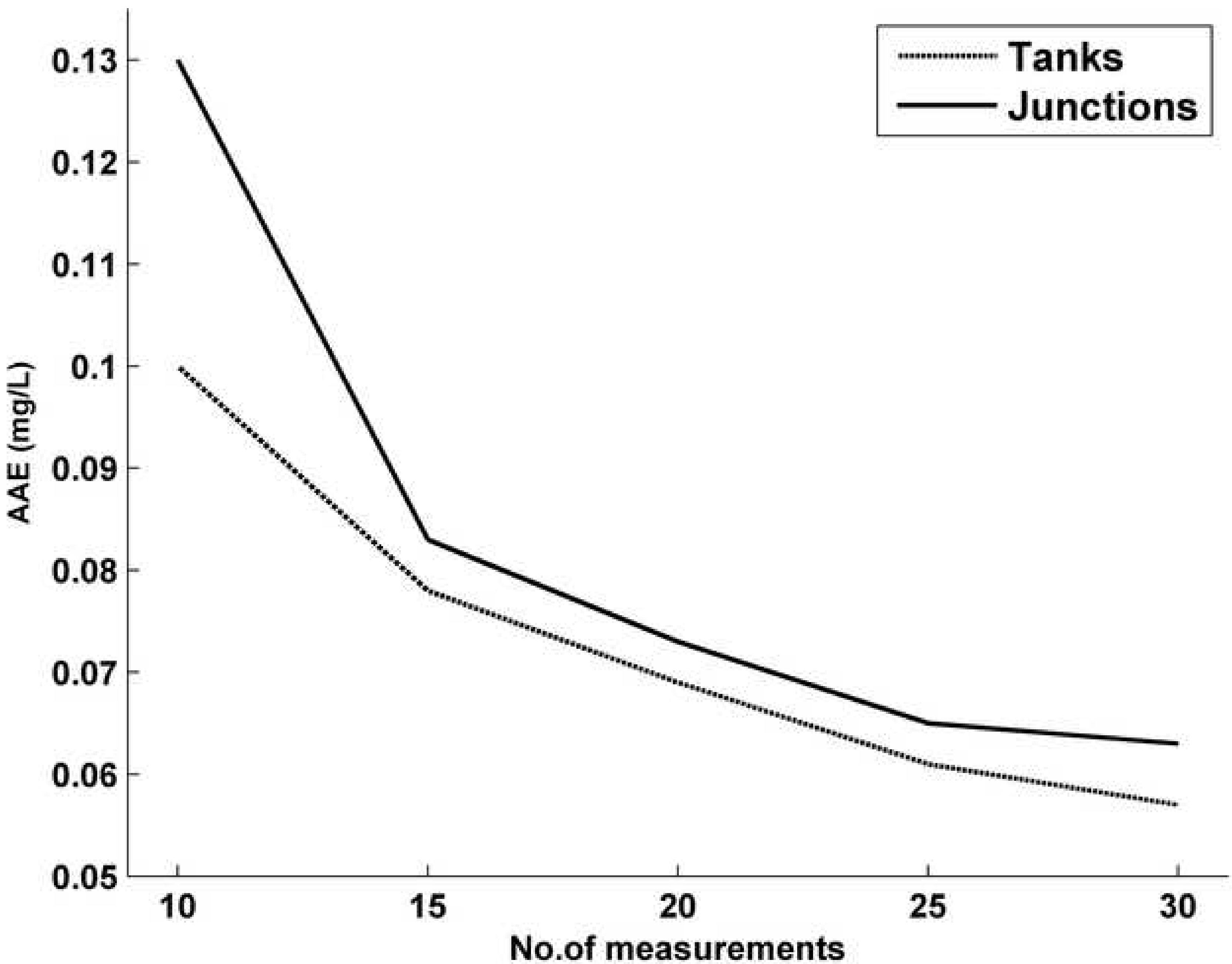

(a)
50
40
30
20
10
0
MAPE (%)
0
100
200
300
400
500
600
700
800
900
Time-mins
α=1
α=0.1
α=0.01
α=0.001
(b)
50
40
30
20
10
0
MAPE (%)
0
100
200
300
400
500
600
700
800
900
Time-mins
α=1
α=0.1
α=0.01
α=0.001

Figure 9

Click here to access/download;Figure;Figure9.tif



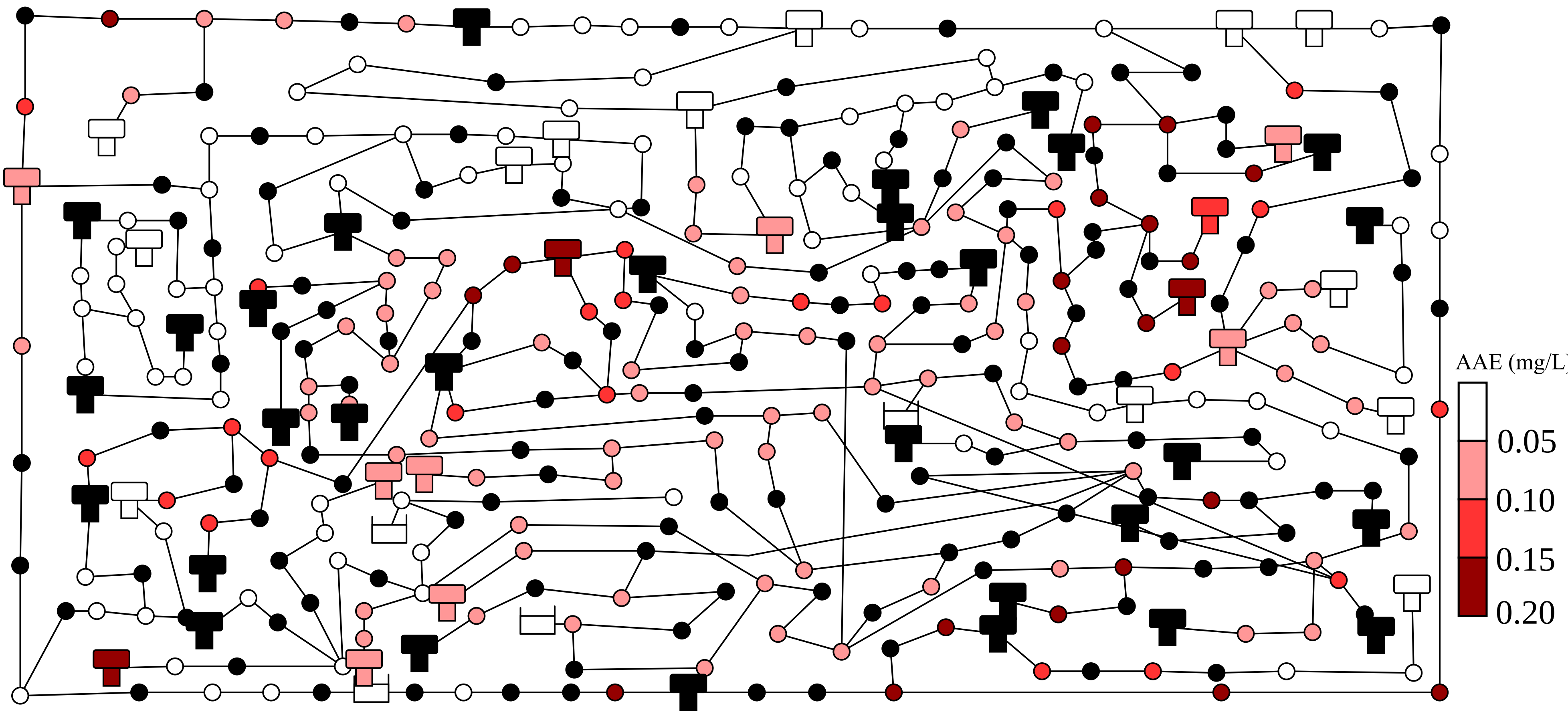

AAE (mg/L)
0.05
0.10
0.15
0.20

Figure 11