Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-11 (2015) 072–077

Water Distribution Network Modeling Based on NARX

Xu Zhe, Yang Jie, Cai Huaqiang, Kong Yaguang, He Bishi

Institute of Information and Control, Hangzhou Dianzi University, Hangzhou, China, (e-mail: xuzhe@hdu.edu.cn)

Abstract: This paper proposes a NARX (Nonlinear Auto-Regressive with Exogenous Inputs) model for the water distribution network real-time prediction and control. The model estimates the time-variable nodal demand equivalently by exploiting the real-time and historical operating data and establishes a functional relationship between the major variables among the network. In addition, a training scheme with a combination of offline training and online training, and corresponding algorithm are proposed. And the NARX model is established for a real distribution network. The results demonstrate the model is applicative and satisfactory, and it shows good tracking and predicting performance.

Keywords: water distribution network, NARX, BP algorithm, forgetting factor.

1. INTRODUCTION

Water distribution network is generally composed of a large number of interconnected pipes, tanks, pumps, valves and other hydraulic elements which carry water to demand nodes from the supply areas, with specific pressure levels to provide a good service to consumers (Cembrano, et al., 2000). And it is characterized by a large variety and complexity, and in particular, their hydraulic dynamics are time-variant, spatially distributed and highly non-linear (Rao & Alvarruiz, 2007). An appropriate model of water distribution network is the basis for the control optimization procedure to improve the economic and/or energetic efficiency (Coelho & Andrade-Campos, 2014).

Water distribution network model generally can be classified into two categories which are microscopic model and macroscopic model (DeMoyer & Horwitz, 1975). Microscopic model, such as the full simulation model, is a kind of first principle model, primarily for the purpose of design, of which the network topology and the exact parameters of pipes and nodes are needed to be known, and the nodal demand must be well estimated in advance. Macroscopic model, such as the regression model, is based on the data-driven method, which can be obtained by constructing and maintaining from the telemetered operating data. It includes only major variables among the network, such as pump flows and pressures, tank flows and depths, and selected pressure points.

In the study of control optimization of water distribution networks in practice, since derivation of optimal control strategies on the basis of complex simulation models is difficult and time-consuming (Cembrano, et al., 2000) (Coelho & Andrade-Campos, 2014) (Razavi, et al., 2012) (Nazif, et al., 2010) (Broad, et al., 2009) (Paluszczyszyn, et al., 2013) (Shamir & Salomons, 2008), the surrogate model has developed recently which is concerned with developing and utilizing cheaper-to-run "surrogates" of the "original" simulation models (Razavi, et al., 2012), such as metamodels (Cembrano, et al., 2000) (Coelho & Andrade-Campos, 2014) (Nazif, et al., 2010) (Broad, et al., 2009) and simplified simulation model (Paluszczyszyn, et al., 2013) (Shamir & Salomons, 2008). Surrogate model is a second level of abstraction, a kind of data-driven model based on the simulation data which can also classified as macroscopic model.

However, the models for water distribution network often require accurate nodal demand data which has limited their application, especially in the context of real time (e.g. with minutes). On the one hand, there are few nodal demand measuring points primarily used for modeling in the network mainly from the perspective of economic efficiency, especially in developing countries such as China. On the other hand, the nodal demand possesses huge uncertainty so that it is difficult to be estimated exactly, especially in real time (Shen, et al., 2014) (Hutton, et al., 2012) (Hatchett, et al., 2010).

In view of the problem above, this paper propose a NARX model based on the theory of data-driven modeling, mainly, for water distribution network real-time prediction and control. The rest of the paper is organized as follows. Section 2 describes and deduces the mathematic model of water distribution network. Section 3 introduces the NARX of water distribution network. Section 4 is a practical case study of which the NARX model is established and analyzed. Finally, section 5 is the concluding remarks.

2. MATHEMATIC MODEL OF WATER DISTRIBUTION NETWORNK

2.1 Model in theory

Based on the control system theory, a description of the dynamic model of a water distribution network, in discrete time, is as follows (Cembrano, et al., 2000):

$$x(t+1) = f(x(t), u(t), d(t), \varepsilon(t)) \qquad (1)$$

It is a MIMO (Multiple Input Multiple Output) system. Where x is a vector of state variable, such as the nodal

10.1016/j.ifacol.2015.09.162

pressure, the pipe flow and the tank depth. u is a vector of control variable, such as the outlet pressure, outlet flow of pumping stations and the flow passing through tele-controlled valves. d is a vector of the nodal demand distributed in the network. $\{\varepsilon(t)\}$ denotes the stochastic disturbance sequence which is assumed to be independent random variables with zero mean values and finite variances. t denotes the instantaneous values at sampling time t. And f is a strongly nonlinear function containing the hydraulic characteristic of the whole network.

By dynamic system (1), if the control vector and state vector of the system are known at present, meanwhile the nodal demand and disturbance are well estimated, and then the states at next moment are identified. But the nodal demand is time-variant, spatially distributed, and it also has a nonlinear relationship with the nodal pressure (e.g. Most of the nodal outflows occur via orifices such as open taps or valves). Therefore, the nodal demand is the function of both time and pressure in the network. And it is difficult or impossible to be estimated exactly, especially in the context of real-time (e.g. within minutes). In view of the problem, we try to estimate the nodal demand indirectly.

Suppose that the system states at t are known, by (1) we can estimate $d(t-1)$ with $\hat{d}(t-1)$ as:

$$\hat{d}(t-1) = g[x(t), x(t-1), u(t-1)] \tag{2}$$

Where, g is an unknown function. And then, the nodal demand at next sampling time t can be predicted by a certain time series method, such as a linear or nonlinear auto-regression model. That is:

$$\hat{d}(t) = h[\hat{d}(t-1), \hat{d}(t-2), \cdots, \hat{d}(t-n_d)] \tag{3}$$

Where h is the regression function, n_d is the regression order. If denoting $g(i) = g(x(t), x(t-i), u(t-i))$, $i = 1, \cdots, n_d$, can be simplified as:

$$\begin{aligned}\hat{d}(t) &= h[g(1), \cdots, g(n_d)] \\ &= hg[x(t), \cdots, x(t-1-n_d), u(t-1), \cdots, u(t-1-n_d)]\end{aligned} \tag{4}$$

Where hg is an unknown function. And by (1), (2), (3) and (4), the water distribution system can be expressed as：

$$x(t+1) = f[x(t), u(t), \hat{d}(t)] \tag{5}$$

The dynamic model in this form does not explicitly contain the nodal demand, which is estimated equivalently by the system state and control information

2.2 Model in practice

As we know, real urban water distribution networks get complex topologies and a large number of nodes and interconnected pipes buried underground. The complexities and uncertainties make it difficult to create a complete and accurate mathematical model.

In practice, for a real water distribution system, we often only set pressure monitoring point at some key nodes, and flow monitoring point in some key pipes from the perspective of economic and modeling efficiency. And the readings of pressure and/or flow at these dominant points, to some extent, represent and describe the dynamics of the corresponding local/component area of the network. By integrating these local dynamics, the comprehensive dynamics of the entire network can be obtained. And when the system is running, the field readings are collected by SCADA (Supervisory Control And Data Acquisition) system and uploaded to the control center for decision making in real time.

According to the control theory, pressure and flow are the state variables of water distribution system, and the pressure and flow have been measured are the observable state variables, the rest are the unobservable state variables. Therefore, for a real water distribution system we operated in practice, its states are only partially observable or known. So we should make some changes on the theory model (5) when put it into practice.

Let the states of the system be partially observable, then the x in system (1) here is a vector of observable state variable, such as the pressure and flow measured at some dominant points and the tank depth. And in the formula (3) and (4), we have estimated the nodal demand $d(t)$ based on the state and control information of the system during the period between $t-1-n_d$ and t. But here since the states are partially known, the state information is incomplete. One possible way is to take the state and control information over a longer period in the past as information compensation. That is:

$$\hat{d}(t) = hg[X(t-n_a), U(t-n_b)] \tag{6}$$

$$X(n_a) = [x(t), x(t-1), \cdots, x(t-n_a)] \tag{7}$$

$$U(n_b) = [u(t-1), u(t-2), \cdots, u(t-n_b)] \tag{8}$$

Where, n_a is the length of the state information, n_b is the length of the control information, both n_a and n_b are integers and such that $n_a, n_b \geq n_d$.

In summary, by (5) and (6), the model of the water distribution network system with the states partially known is:

$$x(t+1) = f[x(t), u(t), hg(X(t-n_a), U(t-n_b))] \tag{9}$$

And (9) can be further simplified as:

$$x(t+1) = f[x(t), \cdots, x(t-n_a), u(t-1), \cdots, u(t-n_b)] \tag{10}$$

Due to f is a complex nonlinear function, an accurate mathematic model for it base on first principle is difficult or impossible. As we know, neural network model gets a very strong non-linear mapping capability, and a good robustness and stability as well. Therefore, in this paper, we consider an appropriate neural network model in place of the function f . And it is not difficult to find out that the form of the model expressed as (10) is quite in line with the NARX model. In other words, the states at a given instant are explained by the evolution of the states at previous instants as well as by the effect of exogenous variables which are the control inputs. Thus, a NARX model base on neural network for water distribution network is introduced in the following section.

3. NARX MODEL FOR WATER DISTRIBUTION NETWOR

3.1 Architecture of NARX

NARX model is widely used for nonlinear dynamic system identification and modeling, and there are two kinds of alternative architectures which are series-parallel architecture and parallel architecture (Anh & Nam, 2014) (Basso, et al., 2005) (Lee & Chang, 2009) (Kadri, 2012). For water distribution system, the analysis in section 2 has shown that the nodal demand is time-variable, thus we have to use the actual state and control information measured to estimate it in real time. Therefore, a series-parallel architecture is quite in line with the actual situation of the system, whether it is in the stage of identification or validation or application.

NARX model mainly consists of two modules, one is composed by Multilayer Perceptron (MLP) which has a strong nonlinear mapping ability, and the other is composed of learning and training methods for Multilayer Perceptron. Fig. 1 shows the framework of using NARX model to identify the water distribution network in the series-parallel architecture.

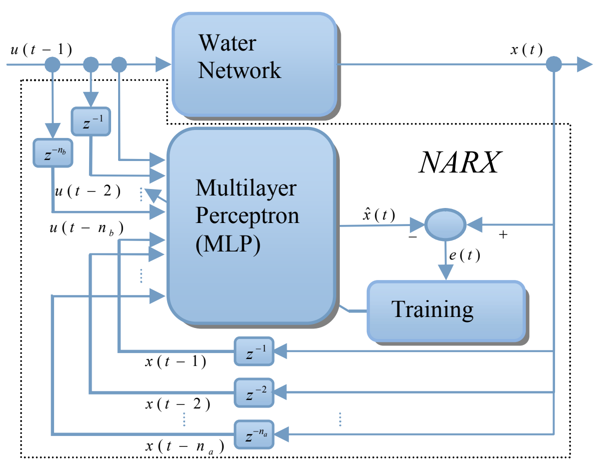

Fig.1. Framework of the identification of the water distribution network by NARX model.

Where, $\hat{x}(t)$ is the estimated value respect to the actual output $x(t)$ at sampling time t, and $e(t) \in R^{q\times 1}$ is the estimation error:

$$e(t) = [e_1(t), e_2(t), \cdots, e_q(t)]^T \quad (11)$$

In addition, the class of MLP considered in this paper is furthermore confined to those having only two layers. It is also known as the two-layered feed-forward neural (FFN) network which is widely used. And it has been proven in theory that the two-layered FFN network, with sigmoid transfer functions in the hidden layer and linear transfer functions in the output layer, can approximate virtually any function with arbitrary precision, provided sufficiently many hidden neurons are available (Hagan, et al., 1996). A fully connected two-layer FFN network with p inputs, n hidden neurons, and q outputs neurons is shown in Fig. 2.

The inputs of the FNN network are composed of the control vector u and the state vector x, as well as their delays. The inputs and outputs of the FNN network are respectively defined as:

$$Inputs(t) = [u(t-1), \cdots, u(t-n_b), x(t-1), \cdots, x(t-n_a)]^T \quad (12)$$

$$Outputs(t) = \hat{x}(t) \quad (13)$$

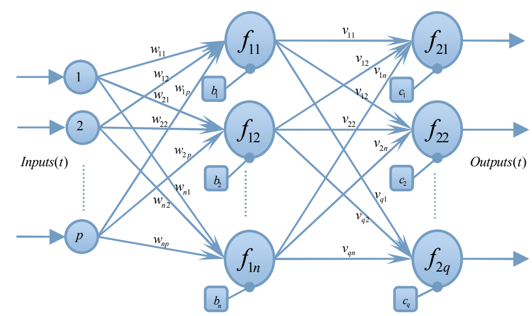

Fig.2. Fully connected two-layered feed-forward neural network

And $W \in R^{n\times p}$ is the connection weight matrix between the input layer and the hidden layer, $V \in R^{q\times n}$ is the connection weight matrix between the hidden layer and the output layer, where

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & \cdots & w_{np} \end{bmatrix}, \ V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ v_{q1} & v_{q2} & \cdots & v_{qn} \end{bmatrix} \quad (14)$$

Respectively, $b \in R^{n\times 1}$ and $c \in R^{q\times 1}$ are the bias vectors of the hidden layer and the output layer:

$$b = [b_1, b_2, \cdots, b_n]^T, \ c = [c_1, c_2, \cdots, c_q]^T \quad (15)$$

And, $F_1 \in R^{n\times 1}$ and $F_2 \in R^{q\times 1}$ are the transfer function vectors of the hidden layer and the output layer respectively:

$$F_1 = [f_{11}, f_{12}, \cdots, f_{1n}]^T, \ F_2 = [f_{21}, f_{22}, \cdots, f_{2q}]^T \quad (16)$$

In addition, the sigmoid function and the linear function are chosen respectively as the transfer functions for the hidden layer neurons and output layer neurons:

$$f_{1k}(net_{1k}) = \frac{1}{1+e^{-net_{1k}}}, \ f_{2i}(net_{2i}) = net_{2i} \quad (17)$$

Where, net_{1k} is the input of kth neuron at hidden layer, and $k = 1, 2, \cdots, n$, net_{2j} is the input of ith neuron at output layer, and $i = 1, 2, \cdots, q$.

In summary, the basic architecture of the two-layered FNN network can be described as:

$$Outputs(t) = F_2\{V(t)F_1[W(t)Inputs(t) + b(t)] + c(t)\} \quad (18)$$

3.2 Training Method

For water distribution network, we propose a training method with a combination of offline training and online training.

Offline training mode is used to establish the fundamental NARX model of water distribution network. Online training mode is a supplementary/enhancement for the fundamental NARX model. The basic training scheme is as follows.

Firstly, take an offline training for the NARX to obtain a fundamental model. Secondly, take an online training argorithm in place of the offline training algorithm to enhance the fundamental NARX. In other words, we firstly provide sufficient knowledge for the NARX under offline mode, and when the model works, we switch it to the online mode for self-learning.

(1) Offline Training

Offline training is mainly used to capture the main knowledge of water distribution network for the NARX model. Besides, it is also used to help to ascertain the structure of NARX model, such as the input vector and the neuron number of the three-layered FNN network in NARX.

Since there is a large amount of historical operating data in the SCADA system which contains abundant dynamics of the water distribution network, we can take full advantage of it with an appropriate offline training algorithm to obtain a fundamental NARX model of it efficiently. There are kinds of offline algorithms can be taken. In this study, Levenberg-Marqurd algorithm is selected.

(2) Online Training

Online training is a supplementary/enhancement for the fundamental NARX model. Although the fundamental NARX model has a sufficient knowledge of the dynamics of water distribution system, there exist some dynamic that the model haven't known. In other words, the training data for constructing the fundamental NARX model does not contain the entire behaviors of the system, which may cause a problem of incomplete identification.

Of course, if more historical operating data is taken to training our model, the problem may be avoided to some extent. But it is still impossible make the model omniscient about the dynamics of a real water distribution network just base on its limited historical operating data. So, an appropriate long-term consecutive learning/trianing mechanism for our model should be built. Meanwhile, the time-variable nodal demand needs a real-time updated estimation function.

Therefore, in order to overcome the shortcomings, an online training algorithm with an exponential forgetting mechanism is introduced (Yu, et al., 2007)]. The online training algorithm is based on the basic BP algorithm, so it is actually an iteration process by minimizing an error function with a forgetting factor:

$$E(t) = \frac{1}{2}\sum_{j=1}^{t}\lambda^{t-j}\sum_{i=1}^{q}e_i^2(j) = \frac{1}{2}\sum_{j=1}^{t}\lambda^{t-j}e^T(j)e(j) \tag{19}$$

Where, λ is the forgetting factor, and such that $0<\lambda<1$. By applying the steepest descent method to the error cost function $E(t)$, we can obtain the gradient of $E(t)$ with respect to the weight matrix W, V and the bias vectors b, c respectively:

$$\begin{aligned}\nabla_W E(t) &= \frac{\partial E(t)}{\partial W(t)} \\ &= -\sum_{j=1}^{t}\lambda^{t-j}\dot{F}_1(j)V^T(j)\dot{F}_2(j)e(j)Inputs^T(j)\end{aligned} \tag{20}$$

$$\nabla_V E(t) = \frac{\partial E(t)}{\partial V(t)} = -\sum_{j=1}^{t}\lambda^{t-j}F_1(j)e^T(j)\dot{F}_2(j) \tag{21}$$

$$\nabla_b E(t) = \frac{\partial E(t)}{\partial b(t)} = -\sum_{j=1}^{t}\lambda^{t-j}\dot{F}_1(j)V^T(j)\dot{F}_2(j)e(j) \tag{22}$$

$$\nabla_c E(t) = \frac{\partial E(t)}{\partial c(t)} = -\sum_{j=1}^{N}\lambda^{t-j}e^T(j)\dot{F}_2(j) \tag{23}$$

Where ∇ is the gradient operator, and $\dot{F}_1$, $\dot{F}_2$ such that:

$$\dot{F}_1 = diag[\dot{f}_{11}, \dot{f}_{12}, \cdots, \dot{f}_{1n}] \in R^{n\times n} \tag{24}$$

$$\dot{F}_2 = diag[\dot{f}_{21}, \dot{f}_{22}, \cdots, \dot{f}_{2q}] \in R^{q\times q} \tag{25}$$

Rewrite (20), (21), (22) and (23) in the recursive form as:

$$\nabla_W E(t) = \lambda\nabla_W E(t-1) - \dot{F}_1(t)V^T(t)\dot{F}_2(t)e(t)x^T(t) \tag{26}$$

$$\nabla_V E(t) = \lambda\nabla_V E(t-1) - F_1(t)e^T(t)\dot{F}_2(t) \tag{27}$$

$$\nabla_b E(t) = \lambda\nabla_b E(t-1) - \dot{F}_1(t)V^T(t)\dot{F}_2(t)e(t) \tag{28}$$

$$\nabla_c E(t) = \lambda\nabla_c E(t-1) - \dot{F}_2(t)e(t) \tag{29}$$

Then, the updated formulae of weight matrixs and bias vectors are given by, respectively:

$$W(t) = W(t-1) + \eta\Delta W(t) + \alpha[W(t-1) - W(t-2)] \tag{30}$$

$$V(t) = V(t-1) + \eta\Delta V(t) + \alpha[V(t-1) - V(t-2)] \tag{31}$$

$$b(t) = b(t-1) + \eta\Delta b(t) + \alpha[b(t-1) - b(t-2)] \tag{32}$$

$$c(t) = c(t-1) + \eta\Delta c(t) + \alpha[c(t-1) - c(t-2)] \tag{33}$$

Where Δ is the incremental operator and such that:

$$\Delta W = \nabla_W E(t),\ \Delta V = \nabla_V E(t) \tag{34}$$

$$\Delta b = \nabla_b E(t),\ \Delta c = \nabla_c E(t) \tag{35}$$

Where η is the online learning rate, here it should be relative small value. Because, on the one hand, the online learning mode emphasizes the learning and tracking the dynamics/ behaviors of the system in real time; On the other hand, the NARX model has captured most of the system dynamics/behaviors when under the offline training mode, so it is not required to update largely or learn much in a short time. And α is the momentum factor, and such that $0<\alpha<1$.

Due to the exponential forgetting mechanism, this online training algorithm has taken both the current operating data of the system and all the operating data in history at every recursive training epoch. And the forgetting factor λ will give a higher weight to more recent operating data. Therefore, the online algorithm emphasizes the learning of the system behaviors in real time, but it also considers the impact of historical factors to some extent.

In addition, we have found that it is appropriate to choose λ and η such that $0.6<\lambda<0.8$, and $0.005<\eta<0.02$ for the water distribution network at Binjiang a practical case studied in section 4.

4. CASE STUDY

In this paper, an NARX model is established and analyzed for a real water distribution network at Binjiang District in Hangzhou city. Binjiang District is now equipped with two water treatment plants--Puyan and Binjiang, which provide daily supply of about 160,000 tons. Although, there are two water treatment plants, the daily regulation of this water distribution network is mainly based on Binjiang plant, and the outlet pressure and outlet flow of pumping stations in plant Puyan is relative constant.

Furthermore, there are 11 pressure monitoring points distributed in the network and the telemetry data is collected by the SCADA system online. Fig. 3 is the layout of the network. And unlike most of the water distribution network, there are no tanks and nodal demand/flow measurement available for model establishment in the network. It is almost a pure pressure flow from water plants to customers.

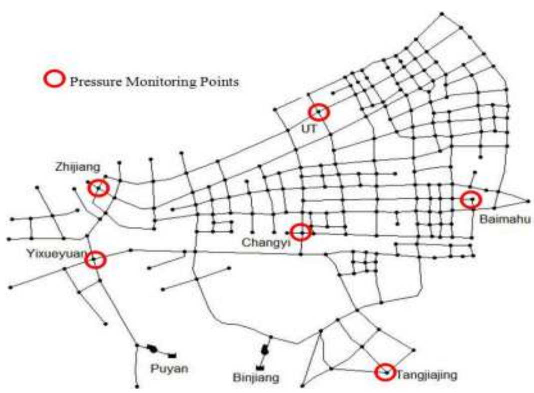

Fig.3. Layout of the water distribution network at Binjiang District

4.1 Model Establishment

From the perspective of the actual daily regulation, this study selects the outlet pressure and outlet flow of pumping stations in plant Binjiang as the control inputs of the NARX model, and 6 representative pressures at monitoring points as the outputs, which are picked by the system operators' experience.

And all the sample data (the pressure of monitoring points, the outlet pressure and outlet flow of pumping stations in water treatment plant) comes from the SCADA system, of which the sampling interval is 15 minutes. And the entire data set covers the period from Oct. 22nd, 2013 to Nov. 3th, 2013 with a total of 1152 observations. Furthermore, the data sets are divided into two periods: the first period covers Oct. 22nd to Oct. 29th with 672 observations, while the second period is from Oct. 30th to Nov. 3th with 480 observations. The first period data, which is assigned to in-sample estimation, is used for model training as the training set. The second period data, which is reserved for out-of-sample evaluation, is used for validation as testing/validation set. And the raw sampling data is preliminary denoised by moving median filter method.

Then, we should ascertain the structure of the three-layered FNN network for the NARX model. There are three parameters related to be determined, those are the length of control and state information defined by integers n_a and n_b, and the number of neurons in the hidden layer denotes by n. In this paper, we will determine these parameters by experience and trial and error method.

First of all, specify $n_a \in \{1,2,3,4\}$ and $n_b \in \{1,2,3,4\}$, which means the time span is limited within an hour (Notice that the sampling period is 15min). Because the longer historical information, the more dimensions of the input vector. It may make the NARX model unnecessary complexity and cause the problem of information redundant. Meanwhile, use the experience formula $n=(p+q)^{1/2}+\gamma$ (γ is an integer constant between 1 to 10) to choose the appropriate number of neurons in the hidden layer. Thus, a structure is ascertained by every given $\{n_a, n_b, n\}$. Then it is the trial and error process.

Specifically, for a given $\{n_a, n_b\}$, firstly obtain the initial value according to the experience formula by setting $\gamma=1$. And then gradually increase the number of neurons by increasing γ. Correspondently, use Levenberg-Marqurd algorithm to train the NARX model with the same training set. At last, select the structure that minimizes the error function (36) and with small n_a and n_b.

$$E(t)=\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{q}e_i^2(j)=\frac{1}{2}\sum_{j=1}^{N}e^T(j)e(j) \tag{36}$$

Where N is the length of the training data set.

For the water distribution system of Binjiang, an NARX model with the optimum structure $\{n_a=2, n_b=3, n=15\}$ is obtained which is the fundamental NARX model. And the structure also means that we will use the control information over the past 30min and state information over the past 45min to estimate the current nodal demand. In summary, the fundamental NARX is consist of 22 inputs, 6 outputs and 15 neurons in the hidden layer, and the structure ascertainment and training process are performed simultaneously.

4.2 Model Validation and Analysis

Now, it is the step of model validation. Firstly, we use the testing data set to validate the fundamental NARX model.

Table 1 shows the quantified model predicting performance (The quantified results are the average of multiple simulation), where RMSE is the root mean-squared error and MAPE is the mean absolute percentage error, i denotes the ith output and N here is the length of testing set. RMSE and MAPE are defined as:

$$RMSE=\sqrt{\sum_{j=1}^{N}\varepsilon_{ij}^2/N}\ ,\ \ MAPE=\sum_{j=1}^{N}|x_{ij}-\hat{x}_{ij}|/(Nx_{ij}) \tag{37}$$

Table 1. Tracking and predicting performance of the offline and online NARX model

Mode	Monitoring Points	Changyi	Baimahu	Tangjiajing	UT	Yixueyuan	Zhijiang
offline	RMSE×100	0.10	0.12	0.18	0.24	0.15	0.16
	MAPE（%）	0.29	0.29	0.45	0.68	0.35	0.42
online	RMSE×100	0.09	0.1	0.14	0.29	0.09	0.13
	MAPE（%）	0.27	0.29	0.40	0.72	0.25	0.39

As shown in Table 1, the fundamental model shows satisfactory predicting performance. The results indicate that the fundamental NARX model has captured the main dynamics of the water distribution network at Bingjiang District.

Secondly, we enable the online training algorithm and switch the fundamental NARX model into online mode, then use the testing data set to validate it.

Table 1 shows the quantified model predicting performance. As it shows, the model shows satisfied predicting performance. And comparing the performance of the different mode in Table 1, we find that the performance of the NARX model on online mode improves slightly as expected compared with the model under offline mode. In term of this case, on the one hand, the NARX under online mode can learn and track the system dynamics, and adapt to the time-variable demand and the real-time dynamics. Therefore, the improvement on performance is reasonable. On the other hand, the fundamental NARX has learned more enough behaviors under offline mode and there is few dynamics new contained in the validation data set. Thus, the performance is marginally improved. In summary, the fundamental NARX model with additional online training mechanism works better than itself.

5. CONCLUSIONS

This article illustrates the application of NARX modeling technique for water distribution network. At the beginning, we propose that using the system control and state information to estimate the variable nodal demand. Thus, an NARX neural network model is derived. Then, we propose a training scheme with a combination of offline training and online training, and corresponding algorithm are proposed. At last, a real case is studied. The results have confirmed the thought of estimating the nodal demand equivalently and shown that the NARX model tracks and predicts dynamics of the water distribution network well. Following study will make further effort to validate the model and combine an optimal control algorithm with it to realize real-time optimal control of the real water distribution network.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61233004. And thanks to Prof. Dr. Lu Renquang for help and suggestions.

REFERENCES

Anh, H. P. H. & Nam, N. T., 2014. Adaptive Line Trajectory Identification of Industrial 5-DOF Robot Arm Using Neural MIMO NARX Model. In: *In AETA 2013: Recent Advances in Electrical Engineering and Related Sciences.* s.l.:Springer Berlin Heidelberg, pp. 605-615.

Basso, M. G. L., Groppi, S. & Zappa, G., 2005. NARX models of an industrial power plant gas turbine. *Control Systems Technology, IEEE Transactions*, pp. 599-604.

Broad, D. R., Maier, H. R. & Dandy, G. C., 2009. Optimal operation of complex water distribution systems using metamodels. *Journal of Water Resources Planning and Management*, pp. 433-443.

Cembrano, G., Wells, G., Quevedo, J. & Pérez, R., 2000. Optimal control of a water distribution network in a supervisory control system. *Control engineering practice*, pp. 1177-1188..

Coelho, B. & Andrade-Campos, A., 2014. Efficiency achievement in water supply systems—A review. *Renewable and Sustainable Energy Reviews*, pp. 59-84.

DeMoyer, R. & Horwitz, L., 1975. Automatic Control of Water Distribution. *Journal of Dynamic Systems, Measurement, and Control*, pp. 126-130.

Hagan, M. T., Demuth, H. B. & Beale, M. H., 1996. *Neural network design.* s.l.:Boston: Pws Pub..

Hutton, C. J., Kapelan, Z., Vamvakeridou-Lyroudia, L. & Savic, D. A., 2012. Real-time demand estimation in water distrubtion systems under uncertainty. In: *In WDSA 2012: 14th Water Distribution Systems Analysis Conference.* s.l.:Engineers Australia, p. 1374.

Lee, Y. W. & Chang, T. L., 2009. Application of NARX neural networks in thermal dynamics identification of a pulsating heat pipe. *Energy Conversion and Management*, pp. 1069-1078.

Paluszczyszyn, D., Skworcow, P. & Ulanicki, B., 2013. Online simplification of water distribution network models for optimal scheduling. *Journal of Hydroinformatics*, pp. 652-665.

Rao, Z. & Alvarruiz, F., 2007. Use of an artificial neural network to capture the domain knowledge of a conventional hydraulic simulation model. *Journal of Hydroinformatics*, pp. 15-24.

Razavi, S., Tolson, B. A. & Burn, D. H., 2012. Review of surrogate modeling in water resources. *Water Resources Research*, pp. 50-62.

Shamir, U. & Salomons, E., 2008. Optimal real-time operation of urban water distribution systems using reduced models. *Journal of Water Resources Planning and Management*, pp. 181-185.

Shen, H., McBean, E. & Wang, Y., 2014. Sensor placement under nodal demand uncertainty for water distribution systems. In: *In Securing Water and Wastewater Systems.* s.l.:Springer International Publishing, pp. 123-133.