Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2015.1065769

57 (2016) 11592–11598
May

Taylor & Francis
Taylor & Francis Group

Chlorine modeling in water distribution networks using ARX and ARMAX model structures

I. Ethem Karadirek[a], Selcuk Soyupak[b], Habib Muhammetoglu[a,*]

[a]*Faculty of Engineering, Department of Environmental Engineering, Akdeniz University, Antalya, Turkey, Tel. +90 2423106327; Fax: +90 2423106306; emails: ethemkaradirek@akdeniz.edu.tr (I.E. Karadirek), muhammetoglu@usa.net (H. Muhammetoglu)*
[b]*Faculty of Engineering, Department of Civil Engineering, KTO Karatay University, Turkey, Tel. +90 3322217207; Fax: +90 3322020044; email: selcuk.soyupak@karatay.edu.tr*

Received 18 March 2015; Accepted 15 June 2015

ABSTRACT

Chlorination is still the main method of disinfection worldwide. Utilization of chlorine modeling enhances the management of water supply systems toward the reduction of the risks of waterborne diseases and cancer from disinfection byproducts formation. In this paper, the results of chlorine modeling efforts were presented as an extension of a project for Konyaalti Water Distribution Network (KWDN) of Antalya City, in the south of Turkey, using autoregressive with exogenous input (ARX) and autoregressive moving average with exogenous input (ARMAX) model structures. The required data-sets were obtained from the existing online monitoring stations. The ARX and ARMAX model structures modified for time series applications were utilized to predict chlorine concentrations at the critical point of KWDN. Non-representative data-sets were initially identified and excluded from the database. Best fit and Akaike's Final Prediction Error techniques were used as model selection criteria. ARX4-5-3 and ARMAX2-3-3-4 were identified to be the best ARX and ARMAX model structures among several structures tested to predict chlorine concentrations at the critical point of KWDN. This study shows that ARX and ARMAX model structures can be considered as potential for managing chlorine levels in water distribution networks especially when the properties of the components and hydraulics of water distribution network are unknown.

Keywords: ARX and ARMAX model structures; Chlorine modeling; Dynamic modeling; System identification; Water distribution network

1. Introduction

Drinking water should be disinfected for both inactivating micro-organisms and providing a residual against possible contamination of water in water distribution networks (WDNs). Chlorination is the most common disinfection method as it is cheap, effective, widely available, and easy to apply. Many waterborne diseases resulting from drinking water contamination might be eliminated using chlorine as a disinfectant [1].

*Corresponding author.

Presented at the 12th International Conference on Protection and Restoration of the Environment (PRE XII) 29 June–3 July 2014, Skiathos Island, Greece

Chlorine reacts with both organic and inorganic substances in water. Absence of chlorine or low chlorine concentration in WDNs increases the risk of waterborne outbreaks in case of contamination while high chlorine concentrations are associated with the formation of disinfection byproducts that some of them have cancer risks as well as other chronic and acute adverse health effects to human beings and animals [2]. Consequently, chlorine concentrations should be kept within certain limits to minimize the health risks. Free residual chlorine (FRC) concentrations in WDNs should not be less than 0.2 mg/L and the concentrations of chloroform, bromoform, dibromochloromethane, and bromodichloromethane must not exceed the values of 300, 100, 100, and 60 μg/L, respectively [3].

Several deterministic models have been developed for quantification of chlorine decay in WDNs, assuming first-order reaction kinetics [4–9]. However, chlorine reactions and their kinetics in actual distribution systems which are not well understood cannot adequately be represented by simple process-based deterministic models [10,11]. Biofilm formation and corrosion are some of the important factors of difficulty and uncertainty for deterministic modeling approach of chlorine concentrations [12]. Furthermore, all physical properties of the components of WDNs must be known precisely and an accurate and a well-calibrated hydraulic model is a prerequisite for deterministic water quality modeling [13–15]. Due to the aforementioned difficulties, artificial neural networks (ANNs), one of the data-driven methods using historical data of the system, has been utilized to predict water quality parameters in WDNs [16,17]. System identification is a methodology for mathematical modeling of dynamic systems that uses input and output signals of the system for prediction of chlorine concentrations in WDNs [16]. Mathematical modeling of dynamic systems based on system identification has come up as an alternative to ANNs model approach as it reduces the number of system input variables. The aim of this approach is to estimate the values of the parameters for selected model structure [18,19].

System identification is an iterative process where models are identified with different model structures from data [19]. The autoregressive with exogenous input (ARX) and autoregressive moving average with exogenous input (ARMAX) linear models have been widely used for system identification processes [20–22]. Rodriguez and Serodes [16] conducted a study to predict chlorine concentrations using ANN and ARX model structures in Quebec, Canada. The study showed that ARX and ANN models had similar prediction results in general; however, the performance of ANN model was better for estimating chlorine concentrations for specific conditions such as very high and low chlorine levels. Huusom et al. [23] reported that ARX model structures were more suitable than ARMAX and state—space model structures for control issues.

Kara et al. [24] carried out a leakage modeling and Karadirek et al. [14] carried out a chlorine modeling study in Konyaalti Water Distribution Network (KWDN) of Antalya City using the well-known hydraulic and water quality model—EPANET. ARX and ARMAX model structures based on system identification of dynamic systems were utilized to control FRC in KWDN [22]. This study is an extension of the previous ARX and ARMAX modeling efforts of the authors. The primary goal of the current study was to obtain additional information about performance and predictive capabilities of these model structures for full-scale WDNs using online continuous monitoring.

2. Material and methods

2.1. Study area

Konyaalti Water Distribution Network (KWDN) which serves about 60,000 people was selected as the study area. KWDN is operated independently from the rest of the city and is one of the major subnetworks of Antalya WDN. The raw water is extracted from five groundwater wells in *Bogacay* region and supplied to the network from *Bogacay Pumping Station* after disinfection process using sodium hypochlorite solution. There is no need for water treatment as the quality of the water abstracted meets the drinking water quality standards set by the relative legislation in Turkey [14,25]. The KWDN has about 200 km pipe network with different pipe materials and diameters and only one balancing reservoir namely *Hurma Balancing Reservoir* with a 15,000 m^3 storage capacity which is used to balance hourly water demand with supply in the region [14,25].

2.2. Online monitoring

There is a supervisory control and data acquisition system (SCADA) infrastructure in the study area [14,25]. The study area was divided into 18 district metered areas (DMAs) which allow easy management applications of both water quantity and quality in the area [25]. Each DMA was equipped with sensors for online continuous measurements of flow rates and water pressures. There were also additional water quality sensors at the chlorine dosing station, balancing reservoir and at the entrance of six DMAs for online continuous measurements of pH, FRC, turbidity, electrical conductivity, and water temperature [14]. The water quality monitoring station, ON-68, was used for ARX and ARMAX modeling efforts in another publication by the authors [22]. ON-68 exhibited relatively low chlorine concentrations, in comparison to the other monitoring stations, because it is the farthest online monitoring station from the chlorine dosing station. That is why the authors considered ON-68 as a critical point in their previous study [22]. Afterwards, deterministic modeling approach, using the well-known hydraulic and water quality model—EPANET, was utilized for management of chlorine concentrations in KWDN [14]. In that particular study, the critical point of chlorine concentrations of KWDN was determined (pointed as "*C*" in Fig. 1) by applying EPANET model taking into consideration the temporal and spatial changes of the model parameters all around the year [14]. The critical point, *C*, always showed the lowest chlorine levels and highest water ages in KWDN [14]. The critical point of chlorine concentrations of KWDN was equipped with a chlorine analyzer in addition to existing water quality monitoring stations to conduct this study. The additional chlorine analyzer was equipped with a data logger which allows set time intervals of choice of chlorine measurements. Main components of KWDN, existing water quality monitoring stations and the location of critical point are depicted in Fig. 1.

2.3. Modeling approach

The ARX and ARMAX model structures, modified for time series applications, were utilized to predict chlorine concentrations at the critical point of KWDN. MATLAB R2011a System Identification Toolbox was used for all model efforts presented in this paper. Modeling strategy consists of the identification of autoregressive and exogenous parameters of ARX model structure, and of autoregressive, moving average and exogenous inputs of ARMAX model structure. Model structures were developed to predict chlorine concentrations for five steps ahead in the future (that is $5 \times \Delta t = 75$ min) utilizing the data-sets collected with time step (Δt) of 15 min for subsequent measurements.

For a single input—single output system, the ARX model is defined as Eq. (1) [19]:

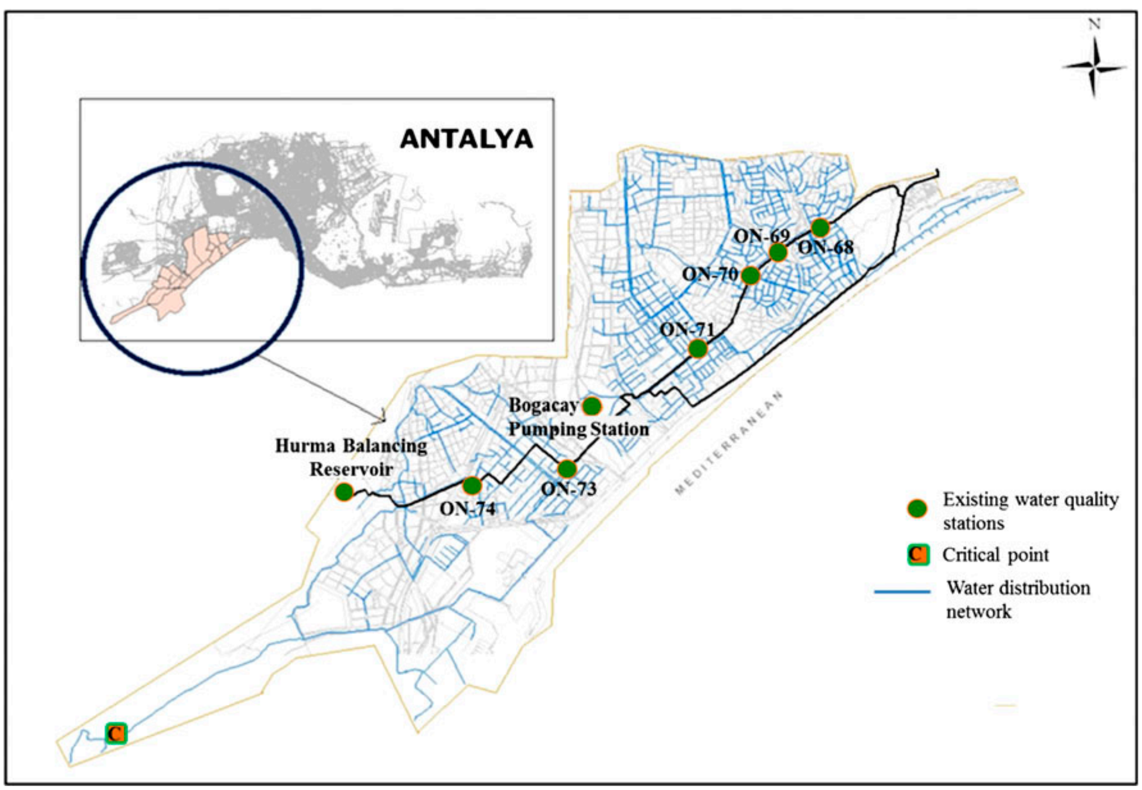

Fig. 1. Main components of KWDN with SCADA water quality monitoring stations (ON), critical point, balancing reservoir, and pumping station [Updated from 14].

$$y(t) + a_1 y(t-1) + ... + a_{na} y(t - na) \\ = b_1 u(t - nk) + ... + b_{nb} u(t - nb - nk + 1) + e(t) \quad (1)$$

where na and nb are the orders of the ARX model (na: Number of poles; nb: Number of zeroes plus 1); nk is the delay (number of input samples that occur before the input affects the output, also called the *dead time* in the system); $y(t)$: output at time t; $y(t-1)$ … $y(t-na)$: previous outputs on which the current output depends; $u(t-nk)$ … $u(t-nk-nb+1)$: previous and delayed inputs on which the current output depends and $e(t)$:white-noise disturbance value.

A more compact way to write the difference equation is

$$A(q)y(t) = B(q)u(t - nk) + e(t) \quad (2)$$

where q is the delay operator. Specifically,

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_{na} q^{-na} \quad (3)$$

$$B(q) = b_1 + b_2 q^{-1} + \ldots + b_{nb} q^{-nb+1} \quad (4)$$

where q is the delay operator; A and B are polynomials; a_1 … a_{na} and b_1 … b_{nb} are the parameters of the polynomials.

For a single input—single output system, the ARMAX model is defined as Eq. (5) [19]:

$$y(t) + a_1 y(t-1) + \ldots + a_{na} y(t - na) \\ = b_1 u(t - nk) + \ldots + b_{nb} u(t - nk - nb + 1) \\ + c_1 e(t-1) + \ldots + c_{nc} e(t - nc) + e(t) \quad (5)$$

where na and nb are the orders of the ARMAX model (na: Number of poles; nb: Number of zeroes plus 1); nc is the number of poles for the disturbance model; nk is the delay (number of input samples that occur before the input affects the output, also called the *dead time* in the system); $y(t)$: output at time t; $y(t-1)$ … $y(t-na)$: previous outputs on which the current output depends; $u(t-nk)$ … $u(t-nk-nb+1)$: previous and delayed inputs on which the current output depends; (a_1 … a_n), (b_1 … b_n), and (c_1 … c_n) are the parameters of the polynomials A, B and C; and $e(t-1)$ … $e(t-nc)$, $e(t)$: White-noise disturbance values.

The parameters na, nb, and nc are the orders of the ARMAX model, and nk is the delay. q is the delay operator. Specifically,

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_{na} q^{-na} \quad (6)$$

$$B(q) = b_1 + b_2 q^{-1} + \ldots + b_{nb} q^{-nb+1} \quad (7)$$

$$C(q) = 1 + c_1 q^{-1} + \ldots + c_{nc} q^{-nc} \quad (8)$$

Data-sets obtained from continuous online measurements of chlorine concentrations both at chlorine dosing station and critical point of KWDN were separated into two different data-sets for model calibration and verification processes. Pre-processing of data-sets, which is essential in modeling processes, was applied to observed data. Non-representative data due to inadequate monitoring response and/or maintenance of chlorine analyzers were identified by analyzing temporal series of data and excluded from database. Best fit (BF) and Akaike's Final Prediction Error (FPE) were used as model selection criteria. FPE provides a measure of model quality by simulating the situation where the model is tested on a different data-set. The most accurate model has the smallest FPE according to the FPE theory and it is defined as Eq. (9) [19]:

$$\mathrm{FPE} = V\left(\frac{1 + \frac{d}{N}}{1 - \frac{d}{N}}\right) \quad (9)$$

where V represents the loss function: The value of the identification criterion at the estimate which is equal to the determinant of the covariance matrix of the prediction errors.

BF is defined as Eq. (10) [19]:

$$\mathrm{BF}\ (\%) = \left|1 - \frac{|y - \hat{y}|}{|y - \bar{y}|}\right| \times 100 \quad (10)$$

where y represents the measured output, $\hat{y}$ denotes predicted model output and $\bar{y}$ is the mean of $\bar{y}$. Hundred percent indicates a perfect fit while 0% indicates that the fit is no better than guessing the output to be a constant ($\hat{y} = \bar{y}$) [19].

3. Results and discussion

3.1. Online monitoring

Continuous online chlorine analyzers helped in acquiring reliable numerous data-sets for system identification dynamic models. The critical point of chlorine concentrations in KWDN was determined by applying deterministic modeling approach [14]. Chlorine concentrations at both chlorine dosing station and the critical point (pointed as "C" in Fig. 1) were

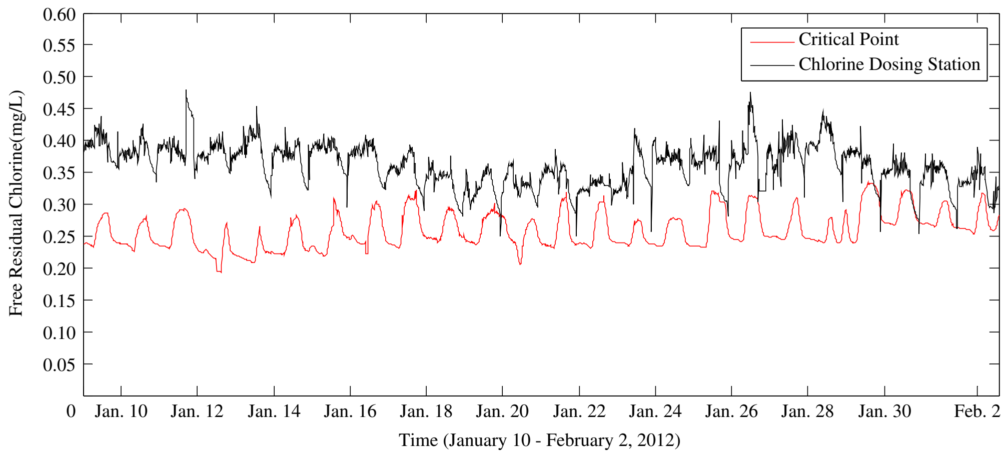

Fig. 2. Chlorine concentrations at the chlorine dosing station and the critical point between 10 January and 2 February, 2012.

continuously measured with an interval of 15 min from January 10–February 2, 2012 as depicted in Fig. 2.

3.2. Modeling results

The data-sets were separated into two parts, one part for calibration and the other part for verification processes. As the first step of modeling efforts, ARX and ARMAX model structures were identified. Many model structures were tested and the most accurate model structure was selected to predict chlorine concentrations at the critical point of KWDN. A comparison of performance criteria of some ARX and ARMAX model structures is given in Table 1.

ARX4-5-3 and ARMAX2-3-3-4, which are the most accurate ARX and ARMAX model structures based on performance criteria, were selected to predict chlorine concentrations at the critical point of KWDN. The least mean absolute error (MAE) of model predictions of chlorine concentrations was found as 0.022 mg/L for ARX4-5-3 model structure while it was found as 0.036 mg/L for ARMAX2-3-3-4 model structure. The MAE obtained from the previously conducted deterministic modeling approach by Karadirek et al. [14] was much higher than the MAE obtained in this study. Fig. 3 provides a comparison of ARX4-5-3 and ARMAX2-3-3-4 model predictions and observations of chlorine concentrations at the critical point of KWDN. It seems that there are differences between model predictions and field observations for both models, although they are not significant as appeared in FPEs and MAE values. The reasons of the observed delays for both models were attributed to the system's complexity and could not be explained by the authors within the scope of the available information about the system. According to the achieved results, ARX4-5-3 model structure was more capable of predicting chlorine concentrations than ARX2-3-3-4 at the critical

Table 1
Model performance criteria of some ARX model structures

Model structure	FPE	BF (%)
ARX4-5-3	0.9815 e-005	70.11
ARX4-5-1	0.9834 e-005	70.06
ARX4-7-1	0.9852 e-005	70.06
ARX4-5-2	0.9827 e-005	70.05
ARMAX2-3-3-4	0.9964 e-005	71.55
ARMAX2-3-3-3	0.9963 e-005	71.54
ARMAX2-3-3-2	0.9996 e-005	71.53
ARMAX3-2-2-2	0.9965 e-005	71.49

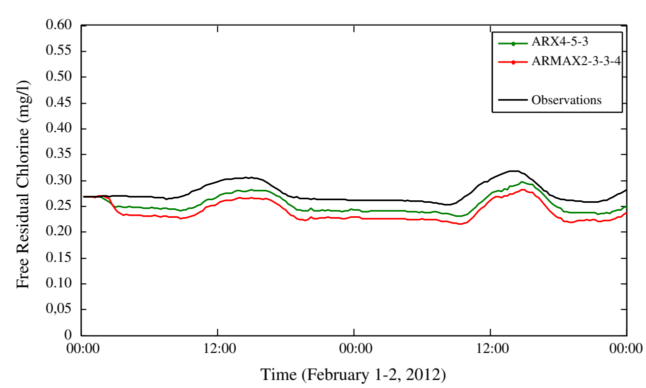

Fig. 3. Comparison of ARX4-5-3 and ARMAX2-3-3-4 model predictions with observations of chlorine concentrations at the critical point of KWDN for 48 h.

Table 2
FPE and BF values for one step and five steps ahead in the future for the critical points ON-68 and C

Critical point	Best of each model structure	FPE	BF (%)
ON-68	ARX38-7-3[a]	3.63 e −005	88.98
ON-68	ARMAX2-2-2-1[a]	5.63 e −005	86.09
C	ARX4-5-3[b]	0.9815e-005	70.11
C	ARMAX2-3-3-4[b]	0.9964 e-005	71.55

[a]Model structure was developed to predict chlorine concentrations at one step ahead in the future [22].
[b]Model structure was developed to predict chlorine concentrations at five steps ahead in the future.

point of KWDN. These results indicate that there may be a scope for better models to improve prediction accuracy.

4. Discussion and concluding remarks

The aim of the previous study of the authors [22] was to predict chlorine concentrations at one step ahead in the future while the model structures presented in this paper were developed to predict chlorine concentrations at five steps ahead in the future. The performance and accuracy of ARX and ARMAX models of this particular study and of the previous study [22] are presented in Table 2. in a comparative way. When the order of magnitudes of FPEs obtained for ARX and ARMAX models are compared, the current study yielded better precision. However, the previous study yielded better model performances, i.e. higher BF values. Although these conclusions seem contradictory, the complexities of the system justify the observed differences. Also, the calibration procedures and instrumentation besides model structures and horizons (number of steps ahead for prediction) were not the same for the two studies. Since the achieved results for both studies produced satisfactory performance values and satisfactory precision levels, both ARX and ARMAX model structures can be utilized in modeling chlorine at critical points of WDNs in the lack of information that enables deterministic modeling.

Acknowledgment

This study was supported by The Scientific and Technological Research Council of Turkey, TÜBİTAK (Project No. 107G088), Antalya Water and Wastewater Administration (ASAT) of Antalya Metropolitan Municipality and Akdeniz University, Antalya, Turkey.

References

[1] U.S. Centers for Disease Control and Prevention, Summary of Notifiable Diseases; Morbidity and Mortality Weekly Report. CDC Surveillance Summaries: Surveillance for Waterborne Disease Outbreaks—USA, 1997.
[2] W.D. King, L.D. Marrett, Case-control study of bladder cancer and chlorination by products in treated water (Ontario, Canada), Cancer Causes and Control 7 (1996) 596–604.
[3] WHO, Guidelines for Drinking Water Quality, fourth ed., WHO, Geneva, Switzerland, 2011.
[4] C.N. Haas, S.B. Karra, Kinetics of wastewater chlorine demand exertion, J. Water Pollut. Control Fed. 56(2) (1984) 170–173.
[5] P. Biswas, C. Lu, R.M. Clark, A model for chlorine concentration decay in pipes, Water Res. 27(12) (1993) 1715–1724.
[6] L.A. Rossman, The EPANET water quality model, in: Bryan Coulbeck (Ed.), Integrated Computer Applications in Water Supply, vol 2, John Wiley & Sons, Chichester, 1994, pp. 79–93.
[7] L.A. Rossman, R.M. Clark, W.M. Grayman, Modelling chlorine residuals in drinking-water distribution systems, J. Environ. Eng. 120(4) (1994) 803–820.
[8] V.K. Chambers, J.D. Creasey, J.S. Joy, Modelling free and total chlorine decay in potable water distribution systems, J. Water Supply: Res. Technol. - AQUA 44(2) (1995) 60–69.
[9] J.J. Vasconcelos, L.A. Rossman, W.M. Grayman, P.F. Boulos, R.M. Clark, Kinetics of chlorine decay, J. Am. Water Works Assoc. 89(7) (1997) 54–65.
[10] R.J. May, G.C. Dandy, H.R. Maier, J.B. Nixon, Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems, Environ. Modell. Softw. 23 (2008) 1289–1299.
[11] G.J. Bowden, J.B. Nixon, G.C. Dandy, H.R. Maier, M. Holmes, Forecasting chlorine residuals in a water distribution system using a general regression neural network, Math. Comput. Model. 44(5–6) (2006) 469–484.
[12] R.M. Clark, R.C. Haught, Characterizing pipe wall demand: Implications for water quality modeling, J. Water Res. PL-ASCE 131(3) (2005) 208–217.
[13] R.M. Males, W.M. Grayman, R.M. Clark, Modeling water quality in distribution systems, J. Water Res. PL-ASCE 114(2) (1988) 197–209.

[14] I.E. Karadirek, S. Kara, A. Muhammetoglu, H. Muhammetoglu, S. Soyupak, Management of chlorine dosing rates in urban water distribution networks using online continuous monitoring and modeling, Urban Water J. (in press), doi: 10.1080/1573062X.2014.992916.

[15] V. Kanakoudis, K. Gonelas, Accurate water demand spatial allocation for water networks modelling using a new approach, Urban Water J. 12(5) (2015) 362–379.

[16] M.J. Rodriguez, J.B. Serodes, Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems, Environ. Modell. Softw. 14 (1999) 93–102.

[17] J.B. Sérodes, M.J. Rodriguez, A. Ponton, Chlorcast©: A methodology for developing decision-making tools for chlorine disinfection control, Environ. Modell. Softw. 16 (2001) 53–62.

[18] L. Ljung, T. Glad, Modeling of Dynamic Systems, PTR PrenticeHall, Upper Saddle River, NJ, 1994.

[19] L. Ljung, System Identification Toolbox™ 7 User's Guide, The MathWorks Inc., 3 Apple Hill Drive Natick, MA, 2010, p. 531.

[20] L. Ljung, Issues in system identification, IEEE Control Syst. 11 (1991) 25–29.

[21] A. Chica, F. Rey, J. Aristizábal, Application of autoregressive model with exogenous inputs to identify and analyse patterns of solar global radiation and ambient temperature, Int. J. Ambient Energy 33(4) (2012) 177–183.

[22] S. Soyupak, H. 'Muhammetoğlu, System identification based dynamic models for control of free residual chlorine levels within urban water supply networks, in: Sixth International Perspective on Water Resources and Environment, IPWE 2013, Paper 185, Proceedings of ASCE Meeting, İzmir, Turkey, 2013.

[23] J.K. Huusom, N.K. Poulsen, S.B. Jørgensen, B.J. Jørgensen, ARX-model based model predictive control with offset-free tracking, Comput. - Aided Chem. Eng. 28 (2010) 601–606.

[24] S. Kara, I.E. Karadirek, A. Muhammetoglu, H. Muhammetoglu, Use of EPANET for leakage modeling in urban water distribution neworks: Antalya case study, in: Twelfth International Conference on Protection and Restoration of the Environment, Skiathos, Greece, June 29–July 3, 2015.

[25] I.E. Karadirek, S. Kara, G. Yilmaz, A. Muhammetoglu, H. Muhammetoglu, Implementation of hydraulic modelling for water-loss reduction through pressure management, Water Resour. Manage 26(9) (2012) 2555–2568.