

# Water quality management using statistical analysis and time-series prediction model

Kulwinder Singh Parmar · Rashmi Bhardwaj

Received: 8 October 2013 / Accepted: 12 February 2014 / Published online: 14 March 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** This paper deals with water quality management using statistical analysis and time-series prediction model. The monthly variation of water quality standards has been used to compare statistical mean, median, mode, standard deviation, kurtosis, skewness, coefficient of variation at Yamuna River. Model validated using R-squared, root mean square error, mean absolute percentage error, maximum absolute percentage error, mean absolute error, maximum absolute error, normalized Bayesian information criterion, Ljung–Box analysis, predicted value and confidence limits. Using auto regressive integrated moving average model, future water quality parameters values have been estimated. It is observed that predictive model is useful at 95 % confidence limits and curve is platykurtic for potential of hydrogen (pH), free ammonia, total Kjeldahl nitrogen, dissolved oxygen, water temperature (WT); leptokurtic for chemical oxygen demand, biochemical oxygen demand. Also, it is observed that predicted series is close to the original series which provides a perfect fit. All parameters except pH and WT cross the prescribed limits of the World Health Organization /United States Environmental Protection Agency, and thus water is not fit for drinking, agriculture and industrial use.

**Keywords** Statistical analysis · ARIMA · Time-series analysis · Prediction model

## Introduction

Yamuna is the largest tributary river of the Ganga in northern India. It originates from the Yamunotri glacier at a height of 6,387 m on the south western slopes of Banderpooch peaks ( $38^{\circ} 59' N 78^{\circ} 27' E$ ) in the lower Himalayas in Uttarakhand. It travels a total length of 1,376 km by crossing several states, Uttarakhand, Haryana, Himachal Pradesh, Delhi, Uttar Pradesh and has a mixing of drainage system of 366,233 km<sup>2</sup> before merging with the Ganga at Allahabad i.e., a total of 40.2 % of the entire Ganga basin. The river accounts for more than 70 % of Delhi's water supplies and about 57 million people depend on river water for their daily usage (CPCB 2006).

Hathnikund is approximately 157 km downstream from Yamunotri and 2 km upstream from Tajewala barrage. Hathnikund barrage is 38 km downstream from Dakpathar and 2 km upstream from Tajewala barrage. Sample location (Hathnikund) provides water quality after joining of the tributaries Tons, Giri and Asan of the lower Himalaya region of Yamuna River. The study of quality of water at Hathnikund is important because after this station Yamuna River enters Delhi (capital of India) and accounts for more than 70 % of Delhi's water supplies and about 57 million people depend on river water for their daily usage (CPCB 2006).

Pollution in river water is continuously increasing due to urbanization, industrialization etc. and most of the rivers are at dying position which is an alarming signal (Parmar et al. 2009; Phiri et al. 2005). Physico-chemical parameters, trace metals have effects of industrial wastes, municipal sewage, and agricultural runoff on river water quality (Akoto and Adiyiah 2007; Alam et al. 2007; Banu et al. 2007; Juang et al. 2008). The analysis of the simultaneous effect of water pollution and eutrophication on the concentration of dissolved oxygen (DO) in a water body shows

K. S. Parmar · R. Bhardwaj (✉)  
Non-linear Dynamics Lab, Department of Mathematics,  
University School of Basic and Applied Sciences, Guru Gobind  
Singh Indraprastha University, Dwarka, Delhi 110078, India  
e-mail: rashmib22@gmail.com

K. S. Parmar  
e-mail: kulmaths@gmail.com

that the decrease in the concentration of DO is much more than when only single effect is present in the water body, thus leading to more uncertainty about the survival of DO-dependent species (Kumar and Dua 2009; Shukla et al. 2008). Trihalomethanes compounds were determined in the drinking water samples at consumption sites and treatment plants of Okinawa and Samoa Islands and observed that the chloroform, bromodichloromethane compounds exceed the level of Japan water quality and World Health Organization (WHO) standards (APHA 1995; Imo et al. 2007; WHO 1971). Water quality modeling using hydrochemical data, multiple linear regression, structural equation, predictability, trend and time-series analysis provides major tools for application in water quality management (Attah and Bankole 2012; Chenini and Khemiri 2009; Fang et al. 2010; Singh et al. 2004; Su et al. 2011; Prasad et al. 2013; Seth et al. 2013). Water quality managers use regression equations to estimate constituent concentrations for comparison of current water quality conditions to water quality standards (Joarder et al. 2008; Korashey 2009; Psargaonkar et al. 2008; Singh et al. 2004; Vassilis et al. 2001; Ravikumar et al. 2013).

Climatic dynamic also plays an important role in determining the water quality standards using fractal dimensional analysis, trend and analyzed time-series data of three major dynamic components of the climate i.e., temperature, pressure and precipitation (Bhardwaj and Parmar 2013a, 2013b; Rangarajan 1997; Damodhar and Reddy 2013). It is analyzed that regional climatic models would not be able to predict local climate as it deals with averaged quantities and that precipitation during the southwest monsoon is affected by temperature and pressure variations during the preceding winter (Kahya and Kalayci 2004; McCleary and Hay 1980; Mousavi et al. 2008; Movahed and Hermanisc 2008; Park and Park 2009; Prasad and Narayana 2004; Rangarajan and Ding 2000; Rangarajan and Sant 2004).

Yamuna River is the most vulnerable polluted water body because of the role in carrying municipal, industrial wastes, and run offs from agriculture lands in vast drainage basins. Detailed water quality management research is needed to maintain water quality standards. The quality of Yamuna River water depends upon the quality of water parameters potential of hydrogen (pH), chemical oxygen demand (COD), biochemical oxygen demand (BOD), dissolved oxygen (DO), water temperature (WT), free ammonia (AMM) and total Kjeldahl nitrogen (TKN). In this paper, statistical analysis, time series, auto regressive integrated moving average (ARIMA), stationary R-square, R-square, root mean square error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE), normalized Bayesian information criterion (BIC) of these



**Fig. 1** River Yamuna route map description of sample site

water parameters have been estimated at Hathnikund (Haryana, India) of Yamuna River as shown in Fig. 1.

## Methodology

The monthly average value of the last 10 years of water quality parameters pH, COD, BOD, AMM, TKN, DO, and WT observed by Central Pollution Control Board (CPCB) at Hathnikund of Yamuna River in Delhi (India) has been considered for the present study.

## Statistical analysis

Statistical analysis is used to calculate mean, median, mode, standard deviation, kurtosis, skewness, and coefficient of variation. Mean explains average value; median gives the middle values of an ordered sequence or positional average; mode is defined as the value which occurs the maximum number of times that is having the maximum frequency; standard deviation gives a measure of “spread” or “variability” of the sample; kurtosis refers to the degree of flatness or peakedness in the region about the mode of a frequency curve; skewness describes the symmetry of the data; coefficient of variation gives the relative measure of the sample (Bhardwaj and Parmar 2013a, 2013b).

## R-squared and stationary R-square

R-squared is an estimate of proportion of total variation in the series which is explained by the model and measure is useful when the series is stationary. Stationary R-squared is a measure that compares stationary part of the model to a simple mean model and is preferable to ordinary R-squared when there is a trend or seasonal pattern. Stationary R-squared can be negative with a range of negative infinity to 1. Negative values mean that the model under consideration is worse than the baseline model. Positive values mean that the model under consideration is better than the baseline model (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980).

## RMSE

RMSE is a measure of variation of the dependent series from its model-predicted level, expressed in the same units as the dependent series (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980). RMSE of an estimator  $\hat{\theta}$  with respect to estimator parameter  $\theta$  is defined as:

$$\text{RMSE}(\hat{\theta}) = \sqrt{E\left[(\hat{\theta} - \theta)^2\right]} \quad (1)$$

## MAPE

MAPE is a measure of variation of dependent series from its model-predicted level. It is independent of the units used and can therefore be used to compare series with different units. It usually expresses accuracy as a percentage and is defined as:

$$M = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (2)$$

where  $A_t$  is the actual value and  $F_t$  is the forecast value. For perfect fit, the value of MAPE is zero, but for upper level the MAPE has no restriction (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980).

## MAE

MAE measures variation of series from its model-predicted level and is reported in the original series units. Also, the MAE is a quantity used to measure variation of forecasts or predictions from the eventual outcome. It is given by

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |F_i - A_i| = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (3)$$

where  $e_i$  is absolute error,  $F_i$  is prediction and  $A_i$  is calculated value. It is a common measure of forecast error in

time-series analysis (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980).

## Maximum absolute percentage error (MaxAPE)

MaxAPE measures largest forecasted error, expressed as a percentage. This measure is useful for imagining a worst-case scenario for the forecasts (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980).

## Maximum absolute error (MaxAE)

MaxAE measures largest forecasted error, expressed in same units as of dependent series. It is useful for imagining the worst-case scenario for the forecasts. MaxAE and MaxAPE may occur at different series points. When absolute error for large series value is slightly larger than absolute error for small series value, then MaxAE will occur at larger series value and MaxAPE will occur at smaller series value (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980).

## Normalized BIC

Normalized BIC is general measure of overall fit of a model that attempts to account for model complexity. It is a score based upon mean square error and includes a penalty for number of parameters in the model and length of series (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980).

$$\text{BIC} = \chi^2 + k \cdot \ln(n) \quad (4)$$

BIC is used to determine the best model the constant ( $k$ ). It can measure the efficiency of parameterized model in terms of predicting the data also it penalizes the complexity of the model where complexity refers to the number of parameters in model.

## Time series

Time series is a sequence of data points, measured typically at successive times spaced at uniform time intervals. Time-series analysis comprises methods for analyzing time-series data to extract meaningful statistics and other characteristics of data and to forecast future events based on known past events to predict data points before these are measured. Time-series model reflect that observations close together in time one closely related than observations further apart. In addition, time-series models will often make use of natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (Lu et al. 2013).

ARIMA

ARIMA model of a time series is defined by three terms ( $p, d, q$ ). Identification of a time series is process of finding integer, usually very small (e.g., 0, 1, or 2), values of  $p, d$ , and  $q$  model patterns in data. When value is 0, element is not needed in model. The middle element,  $d$ , is investigated before  $p$  and  $q$ . The goal is to determine if process is stationary and, if not, to make it stationary before determining the values of  $p$  and  $q$ . A stationary process has a constant mean and variance over time period of study. The representation of an auto regressive model in time series (Box et al. 2008; DeLurgio 1998; McCleary and Hay 1980), well known as AR( $p$ ), is defined as

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \tag{5}$$

where the term  $\varepsilon_t$  is source of randomness and is called white noise,  $\alpha_i$  are constants. It is assumed to have the following characteristics:

$$E[\varepsilon_t] = 0,$$

$$E[\varepsilon_t^2] = \sigma^2,$$

$$E[\varepsilon_t \varepsilon_s] = 0 \text{ for all } t \neq s$$

A series may have both auto regressive and moving average components so both types of correlations are required to model the patterns. If both elements are present only at lag 1, to understand this let the linear equation

$$y_t = x_t \beta + \varepsilon_t \tag{6}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \tag{7}$$

where  $-1 < \rho < 1$ .

where  $v_t$  is independent and identically distributed (iid) and from above expectation values

$$E(v_t, v_{t-1}) = 0 \tag{8}$$

As we learn in a moment, the disturbance or error in this model is said to follow a first order auto regressive (AR1) process. Thus, the current error is part of the previous error plus some shock. So Eq. (6) can be re written as

$$y_t = x_t \beta + \rho \varepsilon_{t-1} + v_t \tag{9}$$

Also, we know that

$$y_{t-1} = x_{t-1} \beta + \varepsilon_{t-1} \tag{10}$$

$$\Rightarrow \varepsilon_{t-1} = y_{t-1} - x_{t-1} \beta \tag{11}$$

From Eq. (9)  $y_t = x_t \beta + \rho(y_{t-1} - x_{t-1} \beta) + v_t$ .

$$y_t = x_t \beta + \rho y_{t-1} - \rho x_{t-1} \beta + v_t \tag{12}$$

Results and discussion

Figure 2 and Table 1 give the detail of statistical analysis including mean, median, mode, range, standard deviation,

kurtosis, skewness, coefficient of variation for all water quality parameters. Table 2 and Eqs. (1)–(12) depict time-series analysis of ARIMA model, stationary R-squared, R-squared, RMSE, MAPE, MaxAPE, MAE, MaxAE, normalized BIC, Ljung–Box analysis, predicted value, lower confidence limit (LCL), upper confidence limit (UCL), residual for all water quality parameters. Figure 3 plots time series of observed data, best fit, LCL, UCL and ARIMA prediction monthly values for next 5 years for all water parameters. It is observed that for:

pH

Average, positional average and mode value of pH is 7.77, 7.72 and 7.7. These values are close to 7.7 thus data exhibit normal behavior. Standard deviation (SD) is 0.422, skewness approximates to 0 thus pH is symmetrical and values are very close to each other. Curve is platykurtic as kurtosis is less than 3. Stationary R-squared and R-squared values exhibit similar behavior thus model is better than baseline model and RMSE values are low so dependent series is closed with its model-predicted level. For all sites, using Ljung–Box Q(18) model statistics is 23.936, significance is 0.121 and degree of freedom is 17. ARIMA (0, 0, 1) model fitted and boundary lines are at 95 % confidence limits. Predicted, LCL, UCL and residual values of pH are 7.7672, 6.9887, 8.5458, and  $-0.00017$ .

COD

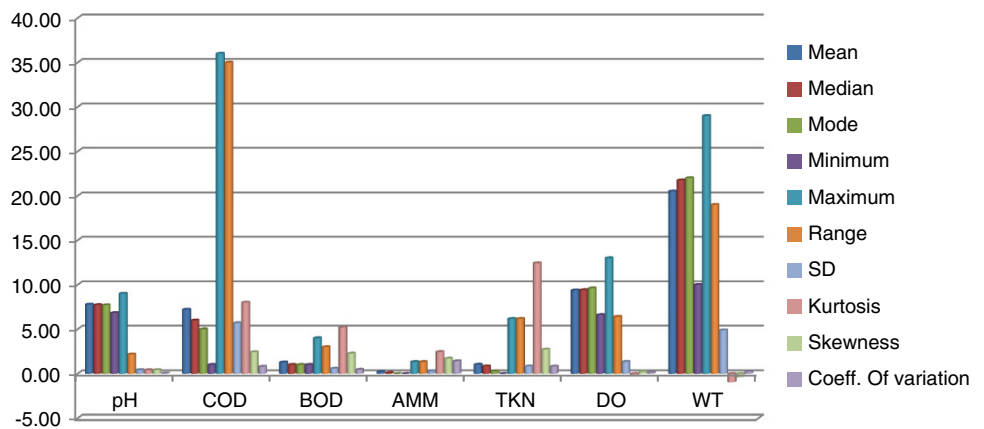
Mean, median and mode value is 7.2, 6.0 and 5.0, respectively, thus curve is not normal. Standard deviation value is high (5.6932807) thus values of COD are not close to each other. It is skewed (2.435) and curve is leptokurtic because kurtosis is more than 3. Stationary R-squared and R-squared values exhibit the similar behavior thus model is better than the baseline model and RMSE values are high so dependent series is not closed with its model-predicted level. Using Ljung–Box Q(18) model statistics is 14.645, significance is 0.477 and degree of freedom is 15. ARIMA (1, 0, 0) model fitted and boundary lines at 95 % confidence limits. Predicted, LCL, UCL and residual values of COD are 8.1648,  $-2.1331$ , 18.4629, and 0.3341.

BOD

Mean and median value is approximately equal but mode value is different so curve is not normal. Standard deviation value suggests that data are spread out and curve is leptokurtic. Stationary R-squared and R-squared values exhibit the similar behavior thus model is better than the baseline model and RMSE values are low so dependent



**Fig. 2** Statistical analysis of Yamuna River water at Hathnikund



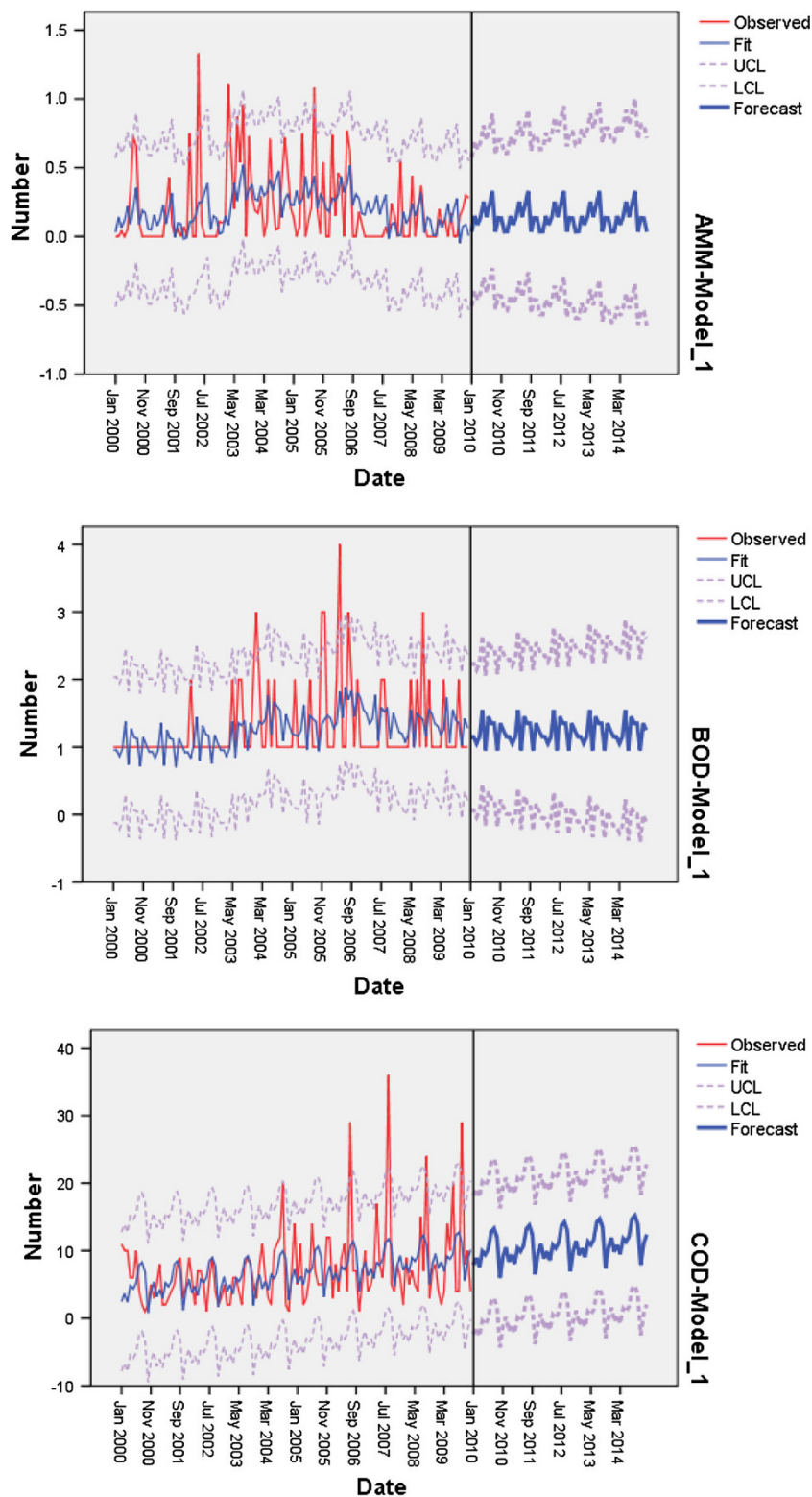
**Table 1** Statistical analysis of water quality parameters

Parameters	pH	COD	BOD	AMM	TKN	DO	WT
Mean	7.77	7.20	1.28	0.20	1.02	9.37	20.51
Median	7.72	6.00	1.00	0.09	0.84	9.40	21.75
Mode	7.7	5	1	0	0.27	9.6	22
Minimum	6.83	1.00	1.00	0.00	0.00	6.60	10.00
Maximum	9.00	36.00	4.00	1.33	6.16	13.00	29.00
Range	2.17	35.00	3.00	1.33	6.16	6.40	19.00
SD	0.4219	5.6932	0.5792	0.2879	0.8304	1.3439	4.8874
Kurtosis	0.4008	8.0001	5.2042797	2.4510	12.4323	-0.2292	-0.9596
Skewness	0.4235	2.4348	2.2676	1.7049	2.7330	0.2411	-0.4004
Coeff. of variation	0.0543	0.7907	0.4542	1.4132	0.8128	0.1434	0.2383

**Table 2** Time-series analysis of water quality parameters

Fit statistic		pH	COD	BOD	AMM	TKN	DO	WT
Stationary R <sup>2</sup>	Mean	0.184	0.778	0.701	0.793	0.691	0.813	0.747
R-squared		0.184	0.182	0.121	0.109	0.149	0.457	0.762
RMSE		0.383	5.192	0.545	0.273	0.769	1.001	2.392
MAPE		3.906	74.282	28.181	110.435	69.896	8.169	9.783
MaxAPE		12.579	588.665	89.094	1.43E + 03	442.932	47.161	45.082
MAE		0.304	3.611	0.386	0.201	0.502	0.75	1.893
MaxAE		1.005	24.257	2.184	1.08	4.83	3.396	6.789
Normalized BIC		-1.84	3.414	-1.133	-2.518	-0.444	0.081	1.824
Statistics	Ljung-Box Q(18)	23.936	14.645	31.755	26.689	16.311	16.161	22.148
DF		17	15	16	16	16	16	16
Sig.		0.121	0.477	0.011	0.045	0.431	0.442	0.138
ARIMA model	Prediction	(0, 0, 1)	(1, 0, 0)	Simple	Simple	(1, 0, 0)	(0, 0, 6)	(1, 0, 13)
Predicted value		7.7672	8.1648	1.2477	0.1832	0.9066	9.3279	20.5947
LCL		6.9887	-2.1331	0.1189	-0.3815	-0.6859	7.2570	15.6442
UCL		8.5458	18.4629	2.3765	0.7480	2.4991	11.3988	25.5453
Residual		-0.00017	0.3341	0.0163	0.00038	-0.0414	-0.0756	-0.0136

**Fig. 3** Time-series (ARIMA model) prediction of Yamuna River water at Hathnikund



series is closed with its model-predicted level. Using Ljung–Box Q(18) model statistics is 31.755, significance is 0.011 and degree of freedom is 16. ARIMA simple model

fitted and boundary lines at 95 % confidence limits. Predicted, LCL, UCL and residual values of BOD are 1.2477, 0.1189, 2.3765, and 0.0163.

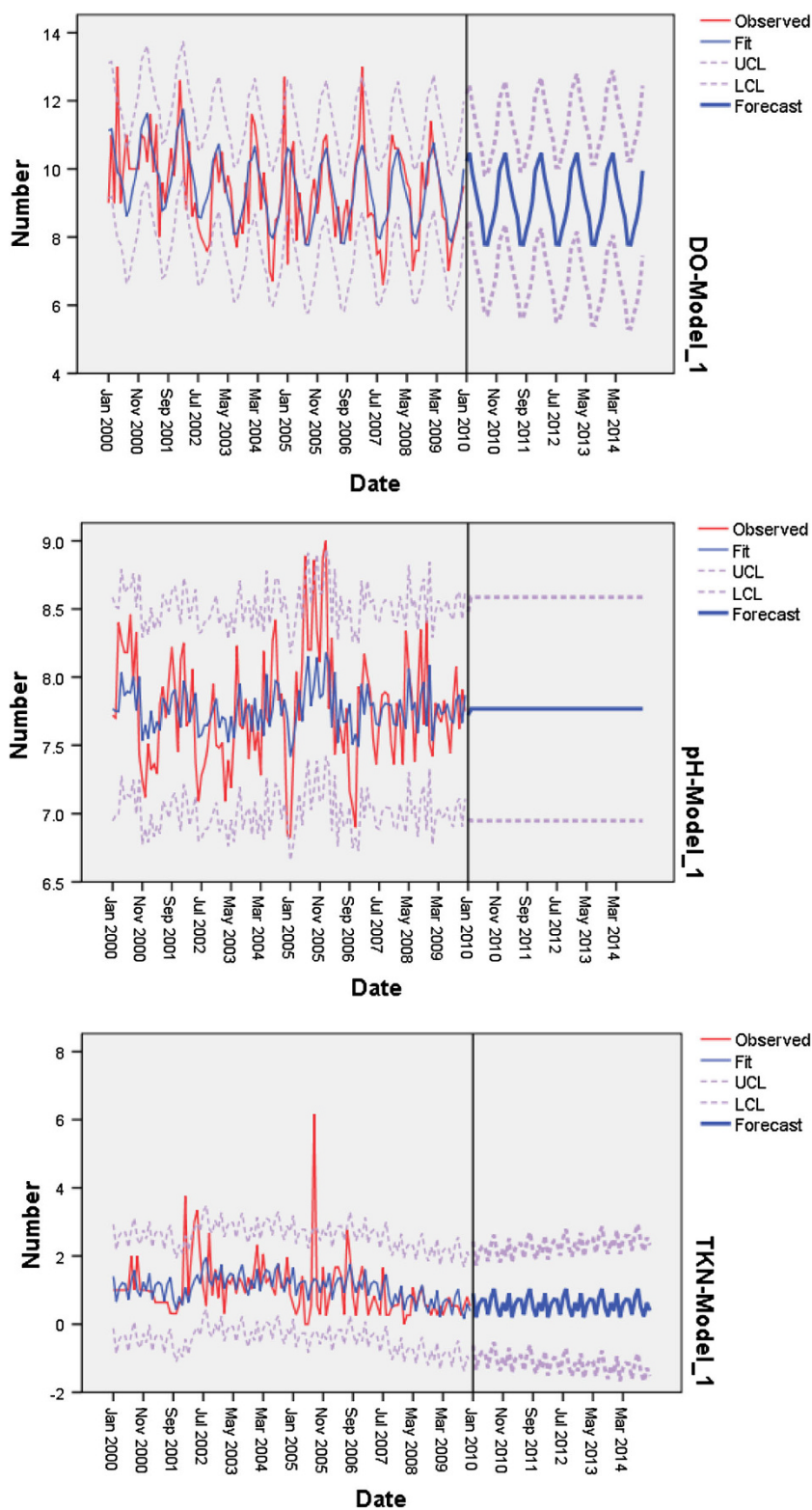
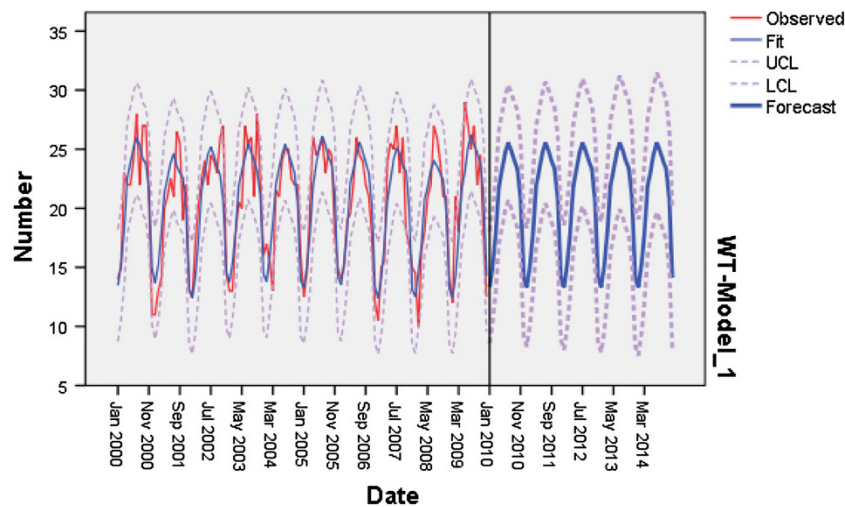


Fig. 3 continued

AMM

Average, median and mode value is approximately equal thus data exhibit normal behavior. Standard deviation

(0.2879) suggests that data are close to each other. Curve is platykurtic. Stationary R-squared and R-squared values exhibit the similar behavior thus model is better than the baseline model and RMSE values is low so dependent



**Fig. 3** continued

series is closed with its model-predicted level. Using Ljung–Box  $Q(18)$  model statistics is 26.689, significance is 0.045 and degree of freedom is 16. ARIMA simple model fitted and boundary lines at 95 % confidence limits. Predicted, LCL, UCL and residual values of AMM are 0.1832,  $-0.3815$ , 0.7480 and 0.00038.

#### TKN

Mean and median value is approximately equal. Standard deviation (0.83) suggests that sample data are close to each other. Skewness value is approximately 0 thus curve is symmetrical and platykurtic. Stationary R-squared and R-squared values exhibit the similar behavior thus model is better than the baseline model and RMSE values are low so dependent series is closed with its model-predicted level. Using Ljung–Box  $Q(18)$  model statistics is 16.311, significance is 0.431 and degree of freedom is 16. ARIMA (1, 0, 0) model fitted and boundary lines at 95 % confidence limits. Predicted, LCL, UCL and residual values of TKN are 0.9066,  $-0.6859$ , 2.4991, and  $-0.0414$ .

#### DO

Average, median and mode value is equal thus data behave normally. Standard deviation value is 1.344 curve is symmetric and platykurtic. Stationary R-squared and R-squared values exhibit the similar behavior thus model is better than the baseline model and RMSE values are low so dependent series is closed with its model-predicted level. Using Ljung–Box  $Q(18)$  model statistics is 16.161, significance is 0.442 and degree of freedom is 16. ARIMA (0, 0, 6) model fitted and boundary lines at 95 % confidence limits. Predicted, LCL, UCL and residual values of DO are 9.3279, 7.2570, 11.3988, and  $-0.0756$ .

#### WT

Mean, median and mode value is approximate to 21 thus data exhibit normal characteristic. Curve is skewed and platykurtic. Stationary R-squared and R-squared values exhibit the similar behavior thus model is better than the baseline model and RMSE values are low so dependent series is closed with its model-predicted level. Using Ljung–Box  $Q(18)$  model statistics is 22.148, significance is 0.138 and degree of freedom is 16. ARIMA (1, 0, 13) model fitted and boundary lines at 95 % confidence limits. Predicted, LCL, UCL and residual values of WT are 20.5947, 15.6442, 25.5453, and  $-0.0136$ .

#### Conclusion

Statistical and time-series analysis of water quality parameters monitored at the Hathnikund of Yamuna River in India has been studied. ARIMA model used for the prediction of the monthly values of water quality parameters for next 5 years. It is observed that curve is platykurtic for pH, AMM, TKN, DO, WT; leptokurtic for COD, BOD and normal for pH, AMM, DO, WT.

For all ARIMA model ( $p, d, q$ ) value of ‘ $d$ ’ i.e., middle value is zero thus process is stationary and has constant mean and variance. It is also observed that RMSE value is comparatively very low which shows that dependent series is closed with the model-predicted level, thus predictive model is useful at 95 % confidence limits. MAPE, MaxAPE, MAE, MaxAE, normalized BIC are calculated for all parameters and it is observed that all water quality parameters have low value. It concludes that the predicted series is close to the original series thus it is a perfect fit. Five year next predicted, LCL–UCL mean values using time series are given as for pH



7.7672, 6.9887–8.5458 with  $-0.00017$  residual error; for COD 8.1648,  $-2.1331$ – $18.4629$  with  $0.3341$  residual error; for BOD 1.2477,  $0.1189$ – $2.3765$  with  $0.0163$  residual error; for AMM 0.1832,  $-0.3815$ – $0.7480$  with  $0.00038$  residual error; for TKN 0.9066,  $-0.6859$ – $2.4991$  with  $-0.0414$  residual error; for DO 9.3279,  $7.2570$ – $11.3988$  with  $-0.0756$  residual error and for WT 20.5947,  $15.6442$ – $25.5453$  with  $-0.0136$  residual error, respectively.

Therefore, using time series and statistical analysis, it is concluded that all parameters except pH and WT cross the prescribed limits of WHO/EPA and water is not fit for drinking, agriculture and industrial use. River is a natural resource of water, prediction results of ARIMA model indicate the increase in pollution, which is an alarming situation and the preventive measure has to be taken to control the same.

**Acknowledgments** Authors are thankful to University Grant Commission (UGC), Government of India for financial support [F. 41-803/2012 (SR)]; Central Pollution Control Board (CPCB), Government of India for providing the research data; Guru Gobind Singh Indraprastha University, Delhi (India) for providing research facilities. First author is thankful to Sant Baba Bhag Singh Institute of Engineering and Technology for providing study leave to pursue research degree.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Akoto O, Adiyiah J (2007) Chemical analysis of drinking water from some communities in the Brong Ahafo region. *Int J Environ Sci Tech* 4(2):211–214
- Alam Md JB, Muyen Z, Islam MR, Islam S, Mamun M (2007) Water quality parameters along rivers. *Int J Environ Sci Tech* 4(1):159–167
- APHA (1995) Standard methods for examination of water and waste water. 19th edn, American Public Health Association, Washington D.C
- Attah DA, Bankole GM (2012) Time series analysis model for annual rainfall data in Lower Kaduna Catchment Kaduna, Nigeria. *Int J Res Chem Environ* 2(1):82–87
- Banu JR, Kaliappan S, Yeom IT (2007) Treatment of domestic wastewater using upflow anaerobic sludge blanket reactor. *Int J Environ Sci Tech* 4(3):363–370
- Bhardwaj R, Parmar KS (2013a) Water quality index and fractal dimension analysis of water parameters. *Int J Environ Sci Tech* 10(1):151–164
- Bhardwaj R, Parmar KS (2013b) Wavelet and statistical analysis of river water quality parameters. *App Math Comput* 219(20):10172–10182
- Box GEP, Jenkins GM, Reinsel GC (2008) Time series analysis: forecasting and control, 4th edn. John Wiley and Sons, Inc, UK
- Chenimi I, Khemiri S (2009) Evaluation of ground water quality using multiple linear regression and structural equation modeling. *Int J Environ Sci Tech* 6(3):509–519
- CPCB (2006) Water Quality Status of Yamuna River (1999–2005): Central Pollution Control Board, Ministry of Environment & Forests, Assessment and Development of River Basin Series: ADSORBS/41/2006-07
- Damodhar U, Reddy MV (2013) Impact of pharmaceutical industry treated effluents on the water quality of river Uppanar, South east coast of India: a case study. *Appl Water Sci* 3:501–514
- DeLurgio SA (1998) Forecasting principles and applications, 1st edn. Irwin McGraw-Hill Publishers, New York
- Fang H, Wang X, Lou L, Zhou Z, Wu J (2010) Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. *Water Res* 44(5):1562–1572
- Imo TS, Oomori T, Toshihiko M, Tamaki F (2007) The comparative study of trihalomethanes in drinking waters. *Int J Environ Sci Tech* 4(4):421–426
- Joarder MAM, Raihan F, Alam JB, Hasanuzzaman S (2008) Regression analysis of ground water quality data of Sunamganj District, Bangladesh. *Int J Environ Res* 2(3):291–296
- Juang DF, Tsai WP, Liu WK, Lin JH (2008) Treatment of polluted river water by a gravel contact oxidation system constructed under riverbed. *Int J Environ Sci Tech* 5(3):305–314
- Kahya E, Kalayci S (2004) Trend analysis of streamflow in Turkey. *J Hydrol* 289:128–144
- Korashey R (2009) Using regression analysis to estimate water quality constituents in Bahr El Baqar drain. *J Appl Sc Res* 5(8):1067–1076
- Kumar A, Dua A (2009) Water quality index for assessment of water quality of river Ravi at Modhopur (India). *G J Env Sci* 8(1):49–57
- Lu WX, Zhao Y, Chu HB, Yang LL (2013) The analysis of groundwater levels influenced by dual factors in western Jilin Province by using time series analysis method. *Appl Water Sci*. doi:10.1007/s13201-013-0111-4
- McCleary R, Hay RA (1980) Applied time series analysis for the social sciences. Sage, Beverly Hills
- Mousavi M, Kiani S, Lotfi S, Naeemi N, Honarmand M (2008) Transient and spatial modeling and simulation of polybrominated diphenyl ethers reaction and transport in air, water and soil. *Int J Environ Sci Tech* 5(3):323–330
- Movahed M, Hermanisc E (2008) Fractal analysis of river flow fluctuations. *Phys A* 387(4):915–932
- Park J, Park C (2009) Robust estimation of the Hurst parameter and selection of an onset scaling. *Stat Sinica* 19(4):1531–1555
- Parmar KS, Chugh P, Minhas P, Sahota HS (2009) Alarming pollution levels in rivers of Punjab. *Indian J Env Prot* 29(11):953–959
- Phiri O, Mumba P, Moyo BHZ, Kadewa W (2005) Assessment of the impact of industrial effluents on water quality of receiving rivers in urban areas of Malawi. *Int J Environ Sci Tech* 2(3):237–244
- Prasad BG, Narayana TS (2004) Subsurface water quality of different sampling stations with some selected parameters at Machilipatnam Town. *Nat Env Poll Tech* 3(1):47–50
- Prasad B, Kumari P, Bano S, Kumari S (2013) Ground water quality evaluation near mining area and development of heavy metal pollution index. *Appl Water Sci*. doi:10.1007/s13201-013-0126-x
- Psargaonkar A, Gupta A, Devotta S (2008) Multivariate analysis of ground water resources in Ganga–Yamuna Basin (India). *J Env Sc Eng* 50(3):215–222
- Rangarajan G (1997) A climate predictability index and its applications. *Geophys Res Lett* 24(10):1239–1242
- Rangarajan G, Ding M (2000) Integrated approach to the assessment of long range correlation in time series data. *Phys Rev E* 61(5):4991–5001
- Rangarajan G, Sant DA (2004) Fractal dimensional analysis of Indian climatic dynamics. *Chaos Solitons Fractals* 19(2):285–291

- Ravikumar P, Mehmood MA, Somashekar RK (2013) Water quality index to determine the surface water quality of Sankey tank and Mallathahalli lake, Bangalore urban district, Karnataka, India. *Appl Water Sci* 3:247–261
- Seth R, Singh P, Mohan M, Singh R, Aswal RS (2013) Monitoring of phenolic compounds and surfactants in water of Ganga Canal, Haridwar (India). *Appl Water Sci*. doi:[10.1007/s13201-013-0116-z](https://doi.org/10.1007/s13201-013-0116-z)
- Shukla JB, Misra AK, Chandra P (2008) Mathematical modeling and analysis of the depletion of dissolved oxygen in eutrophied water bodies affected by organic pollutant. *Nonlinear Anal Real world Appl* 9(5):1851–1865
- Singh KP, Malik A, Mohan D, Sinha S (2004) Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Res* 38(18):3980–3992
- Su S, Li D, Zhang Q, Xiao R, Huang F, Wu J (2011) Temporal trend and source apportionment of water pollution in different functional zones of Qiantang River, China. *Water Res* 45(4):1781–1795
- Vassilis Z, Antonopoulos M, Mitsiou AK (2001) Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. *Hydrol Earth Syst Sc* 5(4):679–691
- WHO (1971) International standards for drinking water. World Health Organization, Geneva