Spatiotemporal Predictions of BOD Levels in River Ganga: Use of NARX Model

Sunayana
NEERI: National Environmental Engineering Research Institute CSIR
Suresh Kumar Gurjar (sgurjar@iitk.ac.in)
IIT Kanpur: Indian Institute of Technology Kanpur https://orcid.org/0000-0002-6212-2543
Vikas Kumar
Central University of Haryana
Vinod Tare
IIT Kanpur: Indian Institute of Technology Kanpur

Research Article

Keywords: BOD, Ganga, NARX, Time Series, Spatiotemporal and Consecutive & Non-consecutive

Posted Date: February 24th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-163110/v1

Word Count: 4296

Spatiotemporal Predictions of BOD levels in River Ganga: use of NARX Model

Sunayana[1]

ORCID no: 0000-0002-9888-6860

Suresh Kumar Gurjar[2,3*]
ORCID no: 0000-0002-6212-2543

Vikas Kumar[4]

Vinod Tare[2,3]

[1]CSIR-NEERI, Delhi Zonal Centre, A-93-94, Naraina Industrial Area Phase I, Naraina Industrial Area, New Delhi, India-110028

[2]Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur, India-208016

[3]Centre for Ganga River Basin Management and Studies (cGanga), Indian Institute of Technology Kanpur, Kanpur, India-208016

[4]Department of Civil Engineering, Central University Haryana, Mahendargarh, Haryana, India - 123031

*corresponding author

Email: sgurjar@iitk.ac.in

Ph: +91-512-2597797

Abstract

Rivers finding their path through different hamlets, industries and residential etc. are the end receivers of liquid and solid waste. This plight of water bodies has been deteriorating the life of the river stretches and greatly impacting the quality of water. The identification of potentially polluted locations and timely action can help in restoring the water quality of the river stretches and improving the life of water body as well. The dynamics of river Ganga stretches has been modeled using Non-linear

Autoregressive network with exogenous inputs (NARX) model spatially and temporally simultaneously. This method helps in developing one single model to understand all the stretches simultaneously which is not possible otherwise. The change in behavior of NARX model under consecutive and non-consecutive time period has also been studied using statistical methods and multi-step ahead predictions. The model developed in this study captures all the seasons together in one single model and has been used to predict BOD values at seven different locations for nine months ahead. The results highlight the performance ability of NARX model to understand future water quality changes of river Ganga stretch owing to continuous pollution adding to the river. The results predicted for 9 months ahead for year 2015 using two different models, a lower root mean square error (RMSE) as 0.026 and higher correlation coefficient 'r' as 0.992 was obtained for model with consecutive days while for other model with non-consecutive days the RMSE was 0.031 and r was 0.989. This information may provide some guidance to the policy makers and water managers to prepare and suggest the pollution mitigation measures for the life line of millions; River Ganga.

Keywords

BOD, Ganga, NARX, Time Series, Spatiotemporal and Consecutive & Non-consecutive

Declarations

Ethics approval and consent to participate: Not Applicable

Consent for publication: Not Applicable

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request

Competing Interests: The authors declare that they have no competing interests

Funding: The authors acknowledge the computational resource and technical help of NEERI, Government of India and cGanga, IIT Kanpur to carry out the research.

Word Count: 4296

Authors contribution

SUN and SKG conceived the research and performed the model analysis. SUN and SKG wrote the manuscript, and VT and VK edited the manuscript.

1 Introduction

Rivers in any civilization have been the lifeline of millions of people who live along the course and depend upon its water for domestic and indusatrial activities, irrigation and other purposes (Gurjar and Tare 2019). With exponential population growth and rapid development, rivers are highly stressed and overused resulting in threat to safe water availability. Unsustainable development practices have made the rivers an end point of discharge for all kinds of waste coming from different sources (NMCG 2020) and hence threatening the life of rivers and dependent ecosystems over time. The discharge of high amount of unchecked waste impedes the water quality and also disturbs the species existing in the pyramid.

In India, sacred river Ganga has also witnessed the deterioration of its water quality over the years. Cities like Kannauj, Kanpur, Prayagraj and Varanasi etc. are situated on its banks, thus, the waste discharges from these cities/towns in river need to be properly inventoried, monitored and suggested of solutions. Considering such plight of river, strategically located monitoring locations in different stretches are needed to understand the changes in water quality over time, identify any threat to aquatic life and possible impacts on human habitation as well (Bhangu and Whitfield 1997; Chang 2008; Chang and Lin 2014; Varekar et al. 2015; Pérez et al. 2017; Ahmadi et al. 2018). In GAP (Ganga Action Plan) phase 1 and 2, water samples from some identified locations were collected and analyzed for a long period to understand the spatiotemporal changes in water quality of river Ganga (Dwivedi et al. 2018). Biochemical Oxygen Demand (BOD) is considered an important parameter by many scientific organizations to measure the water quality of surface water sources (Higashino and Stefan 2017). Central Pollution Control Board (CPCB) India, has defined five classes of surface water based on values of certain water quality parameters including BOD and DO (CPCB 2008). Extreme BOD loads in downstream stretch of river indicate discharge of untreated wastewater from towns and cities. These discharges with very high BOD loads result in low Dissolved Oxygen (DO) values thereby affecting flora and fauna in river stretch (APHA 1995). The other physicochemical parameters like temperature, DO and water flow etc. bears a critical affect over BOD (Dogan et al. 2009; Ferreira et al. 2020). Owing to amount of efforts required in estimating and predicting BOD in water (Singh et al. 2009), there is a need of advanced computing methods to predict BOD concentrations spatially and temporally so that remediating step can be taken in advance for required interventions.

Traditionally before the evolution of soft computing power, deterministic models like QUAL2E, QUAL2K (Park and Lee 2002), SWAT (Jha et al. 2007), integration of QUAL2K and SWAT (Bui et al. 2019), hydrodynamic and transport model (Zhu et al. 2011), empirical or mechanistic model (Reckhow 1994) were used to estimate desired parameters so as to know the water quality. The limitation with such method is inability to capture complex non-linear dynamics existing in the system (Xiang et al. 2006). As a whole, river system in itself is a very complex system and to capture this complexity, methods like neural networks can be adopted for predicting the values spatially and temporally for better understanding of the system (Maier and Dandy 1996). Artificial neural networks are basically designed to mimic the functioning as of human brain to perform tasks in water related fields like water resources, parameter-based water quality prediction, groundwater level prediction and marine water assessment etc. The extraordinary power of this tool has been used in modeling COD, DO (Ranković et al. 2010; Ahmed 2017) and BOD for various river stretches and reservoir water quality across the world (Singh et al. 2009; Alizadeh and Kavianpour 2015; Qaderi and Babanezhad 2017; Jouanneau et al. 2019; Rajaee et al. 2020; Tiyasha et al. 2020). Urban water quality modeling using multi-task multi view learning has been done using learning methods overcoming the shortcomings of existing hydraulic models (Liu et al. 2016) Water quality forecast for coastal waters have also been done using ANN in Singapore (Palani et al. 2008). From the review of available literature on the subject, it can be concluded that AI based models give good accuracy of results in water quality predictions (Kisi and Parmar 2016; Li et al. 2018). Therefore, NARX model has been used in the present study for river water quality modeling to understand spatiotemporal variations.

River Ganga due to its location and nearby settlements witness huge pollution and the river seems to be losing its serenity. BOD is such parameter that it can indicate the quantum of organic load in river during different time periods. The study uses BOD concentrations taken at different identified stations in Kanpur and Kannauj stretch of river Ganga for developing a non-linear autoregressive network with exogenous input (NARX) model to predict BOD for months ahead at different stations. Further, the performance of NARX models under non-consecutive months and consecutive months has been compared. The prediction of BOD at different stations and for different months is a way to capture spatiotemporal dynamics together using AI based model. Therefore, these efficient models can be used for managing and mitigating the pollution level in river stretches.

2 Study area

The Ganga, with a large basin area of about 8,61,452 km^2 area spreads in 11 states of India, supports a large part of population in India for food, water and livelihood (Tare et al. 2017). Rapidly increasing Indian population, Industrialization and Urbanization, changing agricultural practices and damming of river with effects of land use land cover change in basin have impacted the quality and quantity of water in the river. Contentious efforts of measuring water quality in GAP1 and GAP2 drained a lot of money and time without noticing any significant effect on quality of water in the river Ganga. The Environmental Information Systems (ENVIS), New Delhi, estimated a generation of 15, 235 MLD sewage from the urban population of 5 states of Indo-Gangetic Plain (IGP) (Dwivedi et al. 2018). A major portion (about 77%) of this generated sewage goes into river in untreated condition and create high impact on water quality. This untreated sewage discharge has increase about 9-fold in contrast to the discharge prior to the GAP 1 period (Dwivedi et al. 2018). With various point and non-point pollution sources impacting water quality, the most polluted stretch of river Ganga, Kannauj to Kanpur, was selected for present study. Only those stations, having continuous data for model calibration and validation, and covering major part of the most polluted stretch in Ganga, were selected for present study. This helped the model to cover the entire study area, both specially and temporally. The stretch of Ganga in the study covers approximately 48 km; covering trajectory of 18 km in the areas of Kannauj (P1-P3) and approximately 30 km in Kanpur (P4-P7). Fig. 1 and Table 1 give a brief detail of the location of selected stretch along with sampling points.

Figure 1 and Table 1 here….

3 Data collection, cleaning and analysis

The data for the study was received from Centre for Ganga River Basin Management and Studies (cGanga), IIT Kanpur. IIT Kanpur collected and analyzed the water samples collected from various locations on main stem of river Ganga on monthly basis with support received from National River Conservation Directorate (NRCD), GoI. Data for a period from January 2001 to September 2015 was received for all the sampling stations mentioned in Fig. 1.

The monthly variation of BOD values at selected stations shown in Fig. 2 indicates that BOD series is not smooth during year 2001 to 2014. From figure, it is also clear that any type of trend, seasonality or cyclic behavior is not visible and hence difficult to predict BOD values for upcoming months or years based on simple statistical methods (Hyndman and Athanasopoulos 2018).

Figure 2 Here…

3.1 Missing Values

As the NRCD supported monitoring programme was long and there were instances where sampling couldn't be done due to adverse weather conditions and various other reasons at different stations. The missing values were interpolated using multiple regression for different time period and then were used in neural modeling. The locations at which values were missing has been shown in Table 2.

Table 2 here…

Keeping various unknown factors in consideration for missing data an attempt was made to understand how missing data affects the prediction behavior of neural model. Using this data gap as opportunity, the prediction capacity of the NARX model for

BOD (Biochemical Oxygen Demand) was compared for consecutive and non-consecutive days at different locations.

3.2 Methodology

The discharge from various domestic and industrial clusters in the studied segment of the river makes it most polluted and complex trajectory of middle Ganga. The different type of sources and effluent quality in this stretch of river makes it extremely difficult to model the spatial and temporal dynamics simultaneously. Availability of continuous data for time and space dimensions at all selected stations, breakdown during monitoring process or sudden discharge of high pollutant load in the river stretch are additional challenges for any model to predict the outputs (Gazzaz et al. 2012).

3.3 Non-Linear Autoregressive models with Exogenous Input (NARX)

The NARX neural model was used for time series prediction in many fields (Jiang and Song 2011; Chang et al. 2013). NARX models have been proved to be performing better than feed forward back propagated neural models in modeling time series (Chang and Lin 2014; Chang et al. 2016). NARX models are able to capture the non-linear complex dynamics of the system or process. These models are constructed by incorporating two tapped delay passes, one is at the input itself while other is found to pass the predicted output. This tool is powerful in understanding the behavior of past outputs of the system which can affect the outcome at next step.

NARX model has one input, hidden and output layer and makes recurrent connections from the outputs, to delay time steps and again forming new set of inputs. For N-step-ahead the forecasting can be expressed as shown in eqn. (1) where $N \geq 1$.

$$y(t+N) = f[y(t+N-1) \ldots y(t+N-q); X(t)] \qquad \text{Eq. 1}$$

where, $X(t)$and $y(t)$ are the input vectors at the same time step (t). f(.) is the non-linear function and q is output-memory order or feedback delay. The variable $y(t+N-i)$ { i =1 to q } which is output behaves like an autoregressive model while $X(t)$ is exogenous variable in time series prediction. These networks are basically trained by two ways: parallel (P) and series-parallel (SP) mode. In SP mode, the output's regressor at the input layer contains only the target (measured or actual) values of the system to be modeled $z(t)$.

$$y(t+N) = f[z(t+N-1), \ldots., z(t+N-p); X(t)] \qquad \text{Eq. 2}$$

The other mode i.e. in P mode, the estimated or predicted outputs are feedback into the output's regressor at the input layer, as represented in Eq. 1. Therefore, when prediction is conducted at $N \geq 2$, the p antecedent measured values $z(t+N-1)$, $z(t+N-2)$,… $z(t+N-p)$) are future values that are not available at current time. Hence, the NARX model are first trained in SP mode and training accompanied by validation and testing is done to obtain best mode. This model is then used further in P mode to predict values for future times. In this P mode, the open loop in SP mode is closed and same NARX network is used in P mode. The architecture of NARX model is shown in Fig. 3.

Figure 3 here..

In this study, the $X(t)$ is 5 (pH, Temperature, DO, COD and Fecal Coliform) and $y(t)$ is 1 (BOD).

3.4 Data for NARX model

The data used for neural model comprises of monitoring data from all monitoring stations (P1-P7) collected during the sampling programme (Jan, 2001 to Sep, 2015). The statistical parameter of data used for NARX model is given in Table 3. All the water quality parameters were analyzed as per standard procedure of 'Standard Methods for Examination of Water and Wastewater (APHA)'.

Table 3 here…

The different data sets (i.e. matrix size) used for training and prediction for BOD at all stations (P1-P7) is shown in Table 4.

Table 4 here…

3.5 Deciding Input and Feedback delays

For neural networks to deal with temporal dimension, they are supplied with memory i.e. by introducing time delays on connections. Storing inputs for long enough so as to influence the subsequent inputs or storing past inputs assuming to affect the future value falls into the category of time-delay neural network (Dorffner 1996). In NARX model, autocorrelation of the target and the cross-correlation of input and target helps in finding the lags that are significant for input and feedback delays (Matkovskyy and Bouraoui 2019).

3.6 Optimized Parameters for model

The performance of Artificial Intelligence (AI) based models is highly dependent on type of data and the pattern used for training, testing and validation. Therefore, these models are sensitive to input data and hence results may vary with change in pattern of data used the other time (Noori et al. 2015). Therefore, static neural models need to be tested for uncertainty because the selection of data points is random and a pattern change is observed with every run (Aqil et al. 2007; Noori et al. 2010, 2013). For dynamic time series models, the chronology of data is important for forecasting the parameters under study. Therefore, the selection of data for training, testing and validation was divided on chronological basis and the neural model was constructed with and without consecutive days to include pattern change because of inclusion or deletion of some data. By selecting these data pattern, models were optimized and values of certain parameters were obtained and other parameters were kept as default as in LM algorithm (Table 5).

Table 5 here…

4 Results and discussion

In this study, water quality analysis done at the stations (P1-P7) has been used for prediction of BOD for year 2015 (January to September).

4.1 Neural Architecture

NARX network was used to predict BOD concentrations at different locations and at different time. The neural architecture is incomplete without optimization of neurons in hidden layer. Single hidden layer was used in NARX and number of neurons was optimized by calculating the mean square error (MSE) for different neurons during training and testing. The neurons were varied from 5 to 12 as shown in Fig. 4. It can be seen that MSE during testing phase was least when number of neurons were kept 8 in hidden layer. Therefore, the optimal structure obtained for NARX model was 5-8-1 with proper input and feedback delays.

Figure 4 here…

Apart from fixing the neurons in hidden layer for NARX model, the significant lags also need to be identified and incorporated in model. To fix, significant delays after doing auto and cross-correlation, the input delay (ID) was set to zero and feedback delay (FD) was taken as 1. The NARX model after setting the neurons and delays was trained using Levenberg-Marquardt (LM) algorithm and therefore, the optimized NARX model is 5-8-1 with ID =0 and FD =1.

4.2 Model Performance

The optimized neural model with 5-8-1 architecture with significant lag at input as zero and at output as one was trained using LM algorithm under consecutive and non-consecutive time series. The performance of NARX model was assessed using root mean square error (RMSE) and Correlation coefficient (r) under both time series using Eq. 3 and 4 respectively. The performance of NARX model for both the time series is presented in Table 6.

$$RMSE = \sqrt{\left[\sum_{i=1}^{N}(Measured - Predicted)^2 / N\right]} \quad \text{Eq. 3}$$

$$r = \frac{\sum_{i=1}^{N}(Predicted - Predicted_{avg.})(Measured - Measured_{avg.})}{\sqrt{\sum_{i=1}^{N}(Predicted - Predicted_{avg.})^2}\sqrt{\sum_{i=1}^{N}(Measured - Measured_{avg.})^2}} \quad \text{Eq. 4}$$

Table 6 here….

Fig. 5 a, b and c, d, respectively, shows the performance of NARX model for station P1-P3 and P4-P7 during testing phase for consecutive and non-consecutive time series.

Figure 5 here

NARX model has performed well for prediction during testing phase for all the locations despite frequent monthly variations from extreme high to extreme low values. This performance during testing phase indicates that NARX model are able to predict even time series which does not have pattern or trend. The performance of the model shows that time series which are not stationary and need multiple transformations to derive stationarity and future prediction, neural model can be used efficiently in those cases without any transformations.

4.3 Spatiotemporal estimation of BOD for months ahead

The NARX model used in this study, estimate the spatiotemporal BOD concentration in Ganga River. The 5 parameters i.e. pH, temp, DO, COD and Fecal Coliform were used as exogenous inputs to the network. The BOD was predicted for 9 months ahead for next year (i.e. 2015) at all the stations. The prediction at all stations using consecutive and non-consecutive time series through neural model was done and discussed in following paragraphs.

Station (P1-P3)

The BOD prediction at stations P1 to P3 in areas of Kannauj were compared for its predictive power for 9 months (Jan-Sep.) i.e. t=t+9 ahead using consecutive and non-consecutive time series (monthly) for year 2015. In both the time series, the instances of missing values were only present in the training phase while validation and testing phase has continuous values. This therefore helps in understanding that whether the missing values during training phase affects the prediction power of model for time series modeling.

NARX model was used able to predict BOD concentrations at all the stations (Fig. 6) in Kannauj area. At all stations, the results obtained after using consecutive time series in training phase yielded higher ‘r’ value of 0.8984 in contrast to ‘r’ value of 0.8405 for NARX model trained using non-consecutive periods.

Fig 6 here

Station (P4-P7)

The stretch of river which has monitoring stations P4-P7 lies in Kanpur has acute discharges at several instances and hence significantly affect the water quality. The BOD concentrations predicted for year 2015 are shown in Fig. 7.

Fig 7 here.

From Fig. 7 it was observed that in this stretch when there were many missing consecutive months for training of model the predictive results were not good. On the other hand, when NARX model was trained with consecutive time series the predicted results were closer to measured BOD values.

5 Conclusions

In this study, NARX model were used for prediction of BOD in river stretch of Ganga. The stretch from Kannauj to Kanpur was selected for predicting BOD values for upcoming 9 months i.e. Jan to Sep, 2015. The developed NARX model was able to predict closer values of BOD for 9 months ahead in the same stretch and it was concluded that NARX models can be used for modeling the spatiotemporal dynamics of the river. The spatial characteristics were adjusted as per the locations of stations from upstream to downstream and the model was trained based by sequentially aligning locations for the model. Since the river stretch selected was complex in terms of discharges coming from various points and subsequent peaks in BOD values, the future prediction for water quality becomes extremely difficult. This study therefore highlighted that NARX model has the capacity to simulate river water quality spatially and temporally and can be extended to long stretches if good quality data is available. These models showed that with essential parameter measured in river stream over a period of time, can be used for predicting future quality of river stretches. This was also concluded that since these models are data driven models consecutive data for time period is required for training the model so that it has better predictive capability. The study compared two models, one with consecutive period (data gaps filled by interpolation) and other with non-consecutive periods, and it was found that NARX model requires consecutive time period in training phase for better results. From this study, it was found that NARX model can be used for modeling any river stretch and needs to be updated with data for keeping an updated model. This model showed that the NARX model were able to capture the complex non-linear dynamics of the river stretch. The model gave results for different stations in the stretch and thus an information for water quality in that particular stretch was derived. The model is also robust in the way that it can added of some more parameters which are important so that extreme fluctuations can also be taken care of in the model. Therefore, these kinds of model can be constructed for different stretches and combined streams for predicting the water quality of the river.

Word Count: 4296

References

Ahmadi A, Fatemi Z, Nazari S (2018) Assessment of input data selection methods for BOD simulation using data-driven models: a case study. Environ Monit Assess 190:1–17. doi: 10.1007/s10661-018-6608-4

Ahmed AAM (2017) Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs). J King Saud Univ - Eng Sci 29:151–158. doi: 10.1016/j.jksues.2014.05.001

Alizadeh MJ, Kavianpour MR (2015) Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. Mar Pollut Bull 98:171–178. doi: 10.1016/j.marpolbul.2015.06.052

APHA (1995) Standard Mrthods for Examination of Water and Waste Water, 19 th. Americam Public Health Association, Washington, DC

Aqil M, Kita I, Yano A, Nishiyama S (2007) Analysis and prediction of flow from local source in a river basin using a Neuro-fuzzy modeling tool. J Environ Manage 85:215–223. doi: 10.1016/j.jenvman.2006.09.009

Bhangu I, Whitfield PH (1997) Seasonal and long-term variations in water quality of the Skeena River at Usk, British Columbia. Water Res 31:2187–2194. doi: 10.1016/S0043-1354(97)00063-8

Bui HH, Ha NH, Nguyen TND, et al (2019) Integration of SWAT and QUAL2K for water quality modeling in a data scarce basin of Cau River basin in Vietnam. Ecohydrol Hydrobiol 19:210–223. doi: 10.1016/j.ecohyd.2019.03.005

Chang CL, Lin YT (2014) A water quality monitoring network design using fuzzy theory and multiple criteria analysis. Environ Monit Assess 186:6459–6469. doi: 10.1007/s10661-014-3867-6

Chang FJ, Chen PA, Chang LC, Tsai YH (2016) Estimating spatio-temporal dynamics of stream total phosphate concentration by soft computing techniques. Sci Total Environ 562:228–236. doi: 10.1016/j.scitotenv.2016.03.219

Chang FJ, Chen PA, Liu CW, et al (2013) Regional estimation of groundwater arsenic concentrations through systematical dynamic-neural modeling. J Hydrol 499:265–274. doi: 10.1016/j.jhydrol.2013.07.008

Chang H (2008) Spatial analysis of water quality trends in the Han River basin, South Korea. Water Res 42:3285–3304. doi: 10.1016/j.watres.2008.04.006

CPCB (2008) Guidlines for water quality management. Delhi, India, (http://cpcb.nic.in/, Accessed on April 06, 2019)

Dogan E, Sengorur B, Koklu R (2009) Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J Environ Manage 90:1229–1235. doi: 10.1016/j.jenvman.2008.06.004

Dorffner G (1996) Neural networks for time series processing. Neural Netw World 6:447–468

Dwivedi S, Mishra S, Tripathi RD (2018) Ganga water pollution: A potential health threat to inhabitants of Ganga basin. Environ. Int. 117:327–338

Ferreira DM, Fernandes CVS, Kaviski E, Fontane D (2020) Transformation rates of pollutants in rivers for water quality modelling under unsteady state: A calibration method. J Hydrol 585:124769. doi: 10.1016/j.jhydrol.2020.124769

Gazzaz NM, Yusoff MK, Aris AZ, et al (2012) Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar Pollut Bull 64:2409–2420. doi: 10.1016/j.marpolbul.2012.08.005

Gurjar SK, Tare V (2019) Spatial-temporal assessment of water quality and assimilative capacity of river Ramganga, a tributary of Ganga using multivariate analysis and QUEL2K. J Clean Prod 222:550–564. doi: 10.1016/j.jclepro.2019.03.064

Higashino M, Stefan HG (2017) Oxygen uptake prediction in rivers and streams: A stochastic approach. J Environ Manage 203:200–207. doi: 10.1016/j.jenvman.2017.07.059

Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. In: 2nd Ed. OTexts Melbourne, Aust. https://otexts.com/fpp2/. Accessed 20 Oct 2020

Jha MK, Gassman PW, Arnold JG (2007) Water quality modeling for the Raccoon River watershed using SWAT

Jiang C, Song F (2011) Sunspot forecasting by using chaotic timeseries analysis and NARX network. J Comput 6:1424–1429. doi: 10.4304/jcp.6.7.1424-1429

Jouanneau S, Grangé E, Durand MJ, Thouand G (2019) Rapid BOD assessment with a microbial array coupled to a neural machine learning system. Water Res 166:115079. doi: 10.1016/j.watres.2019.115079

Kisi O, Parmar KS (2016) Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. J Hydrol 534:104–112. doi: 10.1016/j.jhydrol.2015.12.014

Li Z, Peng F, Niu B, et al (2018) Water Quality Prediction Model Combining Sparse Auto-encoder and LSTM Network. IFAC-PapersOnLine 51:831–836. doi: 10.1016/j.ifacol.2018.08.091

Liu Y, Yu Z, Yuxuan L, et al (2016) Urban water quality prediction based on multi-task multi-view learning

Maier HR, Dandy GC (1996) The Use of Artificial Neural Networks for the Prediction of Water Quality Parameters. Water Resour Res 32:1013–1022. doi: 10.1029/96WR03529

Matkovskyy R, Bouraoui T (2019) Application of Neural Networks to Short Time Series Composite Indexes: Evidence from the Nonlinear Autoregressive with Exogenous Inputs (NARX) Model. J Quant Econ 17:433–446. doi:

10.1007/s40953-018-0133-8

NMCG (2020) Pollution Threat. In: Natl. Misssion Clean Ganga. https://nmcg.nic.in/pollution.aspx. Accessed 7 Apr 2020

Noori R, Ashrafi K, Karbassi A, et al (2013) Development and application of reduced-order neural network model based on proper orthogonal decomposition for BOD_5 monitoring in river systems: Uncertainty analysis. Environ Prog Sustain Energy 32:344–349. doi: 10.1002/ep.11610

Noori R, Hoshyaripour G, Ashrafi K, Araabi BN (2010) Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. Atmos Environ 44:476–482. doi: 10.1016/j.atmosenv.2009.11.005

Noori R, Yeh H Der, Abbasi M, et al (2015) Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand. J Hydrol 527:833–843. doi: 10.1016/j.jhydrol.2015.05.046

Palani S, Liong SY, Tkalich P (2008) An ANN application for water quality forecasting. Mar Pollut Bull 56:1586–1597. doi: 10.1016/j.marpolbul.2008.05.021

Park SS, Lee YS (2002) A water quality modeling study of the Nakdong River, Korea. Ecol Modell 152:65–75. doi: 10.1016/S0304-3800(01)00489-6

Pérez CJ, Vega-Rodríguez MA, Reder K, Flörke M (2017) A Multi-Objective Artificial Bee Colony-based optimization approach to design water quality monitoring networks in river basins. J Clean Prod 166:579–589. doi: 10.1016/j.jclepro.2017.08.060

Qaderi F, Babanezhad E (2017) Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network. J Clean Prod 161:840–849. doi: 10.1016/j.jclepro.2017.05.187

Rajaee T, Khani S, Ravansalar M (2020) Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. Chemom. Intell. Lab. Syst. 200:103978

Ranković V, Radulović J, Radojević I, et al (2010) Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. Ecol Modell 221:1239–1244. doi: 10.1016/j.ecolmodel.2009.12.023

Reckhow KH (1994) Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecol Modell 72:1–20. doi: 10.1016/0304-3800(94)90143-0

Singh KP, Basant A, Malik A, Jain G (2009) Artificial neural network modeling of the river water quality-A case study. Ecol Modell 220:888–895. doi: 10.1016/j.ecolmodel.2009.01.004

Tare V, Gurjar SK, Mohanta H, et al (2017) Eco-geomorphological approach for environmental flows assessment in monsoon-driven highland rivers: A case study of Upper Ganga, India. J Hydrol Reg Stud 13:110–121. doi: 10.1016/j.ejrh.2017.07.005

Tiyasha, Tung TM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. J. Hydrol. 585:124670

Varekar V, Karmakar S, Jha R, Ghosh NC (2015) Design of sampling locations for river water quality monitoring considering seasonal variation of point and diffuse pollution loads. Environ Monit Assess 187:1–26. doi: 10.1007/s10661-015-4583-6

Xiang S, Liu Z, MA L (2006) Study of Multivariate Linear Regression Analysis Model For Groundwater Quality Prediction. GUIZHOU Sci 24:60–62

Zhu X, Wang JD, Solo-Gabriele HM, Fleming LE (2011) A water quality modeling study of non-point sources at recreational marine beaches. Water Res 45:2985–2995. doi: 10.1016/j.watres.2011.03.015

Figures

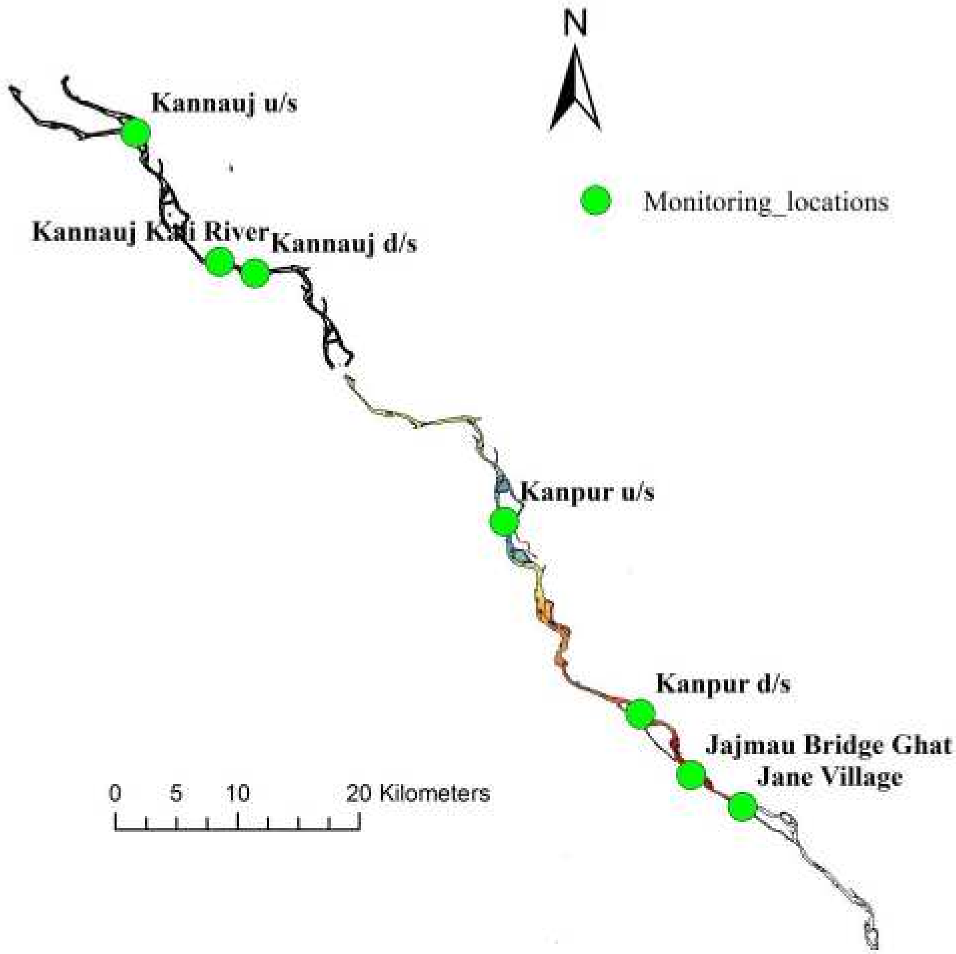


Figure 1

Location of Sampling Stations Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

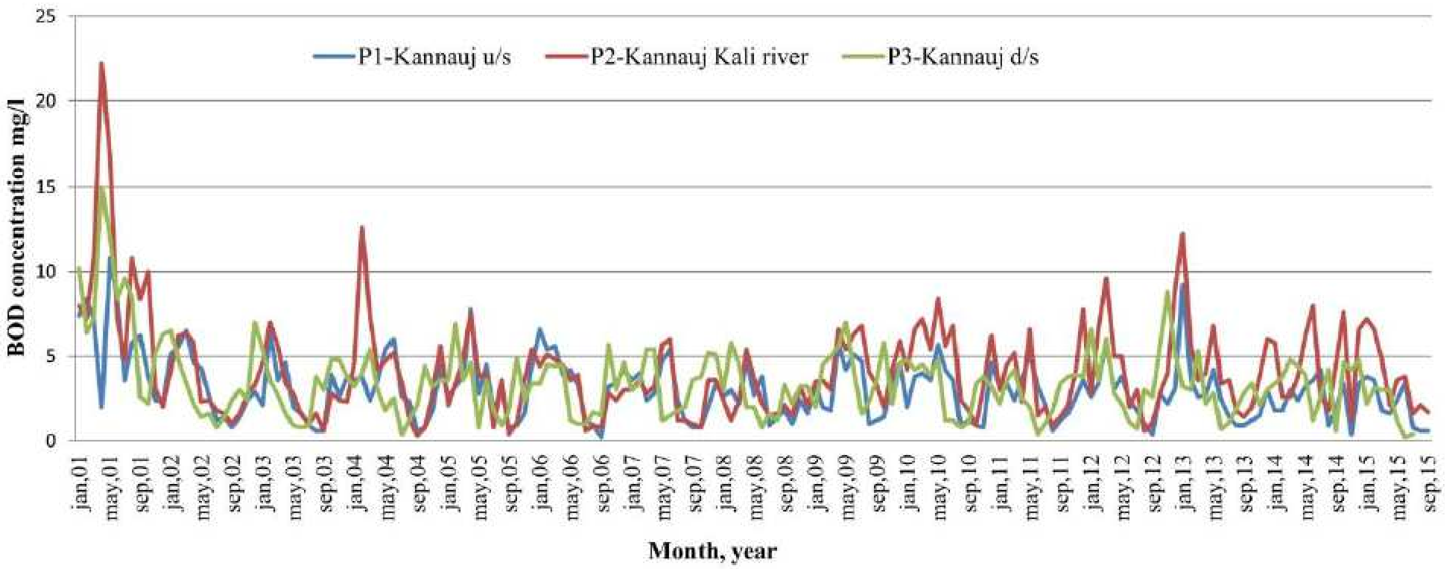
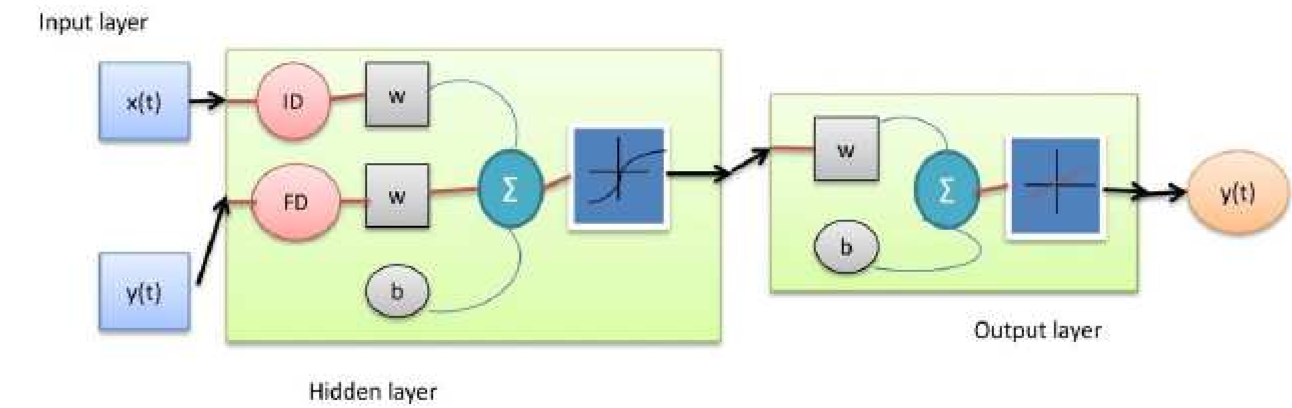
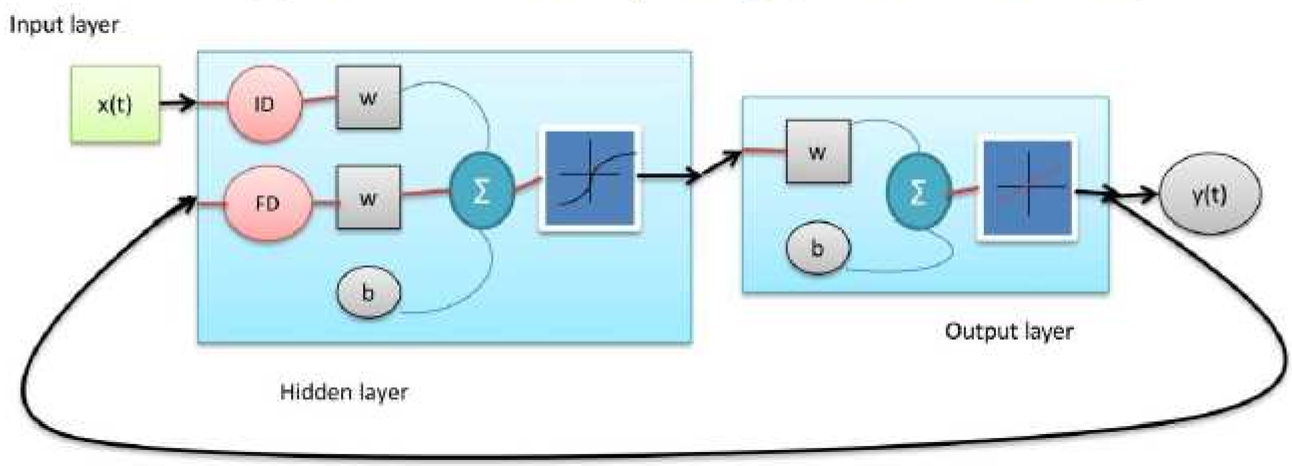


Figure 2

Monthly variation in BOD for stations P1 to P3 from 2001 to 2015



(a) NARX network in open loop (Series Parallel Mode)



(b) NARX network in closed loop (Parallel Mode)

Figure 3

Architecture of NARX network

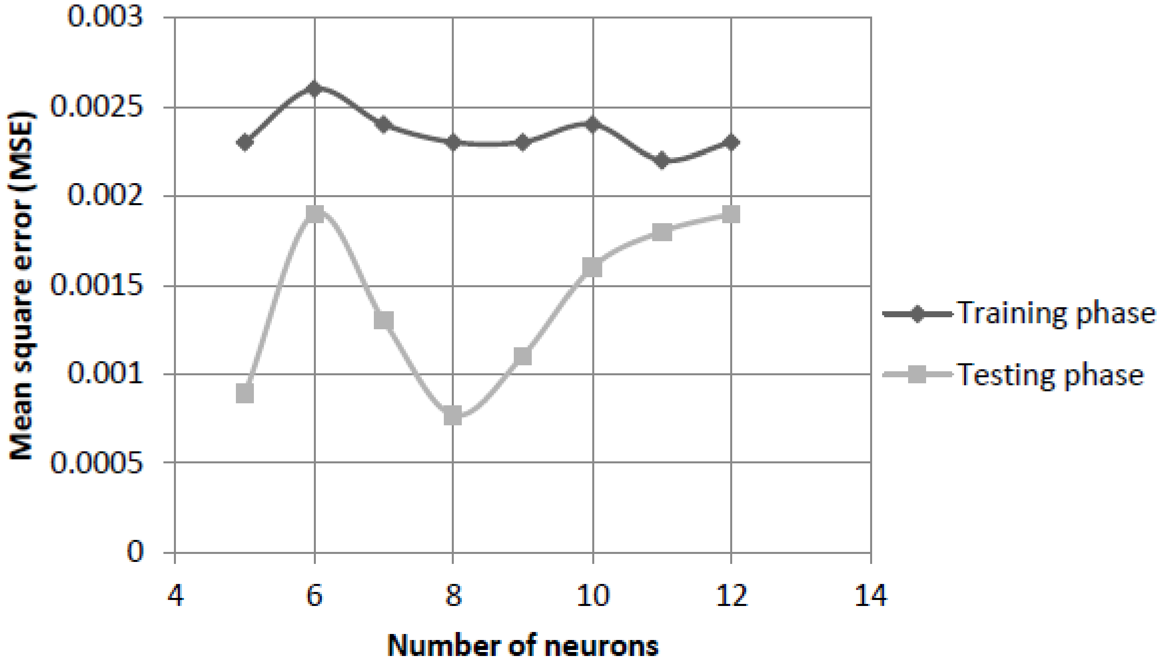
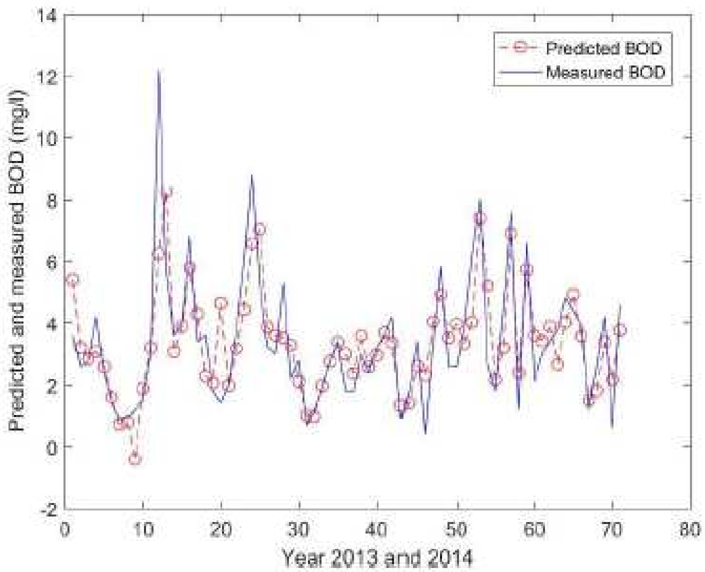
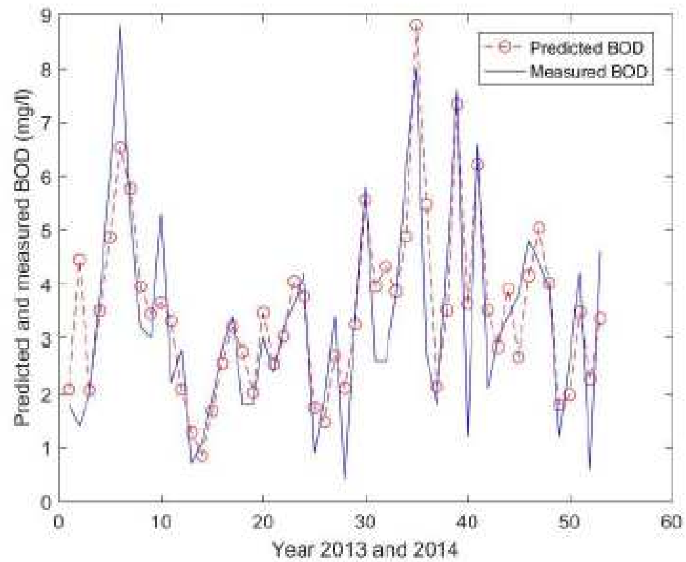


Figure 4

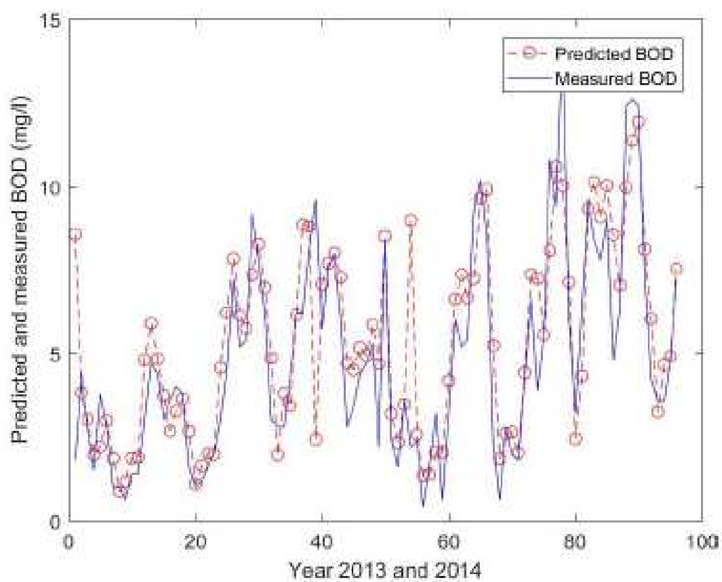
Selecting number of neurons



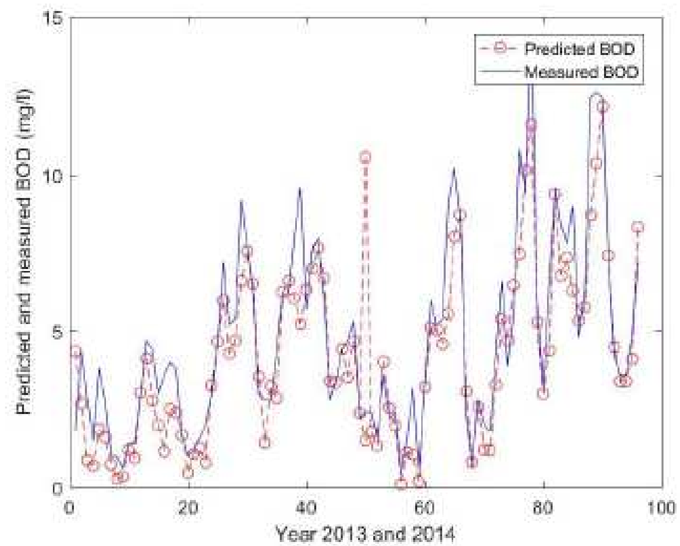
(a) Consecutive days



(b) Non-consecutive days



(c) Consecutive days



(d) Non-consecutive days

Figure 5

Performance of NARX model for station P1-P3 (a, b) and station P4-P7 (c, d) during testing phase

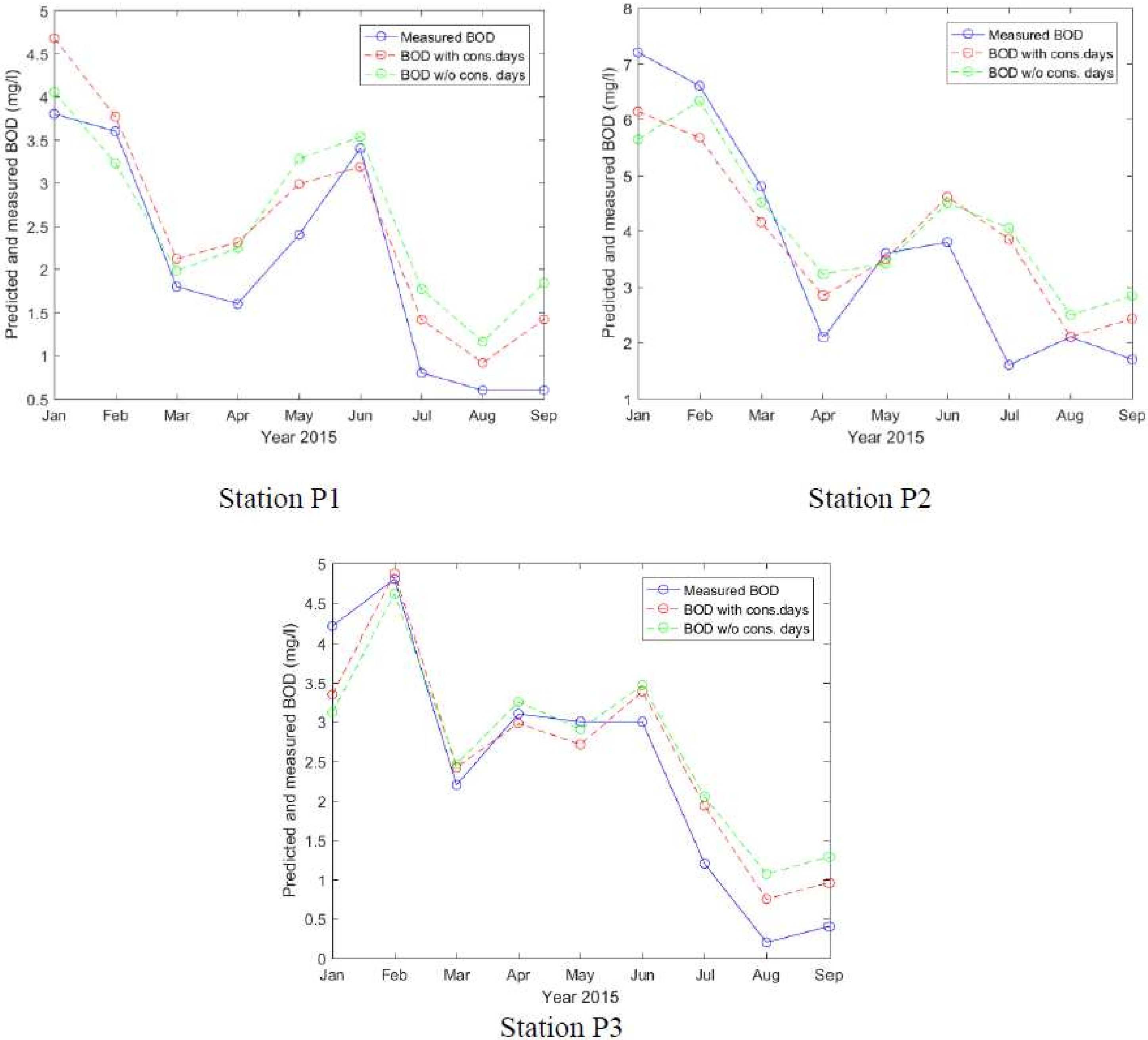


Figure 6

Comparison of predictive performance of NARX model for Kannauj region (P1-P3)

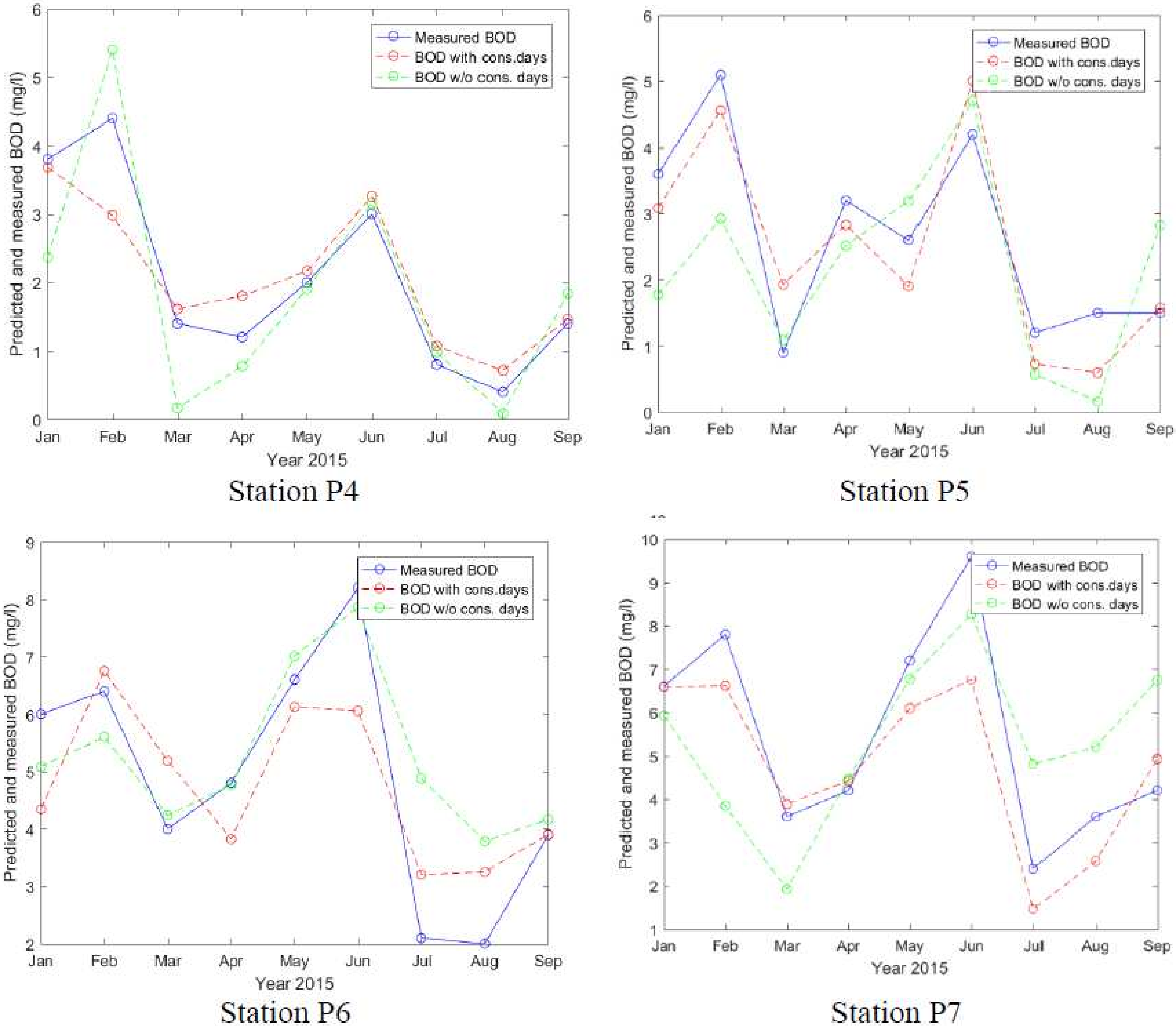


Figure 7

Comparison of predictive performance of NARX model for Kanpur region (P4-P7)