KSCE Journal of Civil Engineering (2018) 22(9):3233-3245

DOI 10.1007/s12205-018-1195-z

Environmental Engineering

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Comparison of ARIMA and NNAR Models for Forecasting Water Treatment Plant's Influent Characteristics

Afshin Maleki*, Simin Nasseri, Mehri Solaimany Aminabad***, and Mahdi Hadi******

Received August 20, 2017/Accepted November 15, 2017/Published Online April 10, 2018

Abstract

A reliable forecasting model for each Water Treatment Plant (WTP) influent characteristics is useful for controlling the plant's operation. In this paper Auto-Regressive Integrated Moving Average (ARIMA) and Neural Network Auto-Regressive (NNAR) modeling techniques were applied on a WTP's influent water characteristics time series to make some models for short-term period (to seven days ahead) forecasting. The ARIMA and NNAR models both provided acceptable generalization capability with R^2s ranged from 0.44 to 0.91 and 0.45 to 0.92, respectively, for chloride and temperature. Although a more prediction performance was observed for NNAR in comparison with ARIMA for all studied series, the forecasting performance of models was further examined using Time Series Cross-Validation (TSCV) and Diebold-Mariano test. The results showed ARIMA is more accurate than NNAR for forecasting the horizon-daily values for CO_2, Cl and Ca time-series. Therefore, despite of the good predictive performance of NNAR, ARIMA may still stands as better alternative for forecasting task of aforementioned series. Thus, as a general rule, not only the predictive performance using R^2 statistic but also the forecasting performance of a model using TSCV, are need to be examined and compared for selecting an appropriate forecasting model for WTP's influent characteristics.

Keywords: *time series analysis, neural network auto-regressive model, auto-regressive integrated moving average model, water treatment plant, forecasting*

1. Introduction

Water characteristics such as temperature, turbidity and pH are important water quality parameters and there is a significant relationship between these parameters and amounts of coagulants and flocculants used in water treatment processes in a Water Treatment Plant (WTP) (Wu and Lo, 2008; Wu and Lo, 2010). Prediction of the influent water characteristics may be helpful in the optimal scheduling of coagulation and flocculation process. To maintain stable performance, it is desirable to know in advance the influent water characteristics of a WTP. Operators' experience and online sensors in practice can provide an estimation of the influent water characteristics. But such estimations or measurements however, may not be enough to manage a WTP, especially for operators that want to manage the WTP performance for one or some days later. The precipitation may cause a large variability in the influent water characteristics like turbidity (Hadi *et al*., 2016) and reduce the efficiency of the WTP. Making some forecasting models for water quality characteristics based on their historical data may be useful in which by time series analysis can be done (Maest *et al.*, 2005). The forecasting of the future states of a time series is possible by analyzing the series pattern using its available observations.

For more than half a century, Box–Jenkins (Box *et al*., 2013) or Auto-Regressive Integrated Moving Average (ARIMA) time series models have dominated many areas of time series forecasting. In an ARIMA model, the future values of a variable are assumed to be a linear function of several past observations and random errors. The ARIMA assumes that the series is generated from a linear process (Khashei and Bijari, 2010). Thus, it may be inappropriate if the underlying mechanism is nonlinear. Indeed, real world systems are often nonlinear (Zhang *et al.*, 1998). In order to consider the non-linear behavior within the variables, more sophisticated tools and methods may be needed. The Artificial Neural Network (ANN) is a technique allows modelling of complex nonlinear relationships among

*Professor, Dept. of Environmental Health Engineering, Environmental Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran (E-mail: maleki43@yahoo.com)
**Professor, Center for Water Quality Research (CWQR), Institute for Environmental Research (IER), Tehran University of Medical Sciences, Tehran, Iran; Dept. of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran (E-mail: naserise@tums.ac.ir)
***M.Sc. Student, Dept. of Environmental Health Engineering, Environmental Health Research Center, Kurdistan University of Medical Sciences, Sanandaj, Iran (E-mail: mehri.solaimany@yahoo.com)
****Assistant Professor, Center for Water Quality Research (CWQR), Institute for Environmental Research (IER), Tehran University of Medical Sciences, Tehran, Iran (Corresponding Author, E-mail: hadi_rfm@yahoo.com; m-hadi@tums.ac.ir)

input regressors and response variable(s) (Thoplan, 2014). One of the most dominant advantages of ANN models over other nonlinear statistical models is that ANN is universal approximator that is able to approximate a large class of functions with a high degree of accuracy (Zhang and Qi, 2005). Feed-forward ANN with backpropagation algorithm have been widely used in different environmental applications such as wastewater treatment plant performance prediction (Khodadadi *et al.*, 2016), estimation of pollutants removal efficiency (Jafari Mansoorian *et al.*, 2017) and prediction of air quality indices (Rostami Fasih *et al.,* 2015). Network Network Auto-Regressive (NNAR) model is one type of ANNs in which the lagged values of time series are used as input predictors to model and the output is predicted values of the series. One of the main differences between NNAR and ARIMA models is the NNAR does not impose any restriction on its parameters to ensure stationarity (Thoplan, 2014).

In this study, we compare Box-Jenkins ARIMA and NNAR models to predict daily influent water characteristic of Sanandaj's WTP. This paper presents an approach to forecast the treatment plant influent water characteristic for a short-term period (to seven days ahead). In this study two prediction models were developed for each influent water characteristics including alkalinity (Alk), pH, calcium (Ca), carbon dioxide (CO_2), temperature (T), total hardness (TH), turbidity (Tur), Total Dissolved Solids (TDS), Electrical Conductivity (EC) and chloride (Cl). Forecasting performance of the selected best ARIMA and NNAR models to forecast series future daily values was compared using several statistics and cross-validation analysis.

2. Data Collection and Model Building

2.1 Data Collection and Preparation

The pre-processed influent data of Sanandaj's - a western city in Iran - water treatment plant as examined in our previous study (Solaimany-Aminabad *et al*., 2013) was again re-analyzed in this study with different approaches and different aims. A summarized description of the variables, units of measure, range of the data, together with the mean and standard deviation of the plant processed data are summarized in Table 1. All statistical modeling and analyses were performed using R software version 3.3.1 (R Core Team, 2016).

Several statistics including the correlation coefficient (r), Root of Mean Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MAE) were calculated to examine the performance of models. The underlying expressions of these statistics are given as follows:

$$r = \frac{\sum_{i=1}^{n}(Y_O - \overline{Y}_O)(Y_P - \overline{Y}_P)}{\sqrt{\sum_{i=1}^{n}(Y_O - \overline{Y}_O)^2 (Y_P - \overline{Y}_P)^2}} \tag{1}$$

$$MSE = \frac{\sum_{i=1}^{n}(Y_O - \overline{Y}_P)^2}{n} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_O - Y_P)^2} \tag{3}$$

$$MAE = \frac{\sum_{i=1}^{n}|Y_P - Y_O|}{n} \tag{4}$$

where n = Number of observations
Y_O = Observed y_t
$\overline{Y}_O$ = Mean of Y_O
Y_P = Predicted y_t
$\overline{Y}_P$ = Mean of Y_P

2.2 Auto-Regressive Integrated Moving Average (ARIMA) model

The Box-Jenkins method (Box *et al*., 2013) is a classical statistical modeling technique for time series forecasting. In this study an ARIMA (p, d, q) model was used to forecast the future value of each influent water characteristics The underlying process that generates the time series, y_t with the mean μ, can be described by following equation:

$$\phi(B)\nabla^d(y_t - \mu) = \theta(B)a_t \tag{5}$$

where, y_t and a_t are the actual value and random error at time period t, respectively; $\nabla = (1-B)$, B is the backward shift operator, $\phi(B)$ and $\theta(B)$ are the regular autoregressive and moving average factors, respectively:

$$\phi(B) = 1 - \sum_{i=1}^{p}\phi_i B^i \tag{6}$$

Table 1. Pre-processed Influent Water Characteristics Data

Variable	Description(unit)	Min.	Median	Mean	Max.	SD	λ
CO_2	Carbon dioxide (mg/l)	0.1	2.3	2.6	8.2	1.5	0.41
TH	Total Hardness (mg/l)	122	155.3	154.0	197.2	10.1	1.86
Cl	Chloride (mg/l)	5.5	9.0	8.9	12.5	1.2	-0.2
Ca	Calcium (mg/l)	32.4	48.0	47.2	59.8	4.0	2.43
TDS	Total dissolved solids (mg/l)	157	214.0	211.0	252.0	12.6	5.05
Alk	Total alkalinity (mg/li)	120.6	160.2	158.2	193.0	14.0	1.76
EC	Electrical conductivity (μmohs/cm)	260	333.5	330.0	393.0	18.5	4.59
pH	-	7.16	8.2	8.2	8.9	0.3	-0.02
Tur	Turbidity (NTU)	0.5	2.0	3.3	24.0	3.5	-0.20
T	Temperature (C°)	2	6.0	11.0	11.4	5.3	0.72

λ: Box-Cox transformation parameter

$$\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j \quad (7)$$

where ϕ_i(i = 1, 2, ..., p) and q_j(i = 1, 2, ..., q) are model parameters, p and q are integers and often referred to as orders of the model. The d in Eq. (5) is an integer and referred to the order of differencing. Random errors, a_t, are often assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

The Box and Jenkins methodology (Box *et al.*, 2013) includes three model building steps and a forecasting step as follows:

1. Identification: The Box–Jenkins model assumes that the time series is stationary. A stationary time series is characterized by its statistical characteristics such as the mean which is being constant over time. In this study the transformation and differencing techniques were used for non-stationary series. The autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) charts for the stationary series were examined to identify the orders of p and q.
2. Estimation: To identify the optimal order of p and q in ARIMA models, the residual sum of squares error (SSE) were compared by F-test among the probational models from the lowest to highest order of p and q. The best model was selected when the residual SSE of the model was not statistically significant at 0.05 level of probability. The least-squares method was used to estimate the parameters of the model.
3. Diagnostic checking: Several diagnostic techniques can be used to examine the goodness of fit of a tentative model. If the model is not adequate, a new model should be examined, which will again be followed by the steps of parameter estimation and model verification. In this study several diagnostic statistics and plots were used to examine the models. The Ljung–Box (LB) test (Ljung and Box, 1978) was used to investigate white noise (zero mean, constant variance, uncorrelated process and normally distributed) in residuals. The significance of model parameters was examined using student t-test. The correlation between parameters of each model was assessed as a measure of models adequacy. The normal pattern of the models' residuals was checked by histogram and normal distribution plot. Moreover, the lack of correlation between the residuals, were assessed using ACF and PACF plots of residuals. The presence of serial correlation between the residuals was also evaluated using the Durbin-Watson (DW) (Durbin and Watson, 1950) statistic. The above three-step model building process was repeated several times until a satisfactory model selected finally. The selected model was used for forecasting purposes.
4. Forecasting: the model with the least Mean Squared Error (MSE) value was chosen for forecasting. The forecasting performance of the best model was then compared with the performance of its equivalent NNAR model using cross-validation technique and MAE statistic.

2.3 Neural Network Auto-Regressive (NNAR) model

Feed forward network is the most widely used neural network model for time series modeling and forecasting (Zhang *et al.*, 1998). The NNAR model is a three layers feedforward neural network which involves a linear combination function and an activation function. The relationship between the model output (y_t) and the inputs (y_{t-1}; ...; y_{t-p}) has the following mathematical representation:

$$y_t = w_0 + \sum_{j=1}^{h} w_j \cdot g\left(w_{0,j} + \sum_{i=1}^{n} w_{i,j} \cdot y_{t-i}\right) + \varepsilon_t \quad (8)$$

where, w_{ij} (i = 0, 1, 2,…, n, j = 1, 2, …, h) and w_j (j = 0, 1, 2, …, h) are model parameters or connection weights; n is number of input nodes; and h is number of hidden nodes. A sigmoid function was used as the hidden layer transfer function that is shown in Eq. (9). The most widely used activation function for output layer is a linear function.

$$Sig(x) = \frac{1}{1 + \exp(-x)} \quad (9)$$

In our previous paper (Solaimany-Aminabad *et al.*, 2013) we only examined NNAR models and the aim was only to find the "prediction performance" of developed NNAR models using training, validation and test data sets. But in this paper all NNAR and ARIMA models were made in R software environment and our aim was to compare "forecasting performance" of these two modeling techniques. Moreover in this study cross-validation technique was used to examine the performance instead of splitting data set. In this paper, NNAR (p, P) networks were considered with one hidden layer. The notation of NNAR (p, P) indicates p non-seasonal and P seasonal lagged inputs into to the models. A common procedure to find the optimal structure for an NNAR model is to test numerous networks with varying numbers of input and hidden units (p, P, h) and then estimate generalization error for each to receive a structure with lowest generalization error (Hosseini *et al.,* 2006). This procedure was examined in our previous work (Solaimany-Aminabad *et al*., 2013).

One of the important issues affecting the performance of an ANN model is the size of the hidden layer (Hagan *et al*., 1996). In practice, a simple ANN structure with a small number of neurons in the hidden layer is goodly capable to predict out-of-sample data (Pankratz, 2009). Thus, a minimum number for neurons in the hidden layer may provide the highest performance for the networks were selected. This is also confirmed in our previous work where the performance of the NNAR network decreased almost by increasing the hidden layer size. Our findings in previous work (Solaimany-Aminabad *et al*., 2013) didn't show a significant difference in the performance of a model by changing the size for neurons in hidden layer up to 10. Thus for simplicity we estimated the h based on the approach recommended by Hyndman and Khandakar (2007): (p + P + 1)/2 (rounding to the nearest integer).

Another important task in NNAR modeling is to determine the appropriate values for p and P lagged inputs. In this study the

data series were first examined for the need to normalization or other transformations and then the number of non-seasonal lags was determined by plotting PACF. The lags in PACF chart with significant correlation coefficient and lowest generalization error were considered as the appropriate number for p. No seasonal lags were import to the models (P = 0) because of the low frequency in seasonal pattern of data.

2.4 Cross-validation Analysis

To compare the time series forecasting performance, a Time Series Cross-Validation (TSCV) analysis was performed. In TSCV analysis a comparison was performed between the forecasted values by ARIMA and NNAR models using the Mean Absolute Error (MAE). The TSCV analysis was performed according to following steps:

1. Out of 700 observations (seven last observations of each series were reserved for checking the forecasting performance of the models) in each series the model was fitted to 50% of the data (350 observations) from Y_{O_1}, …, Y_{O_t} and then the next seven observation from $Y_{O_{t+1}}$ to $Y_{O_{t+7}}$ were forecasted ($Y_{P_{t+1}}$ to $Y_{P_{t+7}}$).
2. One further observation was added to the training dataset (Y_{O_1}, …, $Y_{O_{t+1}}$; 351 observations), then, the observed test set shifted one step ahead ($Y_{O_{t+2}}$ to $Y_{O_{t+8}}$) and they forecasted again ($Y_{P_{t+2}}$ to $Y_{P_{t+8}}$). The observed test set and corresponding forecasted values were stored in $[Y_O]$ and $[Y_P]$ matrices with a dimension of 357 × 7, respectively.
3. Step two was repeated for 350 times and $[Y_P]$ and $[Y_O]$ matrices were completely filled for each model.
4. The Absolute Errors (AE) ($|Y_P - Y_O|$) and errors (E) ($Y_O - Y_P$) were then calculated for the forecasted values and stored in [AE] and [E] matrixes, respectively.
5. The Diebold-Mariano test (Diebold and Mariano, 1995) was then applied on the AE matrices of models to compares the forecast accuracy of ARIMA and NNAR as follows: In this test the forecast absolute errors for models were defined by $AE_{it} = |Y_P - Y_O|$. The loss associated with forecast i (i = 1, 2) is assumed to be a function of the forecast absolute error, AE_{it}, and is denoted by $g(AE_{it})$. The time-t loss differential between the two forecasts 1 and 2 defined by $d_t = g(AE_{1t}) - g(AE_{2t})$.The two forecasts have equal accuracy if and only if the loss differential has zero expectation for all t. Thus, the null hypothesis (H0: E(dt) = $0 \forall t$; two forecasts have the same accuracy) was test versus the alternative hypothesis (H1: E(dt) ≠ 0; two forecasts have different levels of accuracy).

   The Diebold-Mariano test statistic (DM) is:

$$DM = \frac{\bar{d}}{(LRV_{\bar{d}}/T)^{1/2}} \tag{10}$$

   where $\bar{d}$ is the sample mean of the loss deferential series (d_t; t = 1, …, T) and $LRV_{\bar{d}}$ is consistent estimate of the asymptotic (long-run) variance of $\sqrt{T}\bar{d}$. Diebold and Mariano (1995) show that under the null of equal predictive accuracy, the test statistic DM is asymptotically N (0;1) distributed. Thus the null hypothesis of no difference will be rejected at 5% if $|DM| > 1.96$.
5. A graphical comparison was also performed by plotting the actual and forecasted values for seven last observations of each series.

3. Results and Discussion

3.1 Checking the Series Stationarity

In this study, Box and Cox (Box and Cox, 1964) power transformation was used to normalize and stationarize the series variances. In a study conducted by Salas (1980) the Box-Cox transformation was recommended to improve the results predicted by time-series models.

To differentiate a stationary from non-stationary time series, the unit root test can be used (Rumelhart and McClelland, 1986). In this study the Augmented Dicky-Fuller (ADF) test was used to examine the stationarity of the time series processes. Using a nonstationary time series in a regression model may lead to misleading conclusions (Plazl *et al*., 1999). As shown Table 2, all variable except chloride series are non-stationary series. Differencing can help to stabilize the mean of a time series by removing the changes in the level of series, and so eliminating the trend. Moreover, the first order difference of each non-stationary series appears to be a stationary series (p-value < 0.01). Thus for all series the order of differencing (d) or the integrated term of the model (I) in ARIMA model, was considered to be with one unit root.

3.2 Models Structures and Learning

Autocorrelation function is statistically a useful tool to get a description of a time series (Zhang *et al.*, 1998). Analysis of ACF and PACF could be performed to estimate the number of input auto-correlated vectors roughly to create an appropriate model (Kusiak *et al.*, 2012).

The AR or MA terms in ARIMA model were determined by plotting the ACF and PACF plots of the differenced series (Fig. 1).

Table 2. Analysis the Stationarity in Mean of Series using Unit Root Test

Series	First order differenced series		Non-differenced series	
	p-value	ADF statistic	p-value	ADF statistic
CO_2	< 0.01	-12.052	0.06226	-3.3489
TH	< 0.01	-11.727	0.5681	-2.0241
Cl	-	-	0.04375	-3.485
Ca	< 0.01	-12.9432	0.2598	-2.7525
TDS	< 0.01	-13.6288	0.05626	-3.8547
Alk	< 0.01	-9.792	0.4699	-2.2561
EC	< 0.01	-13.7258	0.05998	-3.3621
pH	< 0.01	-11.5507	0.1493	-3.0135
Tur	< 0.01	-12.7716	0.1471	-3.0187
T	< 0.01	-11.3249	0.715	-1.6771

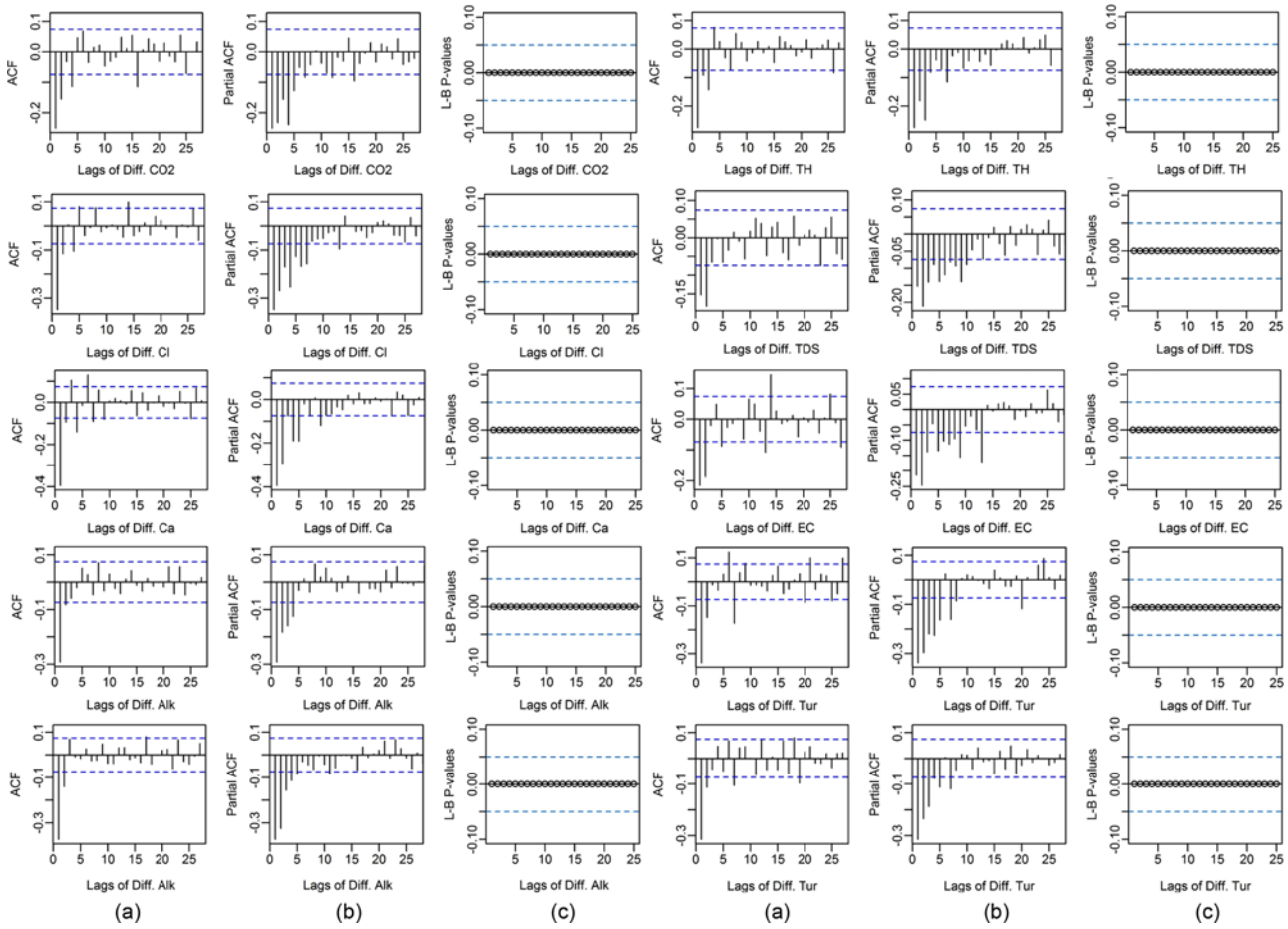

Fig. 1. (a) ACF Plots, (b) PACF Plots, (c) L-B p-values for Several Lags of First-order Differenced Series

ACF is a bar chart of the coefficients of correlation between a time series and its lags, while the PACF is a plot of the partial correlation coefficients between the series and its lags. As a rule the number of lags beyond which the PACF cuts off of the differenced series can be considered as AR terms in ARIMA or p term in NNAR models. In a same way, the lags beyond which the ACF cut off of the differenced series can suggest the number of MA terms. For example according to Fig. 1, in the case of CO_2, Cl, TDS, EC, Tur and T variables the PACF plot suggest atleast one AR term to their models. But selection of the best number of terms in each model should be based on some criteria will be discussed in section 3.4. To test if the residuals of the proposed models are serially uncorrelated, The LB test (Ljung and Box, 1978) could be used to examine the lack of fit of a time series model. In the case of Fig. 1(c) the LB test p-value for all the differenced series lags is well less than 0.05, meaning all series are dependent on their lags.

The feedforward network with one hidden layer is widely used for the prediction aims in most studies on time series modeling (Howard and Mark, 2000). It also has been proven that the neural networks with one hidden layer when sufficient weight is provided in the model, are able to provide a good approximation for any function (Battiti, 1992).

In this study the Multi-Layer Perceptron (MLP) algorithm with one hidden layer was used in all NNAR models. Several different ANN models were examined by Hill *et al.* (1994) to predict the Biochemical Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) parameters in the effluent of a petrochemical wastewater treatment system. They examined several regressor inputs (including Total Suspended Solids (TSS), Total Dissolved Solids (TDS), phenol, ammonia nitrogen, organic carbon and Total

Table 3. Topology and Goodness of Fit Measures for NNAR Models

Variable	Best model	Topology	RMSE	MSE	r	R^2
CO_2	NNAR(6)	6-4-1	0.74	0.55	0.86	0.74
TH	NNAR(4)	4-2-1	4.24	18.00	0.91	0.82
Cl	NNAR(8)	8-4-1	0.81	0.65	0.67	0.45
Ca	NNAR(6)	6-4-1	1.56	2.43	0.89	0.79
TDS	NNAR(7)	7-4-1	5.38	28.91	0.88	0.77
Alk	NNAR(5)	5-3-1	5.27	27.81	0.93	0.86
EC	NNAR(11)	11-6-1	7.16	51.24	0.92	0.84
pH	NNAR(5)	5-3-1	0.14	0.02	0.85	0.73
Tur	NNAR(7)	7-4-1	1.64	2.70	0.86	0.73
T	NNAR(6)	6-4-1	1.54	2.36	0.96	0.92

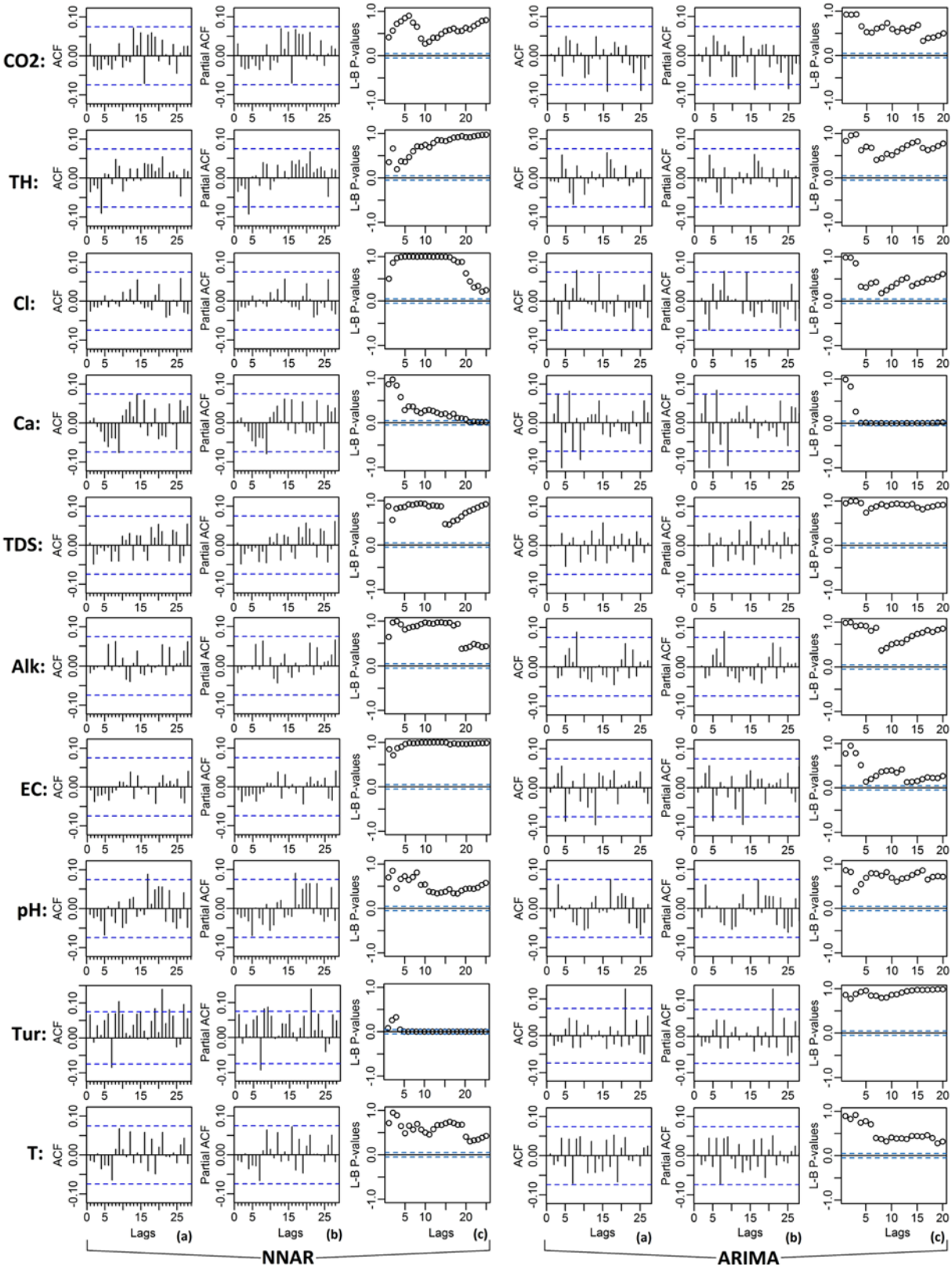

Fig. 2. (a) ACF, (b) PACF, (c) L-B Test p-values Plots of the Models Residuals

Kjeldahl Nitrogen (TKN)) using MLP algorithm to predict the COD and BOD as the outputs of the model.

In this study, the learning algorithm in all MLP networks was Levenberg–Marquardt (LM) algorithm technique. The LM technique was used because of its high speed, high precision and accuracy for the parameters estimation in comparison with other

techniques (Matalas, 1967).

3.3 NNAR Model Characterization

According to the Table 3 the NNAR models for all of the studied variables have been adequately learned to predict the variable within series. Among all, Total Hardness (TH) model gets the lowest input components (4 lags) while the most number of input lags was for EC model (11 lags). As shown in Table 3 and illustrated in Fig. 3(a) the correlation coefficient (r) between the predicted and observed values for all variables except the chloride, found to be higher than 0.8.

The r and coefficient of determination (R^2) obtained for chloride by NNAR model were 0.67 and 0.45, respectively. R^2 implies that only 45% of the series variance is described by eight input regressors and the remaining 55% of the variance could not be attributed to these inputs. Chloride may get into surface water from several manmade sources including wastewater from industries and municipalities, leachate from municipal landfills wastewater from water softening, road salting, animal waste and agricultural runoff (Panno *et al.*, 2002). Thus this variable may be strongly influenced by current or past values of the other driving exogenous series and to better predict the variable, autoregressive exogenous (ARX) models may be more effective if sufficient information be available on influencing exogenous series.

Figure 2(c) confirms the independence of residuals for all NNAR models based on LB test. The LB statistic (Q) is a function of the accumulated autocorrelations up to any specified time lag. A significant Q for residuals indicates a possible problem with the model. As can be seen, the LB test p-value for all the lags contributing to the models is well above 0.05, meaning the residuals are independently distributed (from each other).

Moreover a model is ideal when the ACFs for the residuals are zero. This means that Q should be statistically equal to zero for any m lag. Fig. 2(a) and (b) show the ACF and PACF plots for the residuals of the models fitted to studied variables, respectively. The plots reveal that there isn't any autocorrelation within the residuals and no additional information are exist which not accounted for in the models. The predicted values versus standardized residuals for NNAR models are presented in Fig. 3(b). As presented, the homogeneous variance assumption is not violated in the case of all models and the variance of residuals appears to be independent of the size of fitted values. Accordingly the residuals fall randomly within the horizontal band (between -2 and +2). This is a pattern suggests that the variances of the error terms are equal. Moreover, no one residual stands out from the random pattern thus suggests that there are no outliers.

The DW test (Durbin and Watson, 1950) is used to test the hypothesis that there is no lag one autocorrelation in the residuals of the models. If there is no autocorrelation, the DW test statistic (d) is symmetric around 2. Thus d = 2 indicates no autocorrelation. Field (2009) suggests that d statistics less than unity (positive autocorrelation) or greater than 3 (negative autocorrelation) are a cause for alarm as the model is inadequate. In this study, the d

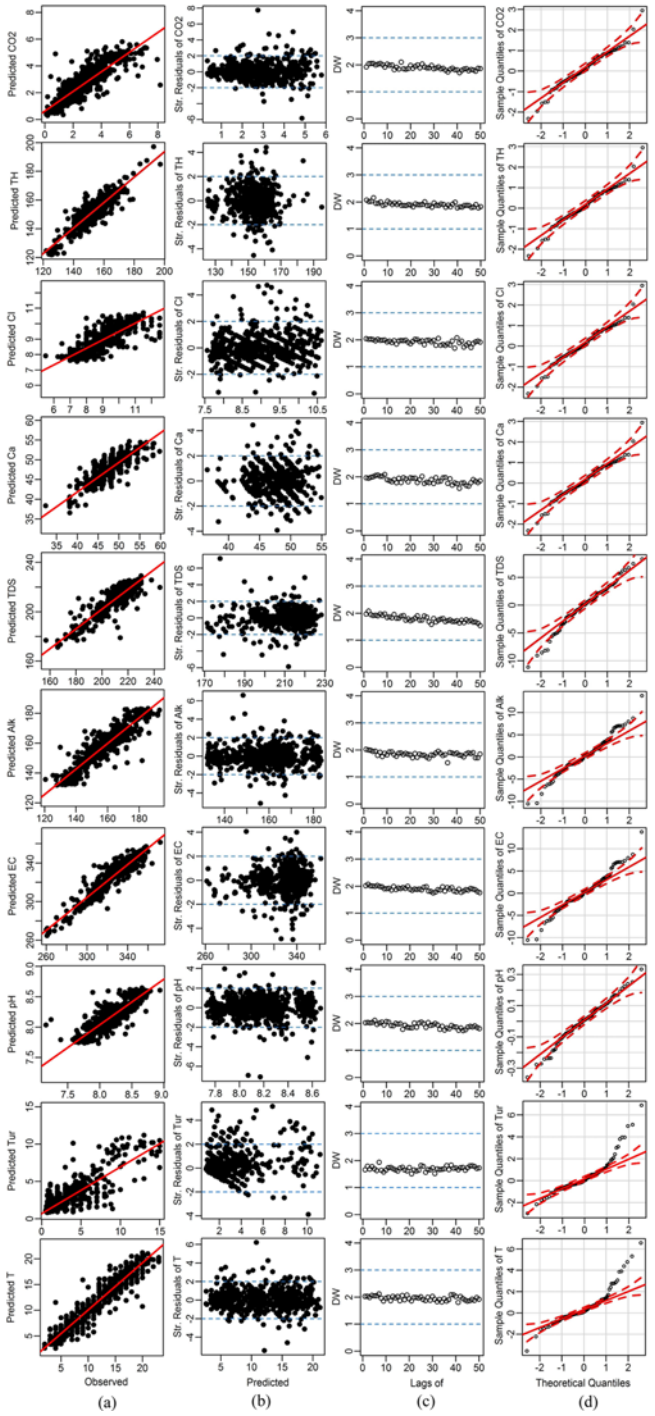

Fig. 3. Diagnostic Plots for NNAR Models: (a) Predicted Versus Observed Plots, (b) Predicted Versus Standardized Residual, (c) DW Test Statistic Plots, (d) Normal Q-Q Plots

statistic for the residuals of NNAR models was calculated at different time lags. As shown in Fig. 3(c) amount of d for all models are close to 2 and in no cases it found larger than 3 or less than unity. So in all proposed NNAR models, serial correlation between the residuals could not be existed.

The Q-Q plots for the residuals of NNAR models are shown in Fig. 3(d). As can be seen, the residuals acceptably follow a normal

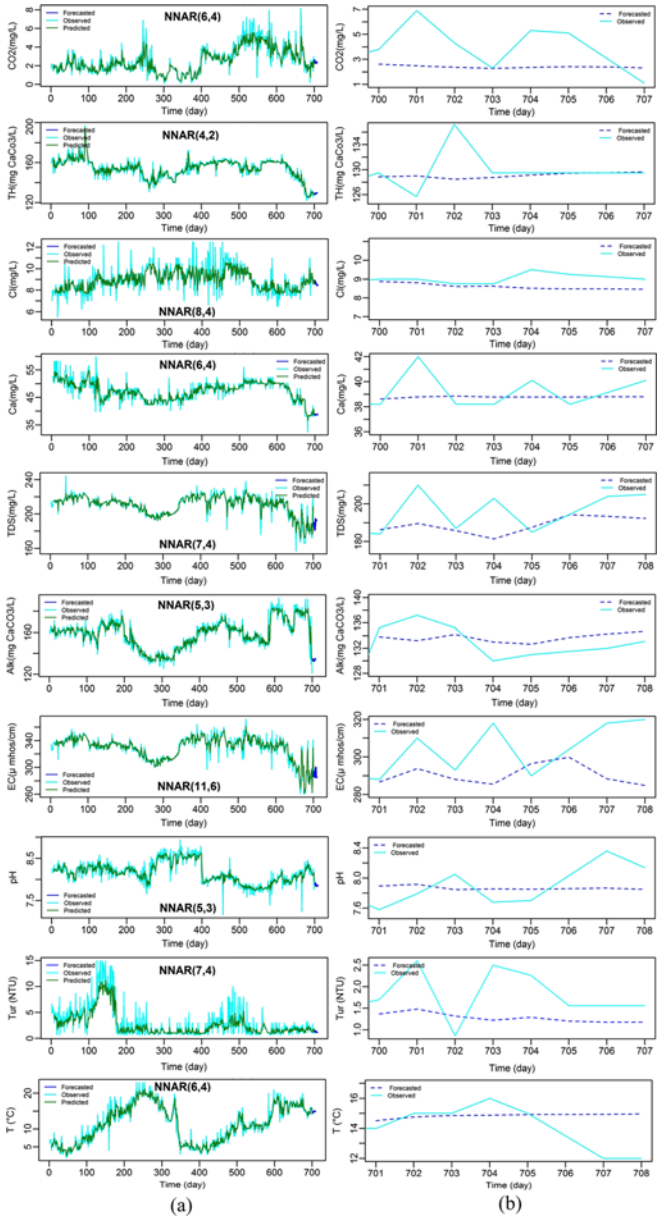

Fig. 4. (a) Predicted Values Response Plot for All Observations, (b) Seven Horizon Daily Plots for Forecasted Values by NNAR Models

distribution. In Fig. 4 predicted values plots for all observations and the plots for forecasted values for seven horizons next days by NNAR models are presented. As shown in Fig. 4(a) the bold lines (green color) represent the values predicted by each model and light lines (blue color) represent observed values. In overall, the proposed NNAR models acceptably can be used for forecasting of the raw water characteristics entering water treatment plant. However, the comparison of forecasting performance of NNAR and their equivalent ARIMA models using TSCV analysis will be discussed in a separate section.

3.4 ARIMA Models Parameters Estimation and Characterization

As presented in Table 2 the non-stationarity in all variables except Cl was confirmed by Dickey-Fuller test. Therefore, only the series those exhibit a general linear trend, were differenced by first-order differencing (d = 1) to achieve stationarity (Fig. 1).

As shown in Fig. 1 for CO_2, Cl, TDS, EC, Tur and T variables the partial autocorrelation PACF of the differenced series display a gradual cutoff which suggest adding one AR(p) term to their models. In the case of alkalinity a sharper cutoff can be seen in PACF plot, thus a greater AR terms may be need in the model. The ACF of the differenced series also displays a sharp cutoff thus it was decided to add at least one MA(q) term to all models. The lags beyond which the ACF and PACF cut off, are the numbers indicate MA and AR terms, respectively. However, in this study we examined different models with different number of AR and MA terms to find the best descriptive one. The judgment in choosing the best was based on providing following conditions in the model.

- Parsimonious and simple model with a minimum number of coefficients,
- Lowest Akaike Information Criteria (AIC) (Akaike, 1974),
- Statistically significant parameters,
- Low correlation between input parameters (not greater than 0.9) (Pankratz, 2009),
- Time-independency of residuals,
- Residual with normal distribution.

A parsimonious model is a model that provides a desired level of prediction with as few predictor variables as possible (Salas *et al.,* 1982). In a time series model with N observation data and K parameters, the ratio N/K is called parsimony index. Salas *et al.* (1982) suggest at least a value of 15 for this index. Based on the number of input regressors in suggested ARIMA models, the parsimony index for all models was found higher than threshold value. The AIC is another selection measure to compare different models. Lowest AICs for the best selected models are summarized in Table 4.

The estimated correlation coefficients for ARIMA models are presented in Table 4 (See Fig. 5(a)). Similar to NAR models, for all ARIMA models except chloride, the value of R^2 is greater than 0.8.

In this study the ARIMA models parameters were determined by Maximum Likelihood (ML) method. In a study by Salas and

Table 4. Goodness-of-fit Measures for Proposed ARIMA Models

Series	Model	RMSE	MSE	r	R^2	AIC	N/K
CO_2	ARIMA(1,1,1)	0.79	0.63	0.84	0.72	869.5263	350
TH	ARIMA(0,1,3)	4.27	18.23	0.91	0.82	13105.56	233
Cl	ARIMA(1,0,2)	0.82	0.67	0.67	0.44	-1985.22	233
Ca	ARIMA(0,1,2)	1.61	2.60	0.89	0.79	6778.019	350
TDS	ARIMA(1,1,1)	5.59	32.40	0.87	0.76	12005.38	350
Alk	ARIMA(2,1,4)	5.44	29.60	0.92	0.85	10658.77	116
EC	ARIMA(1,1,1)	8.71	75.95	0.87	0.76	65407.11	350
pH	ARIMA(0,1,2)	0.14	0.02	0.85	0.73	-3955.28	350
Tur	ARIMA(1,1,1)	1.94	3.77	0.80	0.65	1016.356	350
T	ARIMA(1,1,1)	1.61	2.60	0.95	0.91	1998.03	350

Table 5. Characteristics of Proposed ARIMA Models for Studied Time Series

Series	Parameters	Value	p-value	Correlation coefficient						
				ar_1	ar_2	ma_1	ma_2	ma_3	ma_4	δ
Alk ARIMA(2,1,4)	ar_1	0.22	0.00	1.00	0.36	-0.64	-0.03	-0.04	0.38	-
	ar_2	-0.96	0.00	0.36	1.00	-0.21	-0.61	0.08	0.29	-
	ma_1	-0.63	0.00	-0.64	-0.21	1.00	-0.37	0.75	-0.58	-
	ma_2	0.93	0.00	-0.03	-0.61	-0.37	1.00	-0.62	0.53	-
	ma_3	-0.40	0.00	-0.04	0.08	0.75	-0.62	1.00	-0.56	-
	ma_4	-0.12	0.00	0.38	0.29	-0.58	0.53	-0.56	1.00	-
Cl ARIMA(1,0,2)	ar_1	0.98	0.00	1.00	-	-0.26	-0.22	-	-	-0.10
	ma_1	-0.61	0.00	-0.26	-	1.00	-0.69	-	-	0.03
	ma_2	-0.22	0.00	-0.22	-	-0.69	1.00	-	-	0.02
	δ	1.76	0.00	-0.10	-	0.03	0.02	-	-	1.00
TH ARIMA(0,1,3)	ma_1	-0.38	0.00	-	-	1.00	-0.36	-0.29	-	-
	ma_2	-0.15	0.00	-	-	-0.36	1.00	-0.36	-	-
	ma_3	-0.15	0.00	-	-	-0.29	-0.36	1.00	-	-
Ca ARIMA(0,1,2)	ma_1	-0.53	0.00	-	-	1.00	-0.72	-	-	-
	ma_2	-0.26	0.00	-	-	-0.72	1.00	-	-	-
pH ARIMA(0,1,2)	ma_1	-0.57	0.00	-	-	1.00	-0.67	-	-	-
	ma_2	-0.15	0.00	-	-	-0.67	1.00	-	-	-
CO_2 ARIMA(1,1,1)	ar_1	0.44	0.00	1.00	-	-0.80	-	-	-	-
	ma_1	-0.85	0.00	-0.80	-	1.00	-	-	-	-
TDS ARIMA(1,1,1)	ar_1	0.52	0.00	1.00	-	-0.68	-	-	-	-
	ma_1	-0.89	0.00	-0.68	-	1.00	-	-	-	-
EC ARIMA(1,1,1)	ar_1	0.43	0.00	1.00	-	-0.71	-	-	-	-
	ma_1	-0.87	0.00	-0.71	-	1.00	-	-	-	-
Tur ARIMA(1,1,1)	ar_1	0.25	0.00	1.00	-	-0.59	-	-	-	-
	ma_1	-0.86	0.00	-0.59	-	1.00	-	-	-	-
T ARIMA(1,1,1)	ar_1	0.27	0.00	1.00	-	-0.79	-	-	-	-
	ma_1	-0.75	0.00	-0.79	-	1.00	-	-	-	-

Obeysekera (1982) the ARIMA model parameters were also estimated by ML method. They suggested ML method as an alternative and best approach could be used to estimate the model's parameters in future. The name and the estimated values for different parameters in ARIMA models are given in Tables 5.

The ARIMA (2,1,4) found as the best descriptive model for alkalinity with four MA terms and two AR terms. This is a complex model in comparison with models obtained for other water characteristic time series. The proposed models for total hardness (TH), calcium (Ca) and pH was only characterized by MA terms. For example in ARIMA (0,1,3) for TH, there are three significant (p-value < 0.05) moving average terms (ma_1, ma_2 and ma_3) with estimated magnitude of -0.38, -0.15 and -0.15, respectively. Out of all water quality time series, the chloride as a stationary series has only one AR and two MA terms with no differencing (I) term. Because of its stationarity, a statistically significant intercept term (δ) was also entered to this model. In the cases of CO_2, TDS, EC, Tur and T an ARIMA (1,1,1) model with one AR and one MA terms was described the series as well. For all selected models there were low correlation between input parameters less than Pankratz threshold (0.9) (Pankratz, 2009). The significance of each model parameters needs to be examined by hypothesis test. Hypothesis testing was done using Student's t distribution. The coefficients related to each of parameters of the models are summarized in Table 5. Statistically non-significant parameters should be removed from each model. As shown in Table 5 the input parameters of all models were statistically significant (p-value < 0.05). However, there is always a correlation between the estimated parameters of a Box and Jenkins model. The correlation matrix of the estimated values for the model parameters may show the degrees of multicollinearity and high correlation between two parameters however may reflect the lack of adequacy of the model. The extremely correlation values above 0.9 (Pankratz, 2009) suggest that some pairs of parameters are not providing independent information and the model needs to be simplified, perhaps by deleting one of parameters. As shown in Table 5 the absolute value of the correlation coefficient less than 0.9 between the parameters in all models also confirm the adequacy of the proposed ARIMA models. The normal pattern of the residuals of proposed ARIMA models were also assessed using Q-Q plots which follow as shown in Fig. 5(d). The ACF and PACF plots of residuals of ARIMA models are shown in Fig. 2(c). As this figure shows there is no statistically significant correlation between residual values for all ARIMA models and the correlation coefficient for all time lags is not significant. Lack of correlation

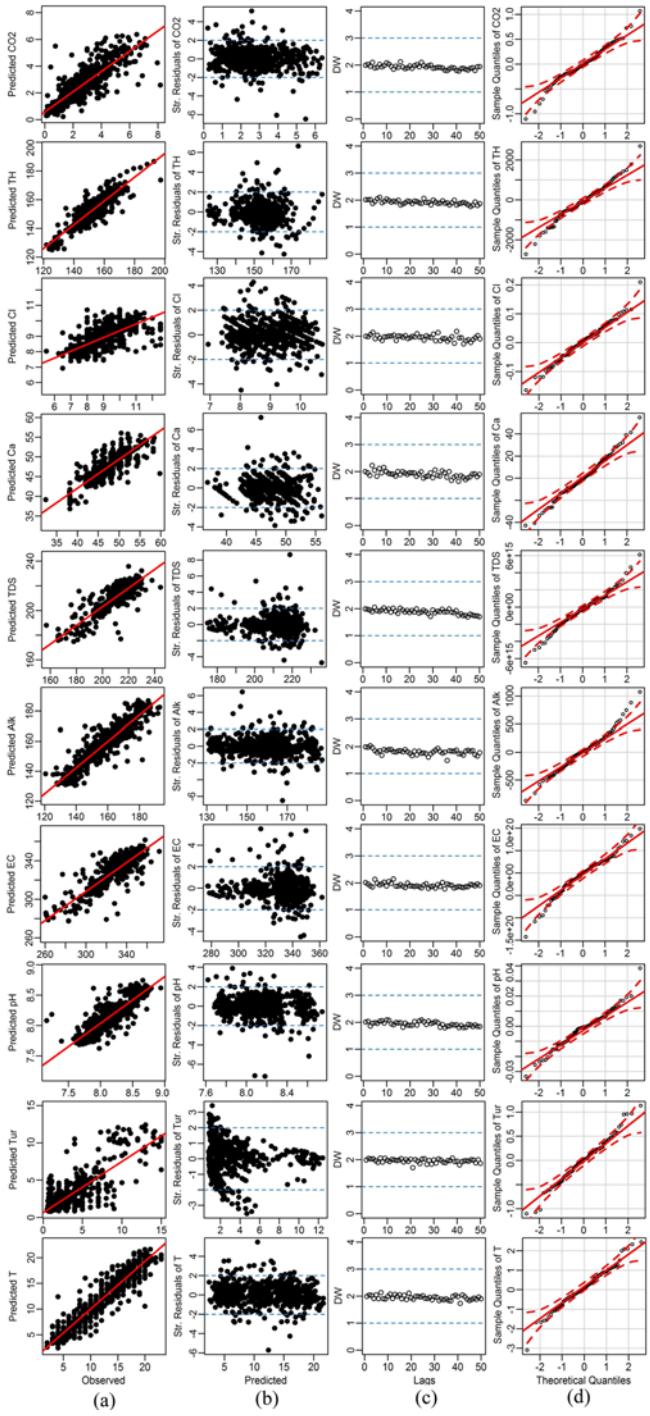

Fig. 5. Diagnostic Plots for ARIMA Models: (a) Predicted Versus Observed Plots, (b) Predicted Versus Standardized Residual, (c) DW Test Statistic Plots, (d) Normal Q-Q Plots

between the residuals confirms the independence of them and indicates models are acceptable. LB test on the residuals of ARIMA models also confirms significantly the independence of the residuals. Fig. 6(a) (se also Fig. 5(a)) show the response plots for the developed various ARIMA models. The bold and light lines represent predicted values by the ARIMA model (Predicted) and observed values (Observed), respectively. The good agreement

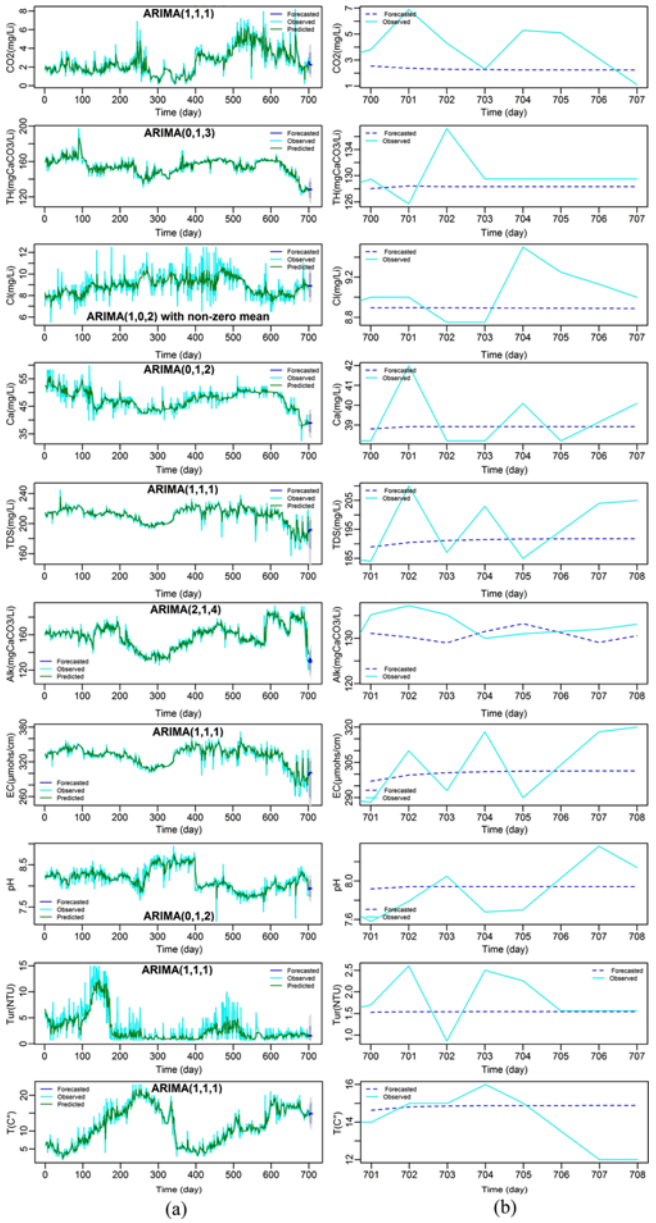

Fig. 6. (a) Predicted Values Response Plot for All Observations, (b) Seven Horizon Daily Plots for Forecasted Values by ARIMA Models

between predicted and observed for all proposed ARIMA models is noteworthy. The DW statistic for ARIMA models residuals at different time lags are shown in Fig. 5(c). The DW statistic was found to be just two and not greater than 3 or less than unity for no cases of time lags. So in all ARIMA models the existence of serial correlation between the residuals could not be confirmed. Although a nonlinear forecasting pattern was evident for NNAR forecasts in Fig. 4(b), ARIMA models horizon-daily forecasted values for next seven days, Fig. 5(b), in most of cases follow a straight line or constant pattern. The forecasting performance of proposed ARIMA and NNAR models was compared and discussed more in next section.

Table 6. Forecasting Performance Results

Series	DM statistic	p-value	R^2	
			ARIMA	NNAR
CO_2	2.1345	0.0167	0.75	0.66
TH	1.5172	0.0650	0.51	0.91
Cl	1.866	0.0314	0.81	0.14
Ca	2.5698	0.0052	0.30	0.10
TDS	1.1037	0.1352	0.77	0.10
Alk	1.104	0.1352	0.27	0.95
EC	1.448	0.0742	0.62	0.56
pH	-0.22469	0.5888	0.41	0.92
Tur	1.1797	0.1195	0.54	0.35
T	-0.13903	0.5552	0.56	0.99

3.5 Models Forecasting Performance Analysis

To assess the forecasting performance of ARIMA and NNAR models a Time Series Cross-Validation (TSCV) analysis was performed because of its simplicity and its universality. The Diebold-Mariano test (Diebold and Mariano, 1995) then was applied to compare the forecast accuracy of two forecast methods (ARIMA and NNAR). The null hypothesis was that the models have same forecast accuracy. The alternative hypothesis was that ARIMA model is more accurate than NNAR model. The results of Diebold-Mariano test which applied on the matrix of Mean Absolute Errors (MAEs) derived (Fig. 7(a)) from TSCV analysis are summarized in Table 6. Fig. 7(b) also shows the actual and forecasted values for seven last observations of each series.

It should be noted that if the p-value for Diebold-Mariano test is significant, with a confidence of 95%, ARIMA would be a model with better accuracy in comparison with NNAR. Otherwise it is not mean necessarily that NNAR is better than ARIMA. Thus three forms of interpretations will be there if we compare DM test with R^2 statistic:

1) DM test (significantly) and R^2 confirm ARIMA: As presented in Table 6 the ARIMA model R^2 statistic for CO_2, Cl and Ca series was obtained 0.75, 0.81 and 0.30, respectively, which were greater than corresponding values for NNAR model, consistently with DM statistic.
2) DM test is not significant and R^2 confirms NNAR: In the case of TH, Alk, pH and T, the DM test p-value is not significant (p-value > 0.05) which mean there is no difference between two models. But R^2 statistic for these variables confirms the better forecasting performance of NNAR.
3) DM test is not significant and R^2 confirms ARIMA: On the other hand, in the case of Tur, EC and TDS, the DM test statistic is not significant while R^2 confirms the better performance of ARIMA model.

Thus to receive a better judgment, the Mean Absolute Error (MAE) as an accuracy measure of model forecasts has been proposed by Hyndman and Koehler (2006). The MAE has favorable properties when compared to other statistics, such as RMSE and R^2, and is therefore recommended for determining comparative accuracy of forecasts (Franses, 2016). Thus any judgment should be based on results of DM test statistic and graphical comparison alone may lead to incorrect interpretations. As shown in Fig. 7(a), the MAEs estimated for forecasted values follow an increasing trend from the first to the seventh days of forecast for both ARIMA and NNAR models. The significance of this difference was confirmed for Cl, Ca, CO_2 time series by Diebold-Mariano test, statistically. These differences confirm the superiority performance of ARIMA at least for Cl, Ca and CO_2 over NNAR. In Fig. 7(b), the forecasted values by ARIMA are following a straight line while for NNAR because of its nonlinearity nature, a deviated pattern in the forecasted values is evident.

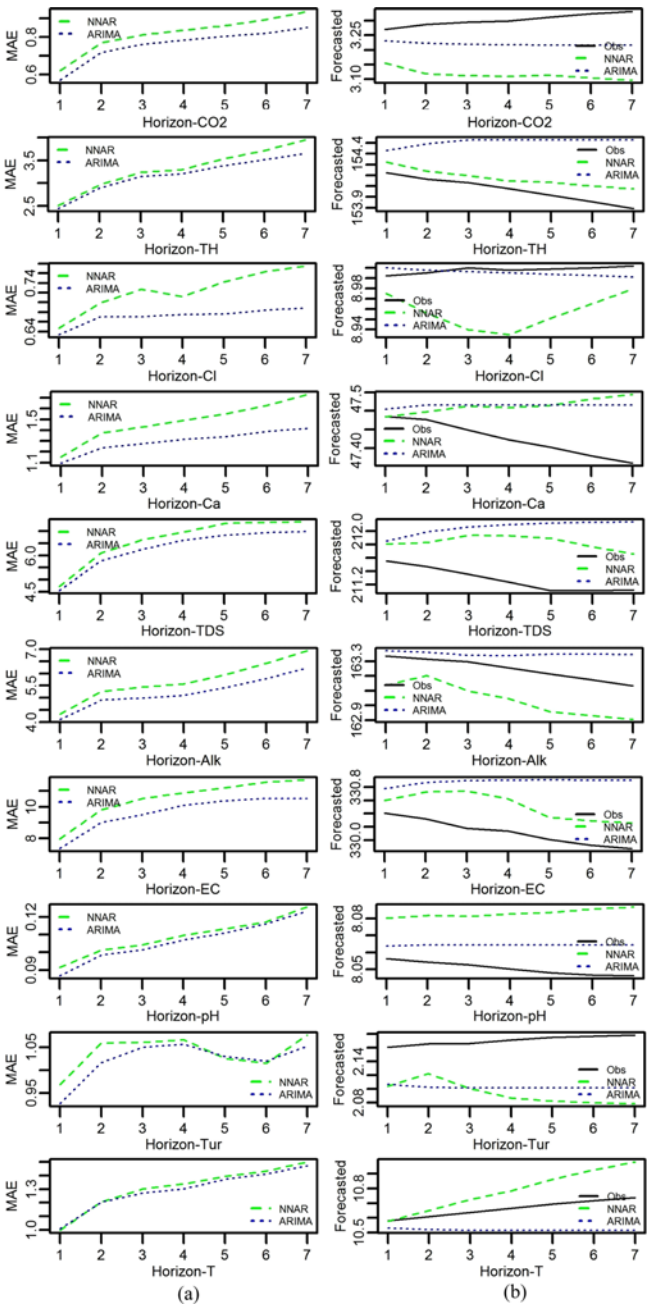

Fig. 7. (a) Estimated MAEs, (b) Forecasted Values versus Horizon Days of Forecasting using TSCV Analysis

Although most of recent research studies suggest that ANNs can be a promising alternative to the traditional linear methods such as ARIMA, based on our findings, these suggestions could not be a general rule and it mainly depends on the nature of the studied variables. For example, in a study conducted by Valipour *et al.* (2013), an artificial neural network model was compared with an ARIMA model and the former was chosen as the best model for forecasting the inflow of Dez Dam's (a dam in Iran) reservoir. In another study on a repairable system failure (Ho *et al.,* 2002), an ANN model was found to give satisfactory performance compared to the ARIMA model. Prybotuk *et al.* (2000) developed an ANN model for forecasting daily maximum ozone level and compared this with conventional regression and Box–Jenkins ARIMA models. They confirmed ANN model superiority over other models. In contrary some other studies claimed that applying ANNs on phenomena with linear behaviors may produce misleading results, and hence; it is not wise to apply ANNs blindly to any type of data (Zhang, 2003; Khashei and Bijari, 2011).

In overall, in the case of linear problems, ARIMA models may be one of the most popular models in time series forecasting and could be widely applied in order to construct more accurate models.

4. Conclusions

Auto-Regressive Integrated Moving Average (ARIMA) and Neural Network Auto-Regressive (NNAR) modeling approaches were used to predict the influent water characteristics of a water treatment plant. The NNAR models show better prediction in terms of R^2 statistic than the ARIMA models to explain the observed values of a time series. However comparison of the forecasting performance of the studies techniques using Time Series Cross-Validation (TSCV) shows ARIMA models could provide better forecast for future values of some series including Cl, CO_2 and Ca. For other series no significant difference was observed between the forecasting performances of the models, statistically. So to choose an appropriate model in cases where the aim is to forecast the future observations of some series, the forecasting performance of the models is recommend to be examined and compared using TSCV approach.

Acknowledgements

This work is part of the Master's Degree Dissertation which financially supported by Kurdistan University of Medical Sciences. Center for Water Quality Research (CWQR) of Tehran University of Medical Science is also acknowledged for analyzing of data.

References

Akaike, H. (1974). "A new look at the statistical model identification." *Automatic Control, IEEE Transactions on*, Vol. 19, No. 6, pp. 716-723, DOI: 10.1109/TAC.1974.1100705.

Battiti, R. (1992). "First-and second-order methods for learning: Between steepest descent and Newton's method." *Neural Computation*, Vol. 4, No. 2, pp. 141-166, DOI: 10.1162/neco.1992.4.2.141.

Box, G. E. and Cox, D. R. (1964). "An analysis of transformations." *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 26, No. 2, pp. 211-252.

Box, G. E., Jenkins, G. M., and Reinsel, G. C. (2013). *Time series analysis: Forecasting and control,* John Wiley & Sons, DOI: 10.1002/9781118619193.

Diebold, F. X. and Mariano, R. S. (1995). "Comparing predictive accuracy." *Journal of Business & Economic Statistics*, Vol. 13, pp. 253-265, DOI: 10.1198/073500102753410444.

Durbin, J. and Watson G. S. (1950). "Testing for serial correlation in least squares regression: I." *Biometrika*, Vol. 37, Nos. 3-4, pp. 409-428, DOI: 10.2307/2332391.

Field, A. (2009). *Discovering statistics using SPSS:(and sex and drugs and rock 'n'roll).* 3rd ed, Sage, Los Angeles.

Hadi, M., Mesdaghinia, A., Yunesian, M., Nasseri, S., Nodehi, R. N., Tashauoei, H., Jalilzadeh, E., and Zarinnejad, R. (2016). "Contribution of environmental media to cryptosporidiosis and giardiasis prevalence in Tehran: A focus on surface waters." *Environmental Science and Pollution Research*, Vol. 23, No. 19, pp. 19317-19329, DOI: 10.1007/s11356-016-7055-9.

Hagan, M. T., Demuth, H. B., and Beale, M. H. (1996). *Neural network design,* Thomson Learning, Singapore.

Hill, T., Marquez, L., O'Connor, M., and Remus, W. (1994). "Artificial neural network models for forecasting and decision making." *International Journal of Forecasting*, Vol. 10, No. 1, pp. 5-15, DOI: 10.1016/0169-2070(94)90045-0.

Ho, S. L., Xie, M., and Goh, T. N. (2002). "A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction." *Computers & Industrial Engineering*, Vol. 42, Nos. 2-4, pp. 371-375, DOI: 10.1016/S0360-8352(02)00036-0.

Hosseini, H. G., Luo, D., and Reynolds, K. J. (2006). "The comparison of different feed forward neural network architectures for ECG signal diagnosis." *Medical Engineering & Physics*, Vol. 28, No. 4, pp. 372-378, DOI: 10.1016/j.medengphy.2005.06.006.

Howard, D. and Mark, B. (2000). *Neural Network Toolbox For Use with MATLAB: Computation Visualization*, Programming, User's Guide Version 4, MathWorks, Inc.

Hyndman, R. J. and Khandakar, Y. (2007). "Automatic time series for forecasting: The forecast package for R." *Journal of Statistical Softwore*, Vol. 27, No. 3, DOI: 10.18637/jss.vo27.i03.

Hyndman, R. J. and Koehler A. B. (2006). "Another look at measures of forecast accuracy." *International Journal of Forecasting*, Vol. 22, No. 4, pp. 679-88, DOI: 10.1016/j.ijforecast.2006.03.001.

Franses, P. H. (2016). "A note on the mean absolute scaled error." *International Journal of Forecasting*, Vol. 32, No. 1, pp. 20-22, DOI: 10.1016/j.ijforecast.2015.03.008.

Jafari Mansoorian, H., Karimaeec, M., Hadi, M., Jame Porazmey, E., Barati, F., and Baziar, M. (2017). "Feed forward artificial neural network model to estimate the TPH removal efficiency in soil washing process." *Archives of Hygiene Sciences*, Vol. 6, No. 1, pp. 69-104.

Khashei, M. and Bijari, M. (2010). "An artificial neural network (p, d, q) model for timeseries forecasting." *Expert Systems with Applications*, Vol. 37, No. 1, 479-489, DOI: 10.1016/j.eswa.2009.05.044.

Khodadadi, M., Mesdaghinia, A., Nasseri, S., Ghaneian, M. T., Ehrampoush, M. H., and Hadi, M. (2016). "Prediction of the waste stabilization pond performance using linear multiple regression and multi-layer perceptron neural network: A case study of Birjand." *Iran.*

Environmental Health Engineering and Management Journal, Vol. 3, pp. 81-89.

Khashei, M. and Bijari, M. (2011). "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting." *Applied Soft Computing*, Vol. 11, No. 2, pp. 2664-2675, DOI: 10.1016/j.asoc.2010.10.015.

Kusiak, A., Verma, A., and Wei, X. (2012). "A data-mining approach to predict influent quality." *Environmental Monitoring and Assessment*, Vol. 185, pp. 2197-2210, DOI: 10.1007/s10661-012-2701-2.

Ljung, G. M. and Box, G. E. (1978). "On a measure of lack of fit in time series models." *Biometrika*, Vol. 65, No. 2, pp. 297-303, DOI: 10.1093/biomet/65.2.297.

Maest, A. S., Kuipers J., Travers C. L., and Atkins D. A. (2005). *Predicting Water Quality at Hardrock Mines: Methods and Models*, Uncertainties and State-of-the-art, Earthworks, Washington, DC.

Matalas, N. (1967). "Time series analysis." *Water Resources Research*, Vol. 3, No. 3, pp. 817-829, DOI: 10.1029/WR003i003p00817.

Pankratz, A. (2009). *Forecasting with univariate Box-Jenkins models: Concepts and cases,* John Wiley & Sons, New Jersey.

Panno, S. V., Keith, C. H., Hwang, H. H., Greenberg, S., Krapac, I. G., Landsberger, S., and O'Kelly, D. J. (2002). "Source identification of sodium and chloride contamination in natural waters: Preliminary results." *In Proceedings, 12th Annual Illinois Groundwater Consortium Symposium*, Illinois Groundwater Consortium.

Plazl, I., Pipus, G., Drolka, M., and Koloini, T. (1999). "Parametric sensitivity and evaluation of dynamic model for single-stage wastewater treatment plant." *Acta Chimica Slovenica*, Vol. 46, No. 2, pp. 289-300.

R Core Team (2016). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org.

Rostami Fasih, Z., Mesdaghinia, A., Nadafi, K., Nabizadeh Nodehi, R., Mahvi, A. H., and Hadi, M. (2015). "Forecasting the air quality index based on meteorological variables and autocorrelation terms using artificial neural network." *Razi Journal of Medical Sciences*, Vol. 22, No. 137, pp. 31-43.

Rumelhart, D. E. and McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Foundations*, MIT Press Cambridge, MA, USA.

Salas, J. D. Obeysekera J. (1982). "ARMA model identification of hydrologic time series." *Water Resources Research*, Vol. 18, No. 4, pp. 1011-1021, DOI: 10.1029/WR018i004p01011.

Salas, J. D. (1980). *Applied modeling of hydrologic time series,* Water Resources Publication.

Salas, J. D., Boes, D. C., and Smith, R. A. (1982). "Estimation of ARMA models with seasonal parameters." *Water Resources Research,* Vol. 18, No. 4, pp. 1006-1010, DOI: 10.1029/WR018i004p01006.

Solaimany-Aminabad, M., Maleki, A., and Hadi, M. (2013). "Application of Artificial Neural Network (ANN) for the prediction of water treatment plant influent characteristics." *Journal of Advances in Environmental Health Research*, Vol. 1, No. 2, pp. 89-100, DOI: 10.22102/jaehr.2013.40130.

Thoplan, R. (2014). "Simple v/s sophisticated methods of forecasting for mauritius monthly tourist arrival data." *International Journal of Statistics and Applications*, Vol. 4, No. 5, pp. 217-223, DOI: 10.5923/j.statistics.20140405.01.

Valipour, M., Banihabib, M. E., and Behbahani, S. M. R. (2013). "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir." *Journal of Hydrology*, Vol. 476, pp. 433-441, DOI: 10.1016/j.jhydrol.2012.11.017.

Wu, G.-D. and Lo, S.-L. (2008). "Predicting real-time coagulant dosage in water treatment by artificial neural networks and adaptive network-based fuzzy inference system." *Engineering Applications of Artificial Intelligence*, Vol. 21, No. 8, pp. 1189-1195, DOI: 10.1016/j.engappai.2008.03.015.

Wu, G.-D. and Lo, S.-L. (2010). "Effects of data normalization and inherent-factor on decision of optimal coagulant dosage in water treatment by artificial neural network." *Expert Systems with Applications*, Vol. 37, No. 7, pp. 4974-4983, DOI: 10.1016/j.eswa.2009.12.016.

Zhang, G., Eddy Patuwo, B. and Y Hu M. (1998). "Forecasting with artificial neural networks: The state of the art." *International Journal of Forecasting*, Vol. 14, No. 1, pp. 35-62, DOI: 10.1016/S0169-2070(97)00044-7.

Zhang, G. P. (2003). "Time series forecasting using a hybrid ARIMA and neural network model." *Neurocomputing*, Vol. 50, pp.159-175, DOI: 10.1016/S0925-2312(01)00702-0.

Zhang, G. P. and Qi, M. (2005). "Neural network forecasting for seasonal and trend time series." *European Journal of Operational Research*, Vol. 160, No. 2, pp. 501-514, DOI: 10.1016/j.ejor.2003.08.037.