

## Drinking water quality control: control charts for turbidity and pH

Sermin Eelevli, Nevin Uzgören, Deniz Bingöl and Birol Eelevli

### ABSTRACT

Water treatment processes are required to be in statistical control and capable of meeting drinking water specifications. Control charts are used to monitor the stability of quality parameters by distinguishing the in-control and out-of-control states. The basic assumption in standard applications of control charts is that observed data from the process are independent and identically distributed. However, the independence assumption is often violated in chemical processes such as water treatment. Autocorrelation, a measure of dependency, is a correlation between members of a series arranged in time. The residuals obtained from an autoregressive integrated moving averages (ARIMA) time series model plotted on a standard control chart is used to overcome the misleading of standard control charts in the case of autocorrelation. In this study, a special cause control (SCC) chart, also called a chart of residuals from the fitted ARIMA model, has been used for turbidity and pH data from a drinking water treatment plant in Samsun, Turkey. ARIMA (3,1,0) for turbidity and ARIMA (1,1,1) for pH were determined as the best time series models to remove autocorrelation. The results showed that the SCC chart is more appropriate for autocorrelated data to evaluate the stability of the water treatment process, since it provides a higher probability of coverage than an individual control chart.

**Key words** | ARIMA, autocorrelation, drinking water, pH, special cause control chart, turbidity

**Sermin Eelevli** (corresponding author)

**Birol Eelevli**

Industrial Engineering Department,  
Ondokuz Mayıs University,  
Samsun,  
Turkey  
E-mail: [sermin.eelevli@omu.edu.tr](mailto:sermin.eelevli@omu.edu.tr)

**Nevin Uzgören**

Department of Business Administration,  
Dumlupınar University,  
Kutahya,  
Turkey

**Deniz Bingöl**

Chemistry Department,  
Kocaeli University,  
Kocaeli,  
Turkey

### INTRODUCTION

Drinking water treatment plants are responsible for providing high quality drinking water to consumers. High quality drinking water refers to esthetically appealing water that is free of both pathogens (disease causing organisms) and chemical contaminants that have been known to cause undesirable health effects. Monitoring and control are indispensable actions for ensuring the production of high quality drinking water. Recently, quality control charts have been used for the detection of changes in biological and chemical water quality parameters (Mitsakos & Psarakis 2005; Smeti *et al.* 2005, 2006, 2007a, 2007b; Usman & Kontagora 2010). The selection of quality parameters for monitoring water depends upon the purpose for which we are going to use that water, and the extent to which we need its quality and purity. pH is one of the most important operational water quality parameters in determining

the corrosive nature of water. pH control is necessary at all stages of water treatment to ensure satisfactory water clarification and disinfection. Turbidity also adversely affects the efficiency of disinfection, and it is also measured to determine what type and level of treatment are needed. Turbidity in water is caused by suspended matter, such as clay, silt, finely divided organic and inorganic matter, soluble colored organic compounds, plankton and other microscopic organisms (Patil *et al.* 2012).

Control charts are used to monitor two types of process variation, common cause variation and special cause variation. The presence of an assignable variation is an indicator of an out-of-control situation, and is characterized by non-randomness in the data. A standard assumption of traditional control charts is that observations taken from the process are independently distributed, with a constant

mean and variance. However, observations depart from the assumption of independence for many industrial processes because of the advanced measurement technology, shortened sampling interval, and the nature of the process. The existence of autocorrelation in observations causes problems: detecting 'special causes' that do not exist and not detecting 'special causes' that truly exist, implying a high probability of false positives and/or false negatives (Smeti *et al.* 2007a). Alwan & Roberts (1988) proposed a method to deal with data autocorrelation. The method consists of selecting and fitting an autoregressive integrated moving average (ARIMA) model to the process, and then applying a traditional control chart to the residuals. The reason for monitoring residuals instead of actual observations is that they are independent and identically distributed with mean zero when the process is controlled, and remain independent of possible differences in the mean when the process gets out of control (Russo *et al.* 2012). Some examples considering data autocorrelation are given by Stone & Taylor (1995), Reynolds & Lu (1997), Young & Winistorfer (2001), Bisgaard & Kulaçci (2005), Mitsakos & Psarakis (2005), Smeti *et al.* (2005, 2006, 2007a), Eelevli *et al.* (2009), Noskievicova (2009), Karaoglan & Bayhan (2011), Martin & Petr (2012), Tasdemir (2012), Kandanond (2014), and Perzyck & Rodziewicz (2015).

Since quality parameters of drinking water need to be monitored to guarantee the final water quality, it is necessary to use the appropriate control chart based on the data structure. The objective of this study is to compare the results of the traditional control chart (individual control (IC) chart) and the chart of residuals from a fitted ARIMA model (special cause control (SCC) chart) when the independence assumption is not valid. The data under investigation are turbidity and pH, which are routinely monitored on a daily basis.

## THEORETICAL BACKGROUND OF CONTROL CHARTS FOR AUTOCORRELATED DATA

### SCC chart

Setting control limits correctly in control charts is one of the main requirements for statistical quality control to verify the

statistical stability of the analyzed process. Autocorrelation is a measure of dependency between observations. Autocorrelated data are data that are 'self-correlated'. If the quality characteristic that is under investigation exhibits various levels of autocorrelation, control charts that are based on the assumption of independence will give a higher false alarm rate. In other words, the main effect of autocorrelation in the process data on control charts is that it produces control limits that are much tighter than desired (Mitsakos & Psarakis 2005). Therefore, if significant autocorrelation in the process data is present, it is necessary to modify traditional methodology to account for this autocorrelation.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) are used to detect whether autocorrelation is present in the data or not. ACF simply shows how correlation between pairs of values varies as a function of the lag or number of periods ( $k$ ) between them. PACF measures correlation between observations that are  $k$  time periods apart after controlling for correlations at intervening lags. In order to visually assess the autocorrelation in a set of time-ordered data, ACF and PACF plots that show non-zero or significant autocorrelations are usually used.

There are two main approaches for constructing control charts for autocorrelated data. The first approach uses standard control charts, but adjusts the method of estimating the process variance so that the true process variance is being estimated. Therefore, in this chart, correlation is taken care of by placing the control limits according to the changed variability (Samanta & Bhattacharjee 2001).

The second approach that has proven useful in dealing with autocorrelated data is to directly model the correlative structure with an appropriate time series model then use that model to remove autocorrelation from the data, and apply control charts to the residuals (Montgomery 1997). This type of control chart is called an SCC chart.

The SCC chart plots the residuals that are obtained after fitting the process to an ARIMA model, rather than plotting the actual observations. All the traditional tools of process control are applicable to this kind of chart (Alwan & Roberts 1988). Since the mean of the residuals is zero, the centerline is zero. The standard deviation used in this case is the standard deviation of the residuals,  $\sigma_a$ . The limits of

the chart are given by:

$$\text{Upper control limit} = L \cdot \sigma_a$$

$$\text{Center line} = 0$$

$$\text{Lower control limit} = -L \cdot \sigma_a$$

where  $L$  is a constant multiplier, and is usually assumed to be equal to 3 (Mitsakos & Psarakis 2005). Since the residuals will be independently and identically distributed random variables, all the assumptions of traditional control charts will be met.

### ARIMA models

A time series is defined as a set of observations that are ordered in time. The time series analysis aims to investigate the mechanism generating the time series, to forecast future values of the series, to describe the behavior of the series (Russo et al. 2012). ARIMA models are flexible and widely used in time series analysis. They are like regression models, but the independent variables are past values of the time series and past and current values of random terms.

ARIMA models are fitted to the time series data either to better understand the data or to predict future points in the series. The model is generally referred to as an ARIMA ( $p,d,q$ ) model where  $p$ ,  $d$ , and  $q$  denote the number of autoregressive terms, the number of times the series has to be differenced before it becomes stationary, and the number of moving average terms, respectively. The error terms are generally assumed to be independent and identically distributed variables sampled from a normal distribution with zero mean. The general mathematical formula of ARIMA models is as follows:

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1)$$

where

$Y_t$  = the observed value at time period  $t$ , as a linear combination of the values observed at the last  $p$  time periods, a random shock to the system at the current time period  $a_t$ , and random shocks that occurred at the previous  $q$  time periods;

$\theta_0$  = a constant;

$\phi_1, \phi_2, \dots, \phi_p$  = autoregressive parameters;

$\theta_1, \theta_2, \dots, \theta_q$  = moving average parameters.

Based on the theoretical background, the methodology followed in this study is summarized in Figure 1.

## ANALYSIS

### Data

The Samsun drinking water treatment plant is located on an area of approximately 300 decares in Tekkekoy District, about 25 km away from Samsun city center. The water coming by gravity from the Cakmak Dam to the treatment plant 13,346 m away is supplied to the city after being subjected to the processes shown in Figure 2. The daily capacity of the plant is 200,000 m<sup>3</sup> and it serves over 500,000 people. Daily routine analyses (for turbidity, pH, and oxidability) and

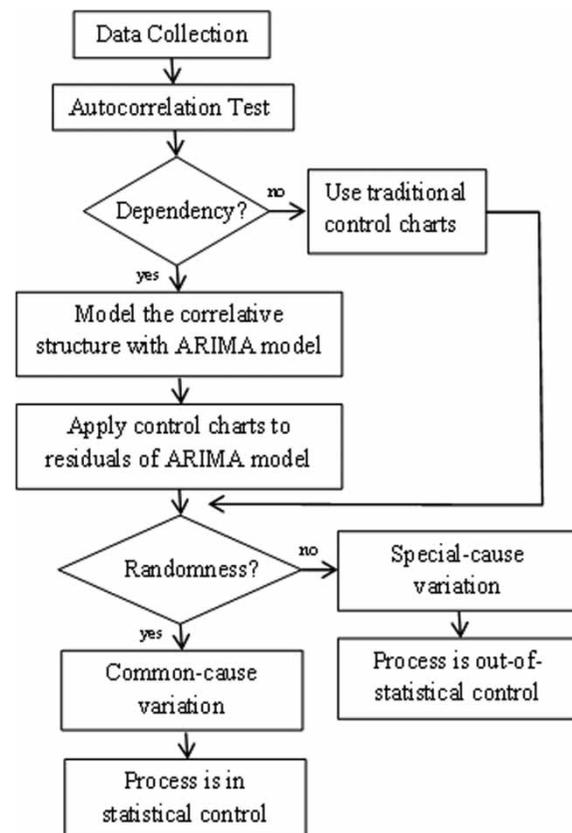


Figure 1 | Research methodology.

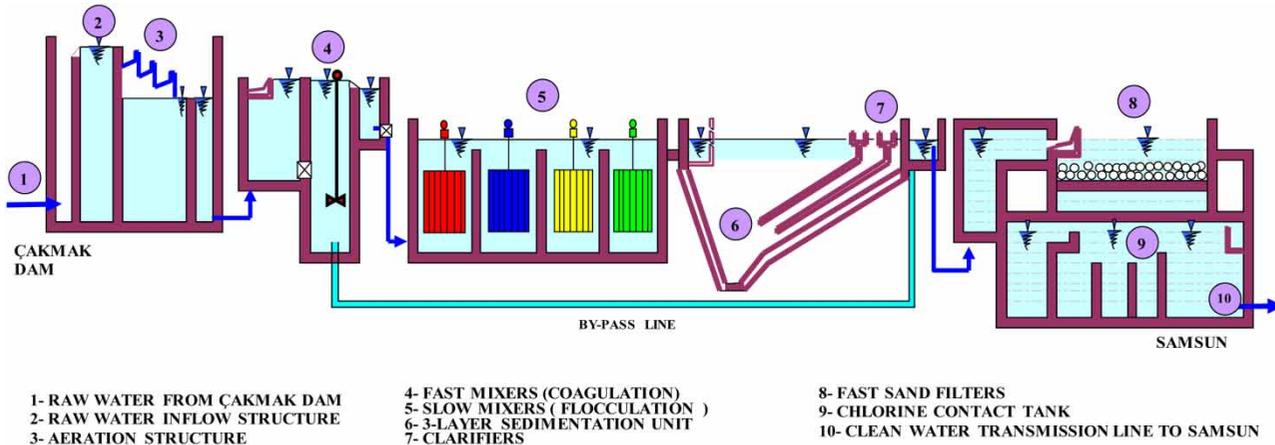


Figure 2 | Flowchart of Samsun drinking water treatment plant.

weekly, biweekly, and monthly chemical and bacteriological analyses are carried out in the Clean Water and Wastewater Laboratories of the General Directorate of Water Treatment to show compliance with the legal standards.

The data gathered on a daily basis covers August and November 2013. Descriptive statistics of turbidity and pH are given in Table 1. STATGRAPHICS Centurion XVI.I was used to analyze data and to construct control charts.

IC charts

IC charts are used to track the process level and detect the presence of assignable causes when the data have been collected one at a time rather than in groups. This type of chart uses the moving range of two successive observations to estimate the process variability.

In order to analyze the variation in turbidity and pH, control charts for individuals have been first established, since the sample size used for process monitoring is n = 1.

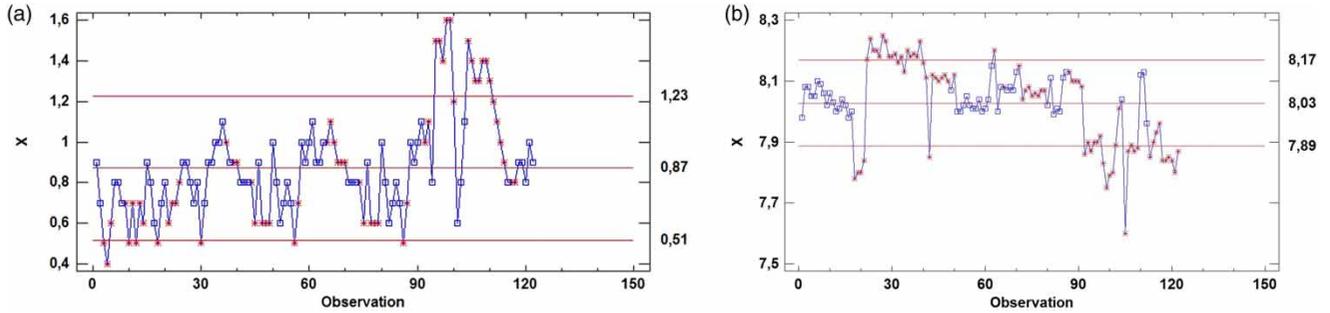
Notice from Figure 3 that the process for both parameters is out of statistical control, since some of the data points fall outside the control limits and the fluctuations of the points around the centerline are irregular and excessive.

Testing autocorrelation and trend

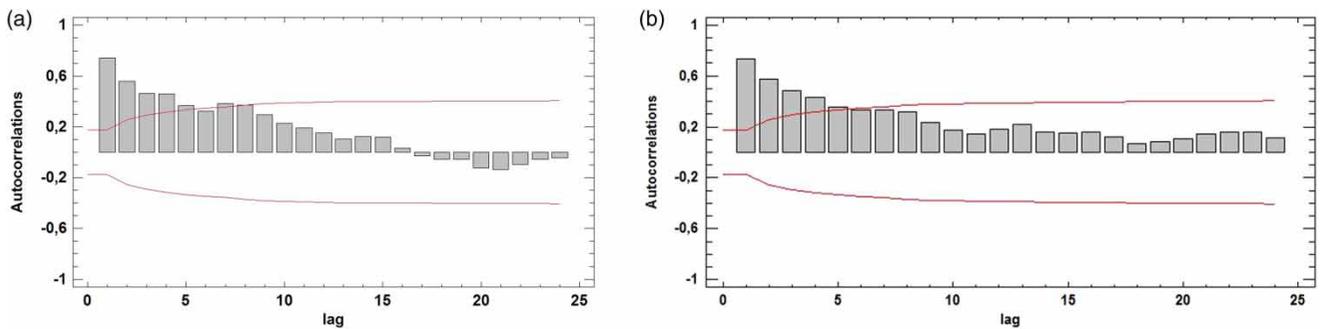
Control charts given in Figure 3 are based on the assumption that there is no correlation between successive observations. However, since the assumption of independence of observations is questionable in practice, as mentioned before, the existence of autocorrelation should be investigated first. The estimated autocorrelations for the data are shown in Figure 4. The straight lines on the graph, which are useful in detecting nonzero correlations, are two standard deviation limits. Bars extending beyond the lines indicate statistically significant autocorrelations for turbidity and pH.

Table 1 | Descriptive statistics

Parameter	Allowed limits	Months	Count	Mean	StDev.	Minimum	Q1	Median	Q3	Maximum	Range
Turbidity	1 NTU	August	31	0.6935	0.1365	0.4000	0.6000	0.7000	0.8000	0.9000	0.5000
		September	30	0.8300	0.1643	0.5000	0.7000	0.8500	1.0000	1.1000	0.6000
		October	31	0.8355	0.1644	0.5000	0.7000	0.9000	1.0000	1.1000	0.6000
		November	30	1.1300	0.2891	0.6000	0.8750	1.1000	1.4000	1.6000	1.0000
pH	6.5–9.5	August	31	8.0610	0.1293	7.7800	8.0000	8.0600	8.1800	8.2500	0.4700
		September	30	8.0873	0.0829	7.8500	8.0175	8.1050	8.1600	8.2300	0.3800
		October	31	8.0723	0.0619	7.8600	8.0500	8.0800	8.1100	8.2000	0.3400
		November	30	7.8867	0.1027	7.6000	7.8400	7.8750	7.9225	8.1300	0.5300



**Figure 3** | IC charts for turbidity (a) and pH (b).

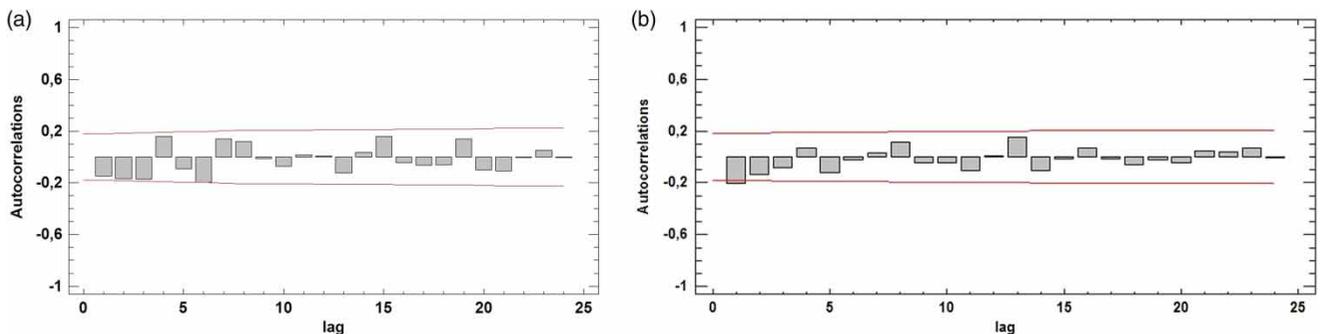


**Figure 4** | Estimated autocorrelations for turbidity (a) and pH (b).

Prior to fitting the ARIMA model, the data should be examined for the presence of any trend, since the identification process for the AR and MA components requires stationary series. A process is considered stationary if its statistical characteristics (mean value, process variance) do not change with time. ACF plots in Figure 4 indicate that the series are non-stationary, since the autocorrelations diminish very slowly. Since the series are not stationary, first order differences were taken. ACF plots in Figure 5 indicated that both of the series are now stationary after first differencing.

### ARIMA model

Transformation of non-stationary process variance of 'random walk' using differential  $d$ -th order is the final process model to describe the stationary ARMA ( $p,q$ ). The  $p$  and  $q$  in ARMA ( $p,q$ ) measure the orders of the autoregressive and moving average components, respectively. To get an idea of what orders to consider, the PACF and ACF were examined. In Figure 5, there is no significant autocorrelation coefficient for turbidity, that is, the ARIMA model for turbidity will not have a moving range component.



**Figure 5** | Estimated autocorrelations for turbidity (a) and pH (b) after first differencing.

However, it is seen that one significant autocorrelation coefficient at lag 1 exists for pH. Therefore, MA(1), having a memory of only one period, should be considered for pH. Notice from Figure 6 that the second and third autocorrelations are statistically significant for turbidity. This suggests the AR (3) model. There is another statistically significant autocorrelation at lag 6, but this can be ignored. Since only the first partial autocorrelation coefficient is significant, the time series for pH has an autoregressive order of 1, called AR (1).

In addition to the above approach based on the ACF and PACF plots of the first-differenced series, several alternative models were also examined for model selection (Table 2). At this stage, statistics calculated from the one-ahead forecast errors were considered. Better ARIMA models have smaller root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) values, which measure the variability of the forecasting errors. If the forecasts are not biased, mean error (ME)

and mean percentage error (MPE) should be close to zero. ARIMA (3,1,0) and ARIMA (1,1,1) were found to be suitable models for turbidity and pH, respectively. The number of differences ( $d$ ) is 1 for both models, since the data were first differenced to render stationarity.

Table 3 provides the final estimates of the parameters. Each of the estimated model coefficients is shown together with a  $t$ -test. Since the  $p$ -values associated with the coefficients are less than 0.05, the coefficients are significantly different from 0 at the 5% significance level.

### SCC chart

SCC charts based on the residuals of ARIMA models show that two points for turbidity and six points for pH are beyond the control limits (Figure 7). This means that the water treatment process for both parameters is not in control and the source of the variation should be investigated. IC charts in Figure 3 also indicated an unstable or

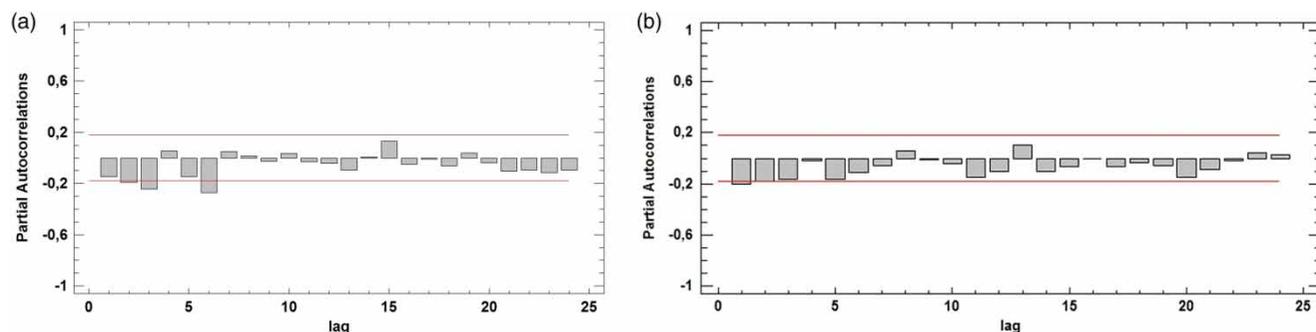


Figure 6 | Estimated partial autocorrelations for turbidity (a) and pH (b) after first differencing.

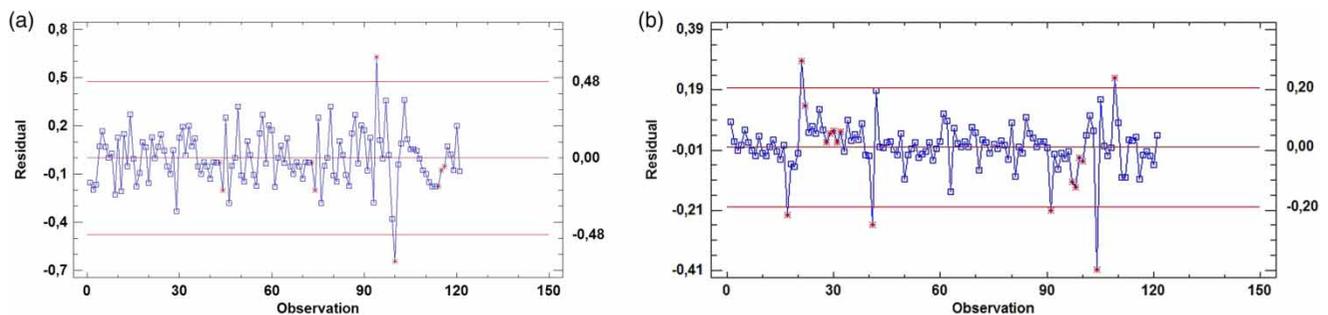
Table 2 | Model comparison

Parameter	Model	RMSE	MAE	MAPE	ME	MPE
Turbidity	ARIMA (3,1,0)	0.1720	0.1278	16.0378	-0.0012	-3.1232
	ARIMA (2,1,0)	0.1775	0.1292	16.4177	-0.0004	-3.0140
	ARIMA (3,2,0)	0.2029	0.1534	18.5821	0.0001	-1.5897
	ARIMA (3,1,1)	0.1729	0.1284	16.1178	-0.0013	-3.1543
	ARIMA (0,1,2)	0.1726	0.1305	16.4740	-0.0015	-3.5570
pH	ARIMA (1,1,1)	0.0844	0.0536	0.6713	0.0016	0.0085
	ARIMA (1,1,2)	0.0845	0.0536	0.6712	0.0017	0.0104
	ARIMA (3,1,1)	0.0847	0.0540	0.6752	0.0056	0.0584
	ARIMA (0,0,1)	0.0993	0.0746	0.9334	0.0003	-0.0137
	ARIMA (1,0,0)	0.0848	0.0529	0.6625	-0.0001	-0.0123

RMSE, root mean squared error; MAE, mean absolute error; MAPE, mean absolute percentage error; ME, mean error; MPE, mean percentage error.

**Table 3** | Parameter estimates

Parameter	Model	Model parameter	Estimate	Std. Error	t	P-value
Turbidity	ARIMA (3,1,0)	AR (1)	-0.232335	0.0886801	-2.61993	0.009961
		AR (2)	-0.250073	0.0880416	-2.8404	0.005316
		AR (3)	-0.261968	0.089041	-2.9421	0.003931
		Mean	0.00133828	0.00913336	0.146527	0.883758
		Constant	0.00233447			
pH	ARIMA (1,1,1)	AR (1)	0.610754	0.098874	6.17709	0.000000
		MA (1)	0.937621	0.0471965	19.8663	0.000000
		Mean	-0.00135921	0.00160166	-0.848629	0.397806
		Constant	-0.000529069			

**Figure 7** | SCC charts for turbidity (a) and pH (b).

out-of-control process. However, they have generated more out-of-control points than the SCC charts because of the violation of the independence assumption. The existence of autocorrelation in the data caused a substantial increase in the false alarm rate.

## RESULTS

Traditional control charts require that observations from the process are independent of one another. Failure to meet this requirement causes the traditional control charts to give a huge number of false alarms. Even a very low degree of autocorrelation in observations has a substantial effect on the performance of these charts. SCC charts, traditional control charts applied to monitor the residuals of the ARIMA model, are more appropriate in a correlated environment as they provide a higher probability of coverage than the individual charts.

In this study, SCC charts for turbidity and pH are based on ARIMA (3,1,0) and ARIMA (1,1,1) models, respectively.

In other studies, Smeti *et al.* (2006, 2007a) used ARIMA (0,1,1). Elevli *et al.* (2009), Noskievicova (2009), and Karooglan & Bayhan (2011) used ARIMA (1,0,0) or AR (1). ARIMA (1,0,1), ARIMA (0,1,1) and ARIMA (0,1,2) were used by Tasdemir (2012) and Kandanand (2014). Different models are the results of ACF and PACF plots used during determination of models.

According to the SCC charts given in Figure 7, the water treatment process is not in control statistically for both of the parameters. The outside points for turbidity are observed after the 90th data point. This means that the variability in November was increased compared to other months. According to the descriptive statistics of Table 1, the process is also not capable of meeting the allowed upper limit (1 NTU). Consequently, some out of specification drinking water is supplied to consumers from the turbidity point of view. A number of out-of-control points for pH are higher than turbidity. However, the observed variation for pH is narrower than the specifications (6.5–9.5). This situation is an indication that the process is capable. Since the objective of water treatment plants is to provide drinking water that is

safe and acceptable to consumers, it is necessary to investigate the assignable causes of process variation and to eliminate them.

## REFERENCES

- Alwan, L. C. & Roberts, H. V. 1988 Time-series modeling for statistical process control. *Journal of Business & Economic Statistics* **6** (1), 87–95.
- Bisgaard, S. & Kulahci, M. 2005 Quality quandaries: the effect of autocorrelation on statistical process control procedures. *Quality Engineering* **17**, 481–489.
- Elevli, S., Uzgoren, N. & Savas, M. 2009 Control charts for autocorrelated colemanite data. *Journal of Scientific & Industrial Research* **68**, 11–17.
- Kandanand, K. 2014 Guidelines for applying statistical quality control method to monitor autocorrelated processes. *Procedia Engineering* **69**, 1449–1458.
- Karaoglan, A. D. & Bayhan, G. M. 2011 Performance comparison of residual control charts for trend stationary first order autoregressive processes. *Gazi University Journal of Science* **24** (2), 329–339.
- Martin, K. & Petr, K. 2012 The usage of time series control charts for financial process analysis. *Journal of Competitiveness* **4** (3), 29–45.
- Mitsakos, J. & Psarakis, S. 2005 On some applications of SPC techniques on water data. In: *7th Hellenic European Conference on Computer Mathematics and Its Applications*, 22–24 September, Athens, Greece.
- Montgomery, D. C. 1997 *Introduction to Statistical Quality Control*, 3rd edn. John Wiley & Sons, New York, USA.
- Noskievicova, D. 2009 Statistical analysis of the blast furnace process output parameter using ARIMA control chart with proposed methodology of control limits setting. *Metalurgija* **48** (4), 281–284.
- Patil, P. N., Sawant, D. V. & Deshmukh, R. N. 2012 Physico-chemical parameters for testing of water – a review. *International Journal of Environmental Sciences* **3** (3), 1194–1207.
- Perzyk, M. & Rodziewicz, A. 2015 Application of special cause control charts to green sand process. *Archives of Foundry Engineering* **15** (4), 55–60.
- Reynolds, M. R. & Lu, C.-W. 1997 Control charts for monitoring processes with autocorrelated data. *Nonlinear Analysis, Theory, Methods & Applications* **30** (7), 4059–4067.
- Russo, S. L., Camargo, M. E. & Fabris, J. P. 2012 Applications of control charts Arima for autocorrelated data. In: *Practical Concepts of Quality Control* (M. S. F. Nezhad, ed.). Intech, Croatia.
- Samanta, B. & Bhattacharjee, A. 2001 An investigation of quality control charts for autocorrelated data. *Mineral Resources Engineering* **10** (1), 53–69.
- Smeti, E., Koronakis, D. & Kousouris, L. 2005 Statistical process techniques on water toxicity data. In: *7th Hellenic European Conference on Computer Mathematics and Its Applications*, 22–24 September, Athens, Greece.
- Smeti, E. M., Kousouris, L. P., Tzoumerkas, P. C. & Golfinopoulos, S. K. 2006 Statistical process control techniques on autocorrelated turbidity data from finished water tank. In: *Proceedings of an International Conference on Water Science and Technology Integrated Management on Water Resources – AQUA 2006*. Athens, Greece.
- Smeti, E. M., Koronakis, D. E. & Golfinopoulos, S. K. 2007a Control charts for the toxicity of finished water – modeling the structure of toxicity. *Water Research* **41**, 2679–2689.
- Smeti, E. M., Thanasoulis, N. C., Kousouris, L. P. & Tzoumerkas, P. C. 2007b An approach for the application of statistical process control techniques for quality improvement of treated water. *Desalination* **213**, 273–281.
- Stone, R. & Taylor, M. 1995 Time series models in statistical process control: considerations of applicability. *The Statistician* **44** (2), 227–234.
- Tasdemir, A. 2012 Effect of autocorrelation on the process control charts in monitoring of a coal washing plant. *Physicochemical Problems of Mineral Processing* **48** (2), 495–512.
- Usman, A. & Kontagora, N. M. 2010 Statistical process control on production: a case study of some basic chemicals used in pure water production. *Pakistan Journal of Nutrition* **9** (4), 387–391.
- Young, T. M. & Winistorfer, P. M. 2001 The effects of autocorrelation on real-time statistical process control with solutions for forest products manufacturers. *Forest Products Journal* **51** (11/12), 70–77.

First received 26 January 2016; accepted in revised form 17 August 2016. Available online 26 September 2016