Prediction of surface water quality of Allahabad and Varanasi inUTTAR PRADESH

Ambuj Verma[1], Dr. Anirudh Gupta[2]

[1]*M.Tech. Student, Environmental Engineering, Institute of Engineering and Technology,Lucknow, Uttar Pradesh, India*

[2]*Assistant Professor, Civil Engineering Department, Institute of Engineering and Technology,Lucknow, Uttar Pradesh, India*

Abstract - *Surface Water quality is gradually deteriorate due to population growth and a faster rate of industrialization. Rivers in India are the country's primary source of water. Half of India's population, including two- thirds of the country's poor, lives in the Ganga River Basin, which is the most populous region in the world. This paper highlights the utility of statistical techniques for evaluating, interpreting complex data sets and recognizing spatial differences in water quality for effective management of river water quality. Time-series data and statistical analysis are used by the Autoregressive Integrated Moving Average (ARIMA) model to understand the data and forecast the future. 6 water quality parameters Dissolved Oxygen, BOD, pH, Temperature, Electrical Conductivity and Total Coliform areanalyzed and predicted. In this work 4 monitoring station is taken for the prediction analysis in Allahabad and Varanasi,data is taken from the CPCB. In this work ARIMA model is giving the better prediction of BOD, total coliform and conductivity in compare of other water quality parameter pH, temperature and DO. The max value for correlation coefficientfor Dissolved Oxygen, BOD, pH, Temperature, Electrical Conductivity and Total Coliform are respectively 0.65, 0.90, 0.68, 0.68, 0.86 and 0.84*

Key Words: **ARIMA model, surface water quality prediction, Allahabad, Varanasi**

1. INTRODUCTION

Since ancient times, the river Ganges has been revered as one of the most sacred and holy rivers in the world. Since thebeginning of time, people have revered the Ganges as one ofthe most holy and sacred rivers in the world. The evaluation of river water quality is a critical element in the assessment of water resources. The quality of water that is consumed defines the base line of protection against many diseases and infections. The Ganga, with over 2,525 km long main-stem along with her tributaries has constantly provided material, Spiritual and cultural sustenance to millions of people living in and around its basin. The riverine water resources provide irrigation, drinking water, economical transportation, electricity, recreation and religious fulfilment, support to the aquatic ecosystem as well as livelihoods for many stakeholders. This river has both emotional and spiritual value among Indians. The water of Ganga carries religious sentiments and is considered as the purest water which can wash off all the sins of the human being. However, present study is carried out with an objective to assess the water quality of the Ganga water at its descendent point on the plains where it is supposed to be least polluted. The river Ganges in India is regarded as the most holy and sacred rivers of the world by Hindus from time immemorial. Bhagirathi is the source stream of Ganga. The river has been the focus of national and international intervention and study for past several decades to identify and establish causes and impact of anthropogenic activities on river water quality.

1.1 Study Area

Ganges River, Great River of the plains of the northern Indian subcontinent. Although officially as well as popularly called the Ganga in Hindi and in other Indian languages, internationally it is known by its conventional name, the Ganges.

In this paper we have selected four locations in the upper Ganga stretch in the Uttar Pradesh. These locations are given below:-

1. GANGA AT ALLAHABAD U/S (1046)
2. GANGA AT ALLAHABAD D/S (1049)
3. GANGA AT VARANASI U/S (ASSIGHAT) (1070)
4. GANGA AT VARANASI D/S (MALVIYA BRIDGE) (1071)

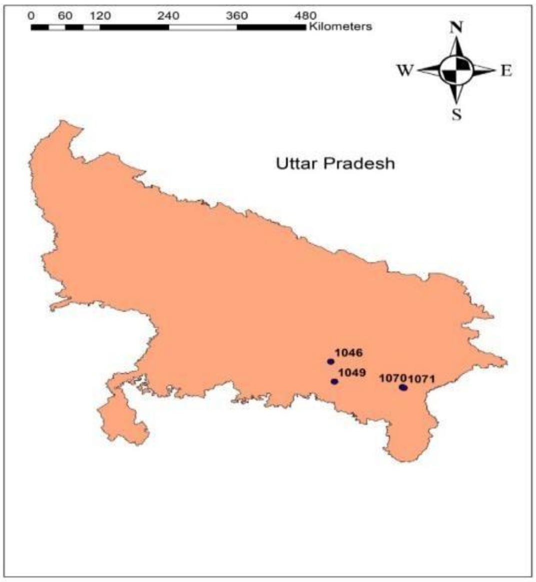


Figure 1; location of station in Uttar Pradesh

3. Materials and methods

Time series forecasting is one of many data analysis techniques that is often utilised in a variety of fields. One of the most significant and often used time series models is the autoregressive integrated moving average (ARIMA) model. ARIMA models can also implement a variety of exponential smoothing techniques. It is also referred as the Box-Jenkins methodology, which consists of a series of steps for identifying, estimating and diagnosing ARIMA models with time series data. ARIMA models have proven to be capable of producing precise short-term estimates.

Every component achieved are provided as a parameter in this model. ARIMA (p, d, q) is a standard notation in which the parameters are replaced by integer values to immediately indicate the ARIMA model being utilized. The ARIMA model's parameters are (p) it defines the number of lagged observations in the model, also referred as the lag order, (d) it can be understood as number of raw observations differenced, also referred as the degree of differencing, and (q) it is referred as the moving average order or the size of window in moving average.

In this work ARIMA (1, 1, 1) is used for the prediction of the surface water quality.

Mathematically,

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon t + \emptyset_1 \epsilon_{t-1} + \emptyset_2 \epsilon_{t-2} + \cdots + \emptyset_q \epsilon_{t-q}$$

Where, Yt is the predicted value, Yt-1 is the lag 1 of the series, β1 is the coefficient of lag 1, α is the constant or the intercept term, ϵt is the random error at t, ϵt-1 is the lagged error term at t-1, ∅ is the coefficient of lagged error at t-1.

a. Data source

The data used in this research and prediction is from the Central Pollution Control Board (CPCB). Monthly data is taken from the 2016 to 2020 for the 6 surface water quality parameter and it is arranged in the MS Excel for the further calculation and the analysis. Dissolved oxygen, BOD, pH, temperature, electrical conductivity, and total coliform are the next parameters that this study is measuring.

b. Data statistics

The following data is observed on the stations:-

Station	Parameter	Temperature	DO	pH	Conductivity	BOD	TotalColiform
1046(S1)							
	Minimum	18.00	6.80	7.45	154.00	2.10	7900.00
	Maximum	32.50	12.20	8.49	385.00	5.60	58000.00
	Mean	25.80	8.55	8.10	259.63	3.52	25851.67
	Std. deviation	3.84	1.22	0.28	75.07	0.82	10822.11
1049(S2)							
	Minimum	18.20	6.80	7.49	204.00	2.00	8400.00
	Maximum	32.80	11.50	8.38	488.00	5.80	63000.00
	Mean	26.08	8.27	8.06	324.92	3.57	27140.00
	Std. deviation	3.89	1.05	0.23	75.45	0.90	12265.11
1070(S3)							
	Minimum	20.00	7.20	7.35	320.00	1.70	1100.00
	Maximum	31.50	10.00	8.60	535.00	3.50	3600.00
	Mean	25.90	8.24	8.28	421.41	2.85	2514.29
	Std. deviation	3.14	0.69	0.20	55.63	0.38	760.11
1071(S4)							
	Minimum	20.00	6.10	7.30	338.00	3.40	17000.00
	Maximum	31.50	8.60	8.80	594.00	6.70	70000.00
	Mean	25.92	6.93	8.28	458.80	4.79	42732.14
	Std. deviation	3.15	0.51	0.25	69.05	0.98	11395.05

Figure 2; Stastical analysis of station S1, S2, S3 and S4

4. Result and discussion

Model performance was estimated by RMSE, MAPE, AIC and R^2 for the water quality parameters Dissolved Oxygen, BOD, pH, Temperature, Electrical Conductivity and Total Coliform. In summary, the ARIMA model performed significantly better prediction of temperature, total coliform and conductivity in compare of other water quality parameter pH, BOD and DO. Furthermore, different prediction performance can be found for the four sites. The results for monthly surface water quality are shown in Figure 3, 4, 5 and 6.

The prediction of station S1 are –

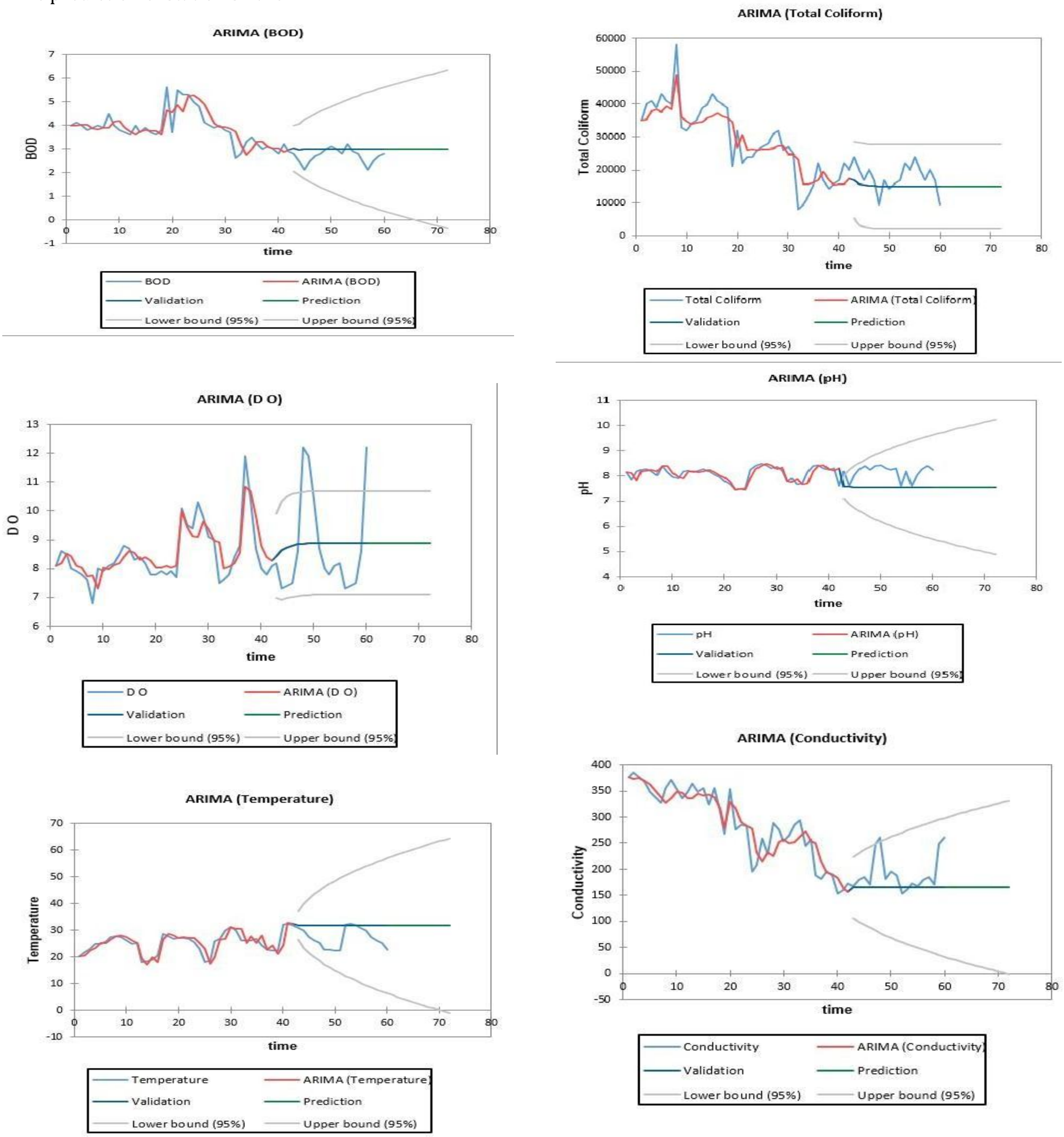
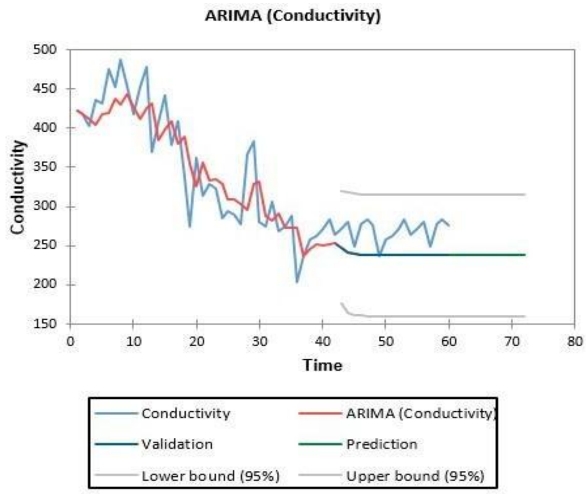
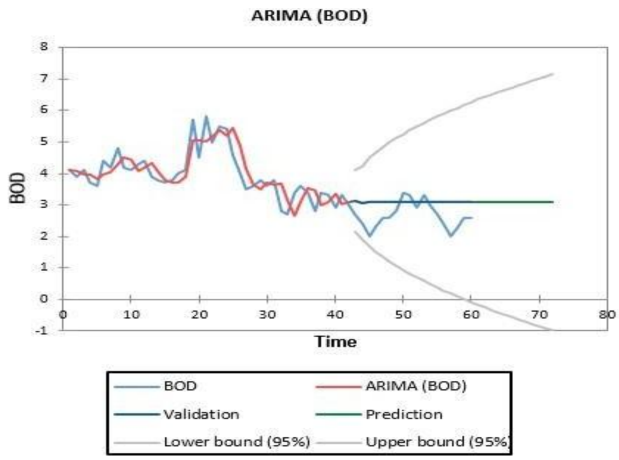
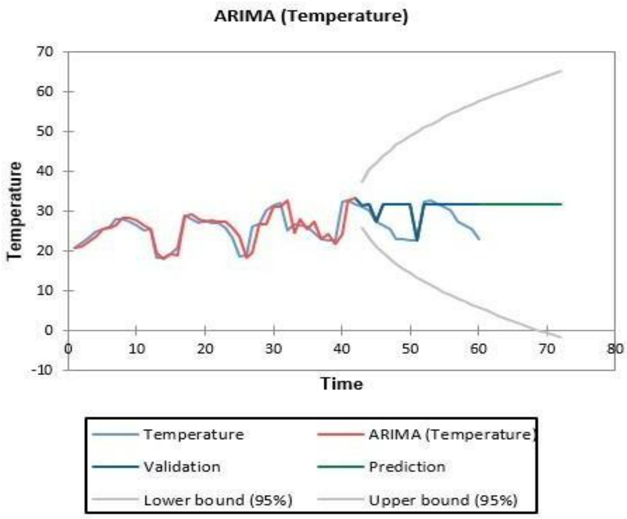


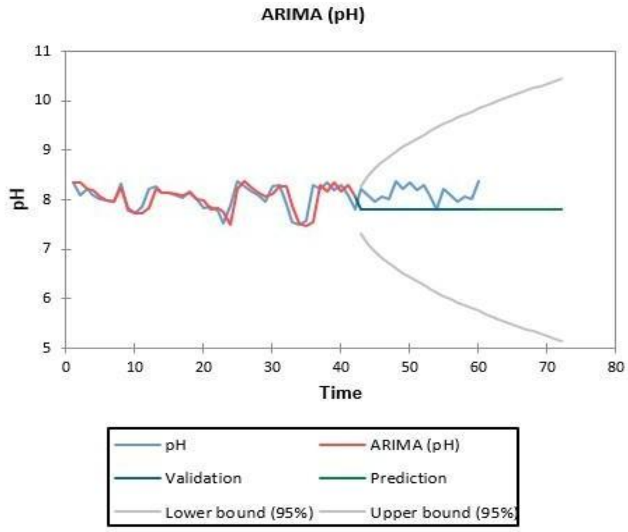
Figure 3; Performance of ARIMA model for station S1

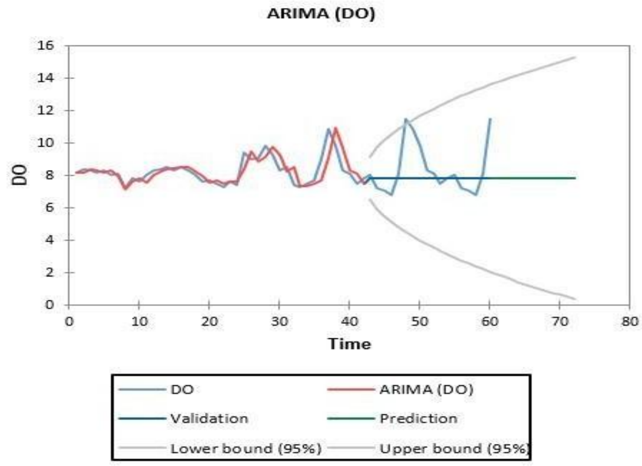
The prediction of station S2 are –











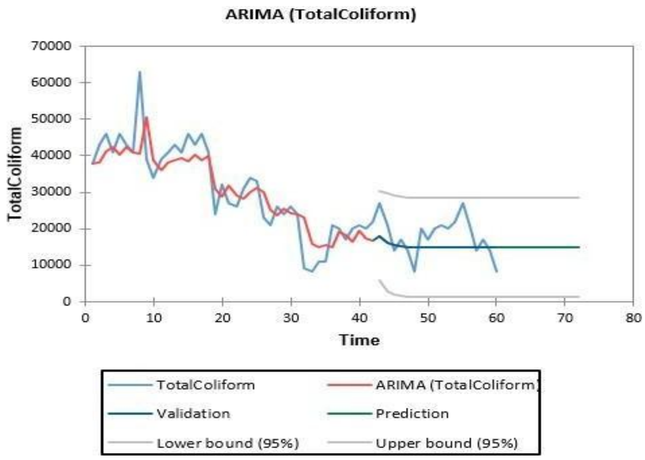
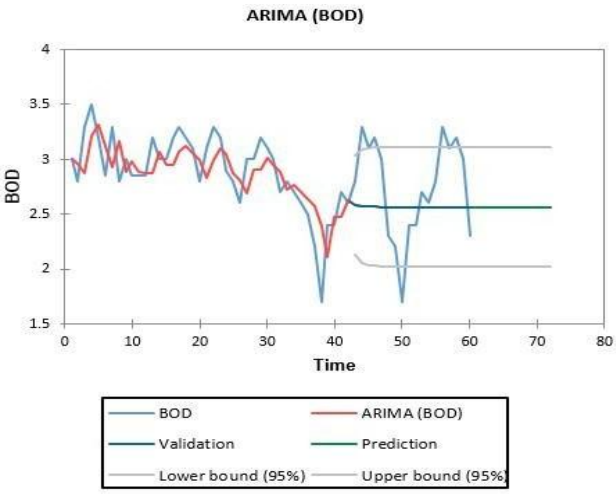
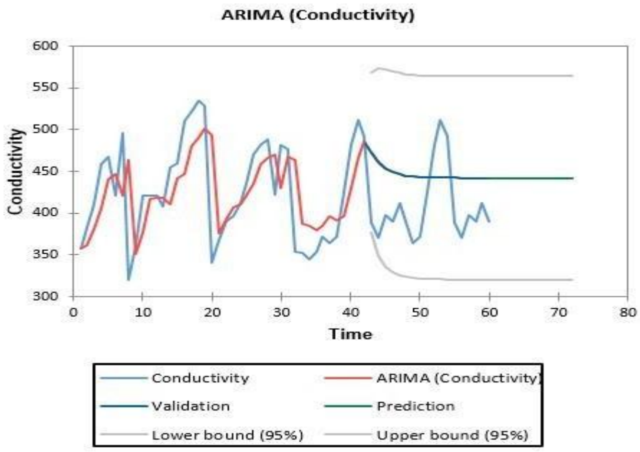
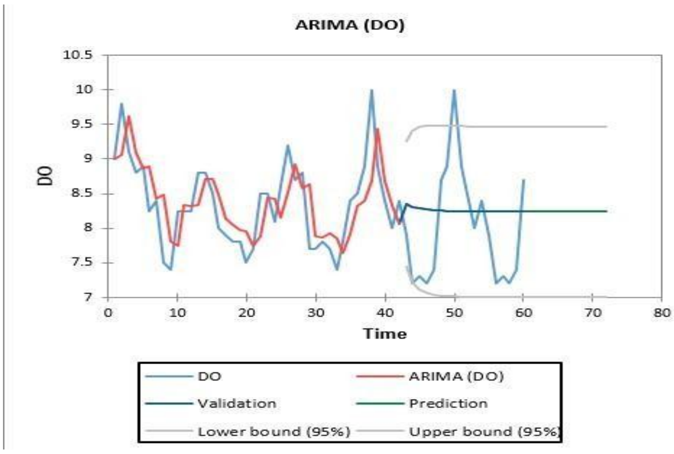


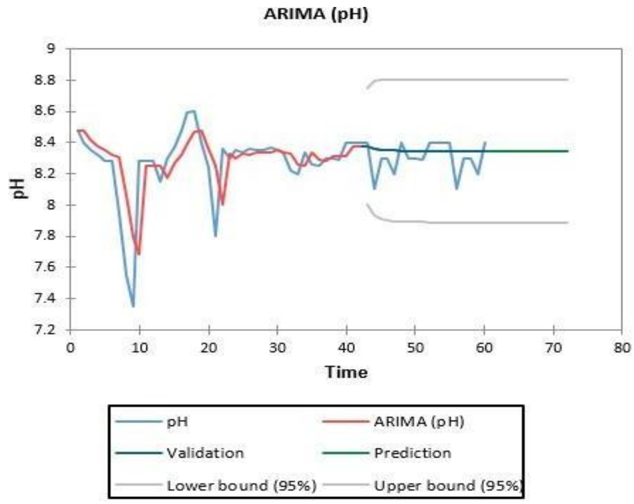
Figure 4; Performance of ARIMA model for station S2

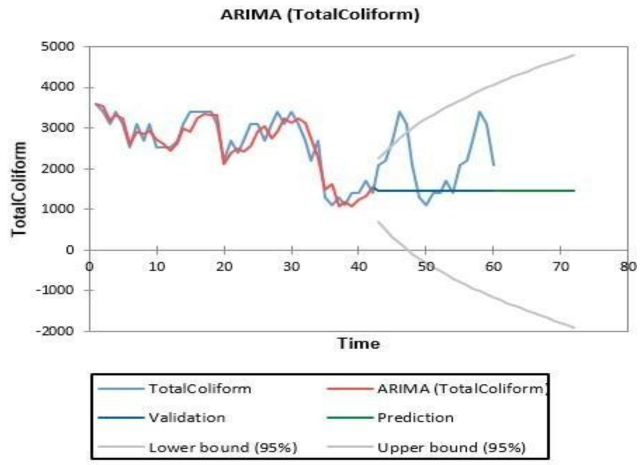
The prediction of station S3 are –











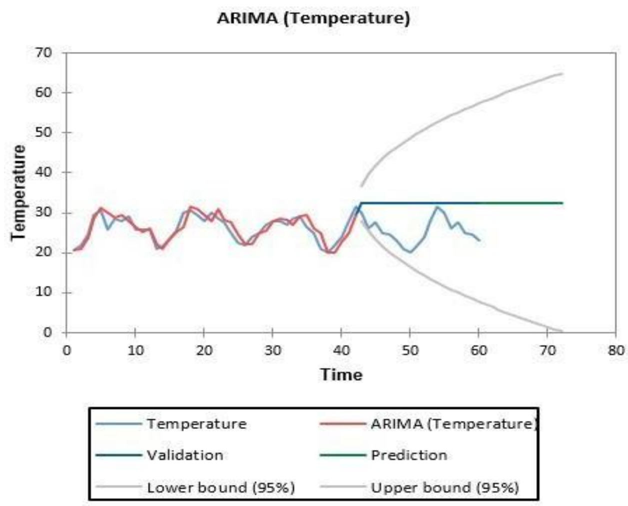
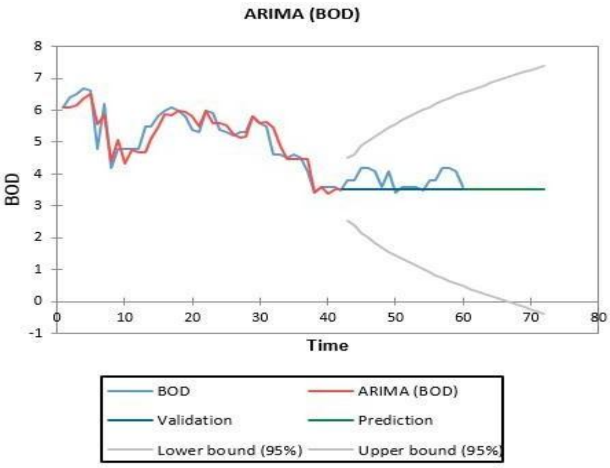
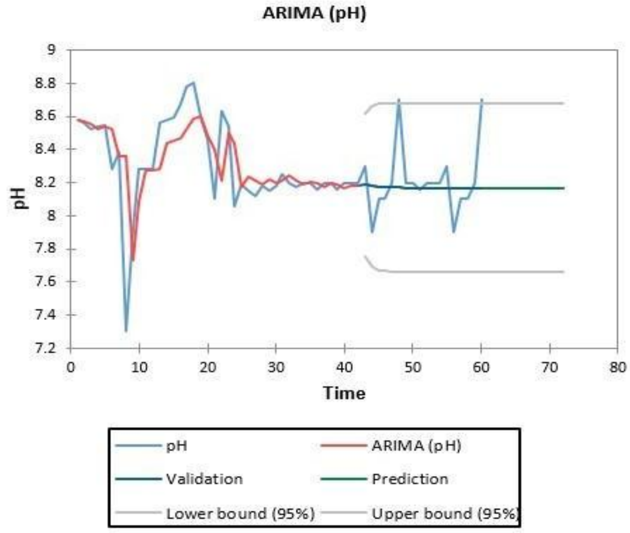
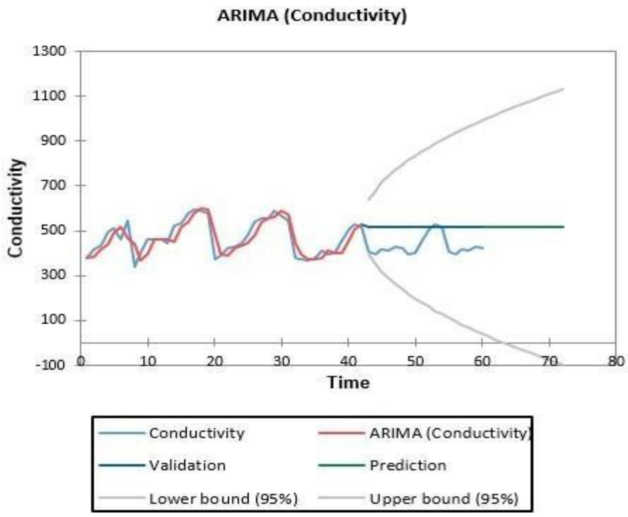


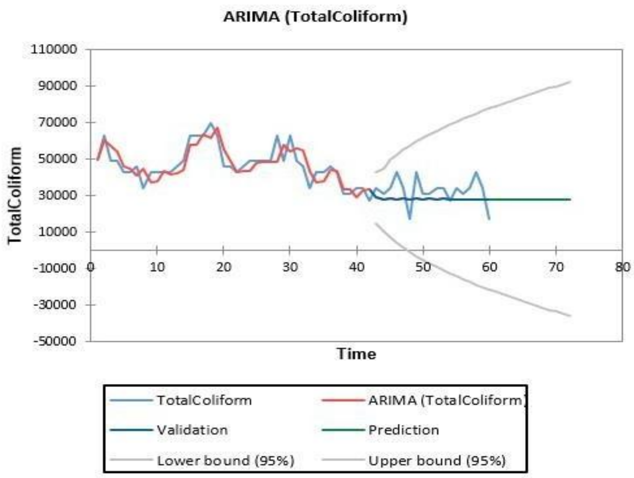
Figure 5; Performance of ARIMA model for station S3

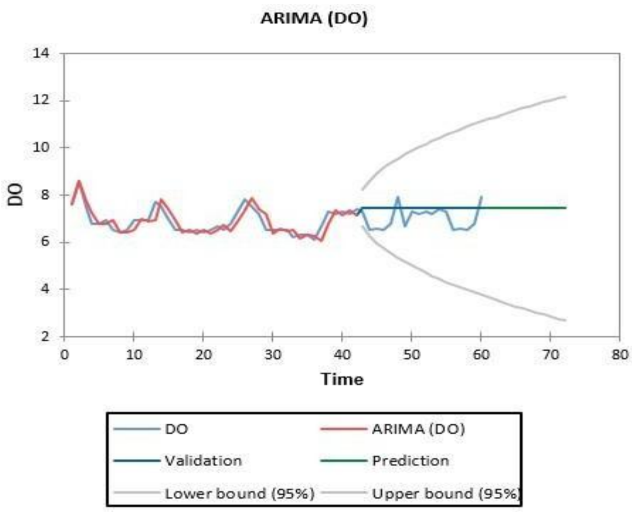
The prediction of station S4 are –











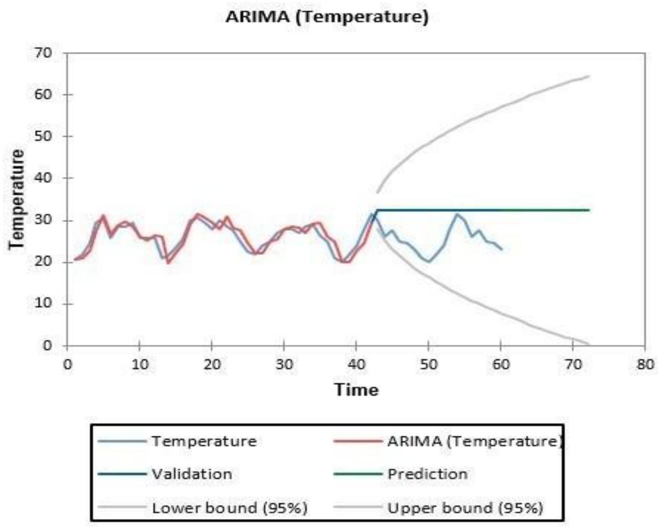


Figure 6; Performance of ARIMA model for station S4

5. CONCLUSIONS

The study displays prediction accuracy and error for all chosen sites for which analysis is conducted. The results shows for S1 station that this model is able to predict the best value for the conductivity with 0.83 R^2 value and pH has lower value 0.52 R^2 . However, station S2 has highest efficiency 0.86 for the conductivity and min value 0.52 for the pH. Station S3 has highest efficiency .75 for the total coliform and min value 0.54 for the temperature. Station S4 has highest efficiency .90 for the BOD and min value 0.55 for the temperature and conductivity.

Station	Parameter	Temperature	DO	pH	Conductivity	BOD	TotalColiform
1046(S1)							
	RMSE	2.75	0.75	0.23	30.11	0.50	5954.63
	MAPE	7.50	5.20	2.01	8.54	8.13	18.81
	AIC	206.94	101.00	3.49	401.86	65.23	838.02
	R^2	0.64	0.54	0.52	0.83	0.77	0.81
1049(S2)							
	RMSE	2.94	0.67	0.24	36.52	0.50	6229.06
	MAPE	7.41	5.40	2.12	8.71	9.78	19.30
	AIC	212.09	90.08	5.09	420.32	66.10	841.62
	R^2	0.68	0.54	0.52	0.86	0.74	0.84
1070(S3)							
	RMSE	2.23	0.46	0.19	49.08	0.23	395.85
	MAPE	6.83	4.33	1.43	8.59	6.72	14.14
	AIC	188.37	61.74	-11.10	444.00	5.65	612.88
	R^2	0.54	0.61	0.68	0.55	0.70	0.75
1071(S4)							
	RMSE	2.25	0.40	0.22	61.99	0.50	7244.57
	MAPE	2.25	4.01	1.41	8.95	6.80	12.20
	AIC	188.78	46.29	0.56	460.78	65.12	851.35
	R^2	0.55	0.65	0.60	0.55	0.90	0.71

Figure 7; Performance of ARIMA model for different water quality parameter

In conclusion, it is clear that the ARIMA model performs better in predicting BOD, total coliform, and conductivity than other water quality parameters such as pH, temperature, and DO. This complete analysis provides an information base to be used by regulators and policymakersfor reconciling the competing interests in the Ganga river through delivering solutions to improve, monitor clean up, maintain water quality and restore its ecosystem.

ACKNOWLEDGEMENT

The author wishes to express his paramount gratitude to Asst. Prof. Dr. Anirudh Gupta CED, IET Lucknow, their scholarly guidance and constructive feedback have been a source of inspiration and motivation to complete this paper and the Central Pollution Control Board (CPCB) for providing the essential data for conducting this paper.

REFERENCES

[1] Bagla, P., Kumar, K., Sharma, N., & Sharma, R. (2021). Multivariate Analysis of Water Quality of Ganga River. *Journal of The Institution of Engineers (India): Series B*, *102*(3), 539-549.

[2] Bhutiani, R., Khanna, D. R., Kulkarni, D. B., & Ruhela, M. (2016). Assessment of Ganga river ecosystem at Haridwar, Uttarakhand, India with reference to Water Quality Indices. *Applied Water Science*, *6*(2), 107-113.

[3] Ighalo, J. O., Adeniyi, A. G., & Marques, G. (2021). Artificial intelligence for surface water quality monitoring and assessment: a systematic literature analysis. *Modeling Earth Systems and Environment*, *7*(2), 669-681.

[4] Zhang, Y. F., Thorburn, P. J., Xiang, W., & Fitch, P. (2019). SSIM—A deep learning approach for recovering missing time series sensor data. *IEEE Internet of Things Journal*, *6*(4), 6618-6628.

[5] Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient water quality prediction using supervised machine learning. *Water*, *11*(11), 2210.

[6] Matta, G., Kumar, A., Nayak, A., Kumar, P., Kumar, A., & Tiwari, A. K. (2020). Determination of water quality of Ganga River System in Himalayan region, referencing indexing techniques. *Arabian Journal of Geosciences*, *13*(19), 1-11.

[7] Hassan, M. M., Hassan, M. M., Akter, L., Rahman, M. M., Zaman, S., Hasib, K. M & Mollick, S. (2021). Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms. *Human-Centric Intelligent Systems*, *1*(3-4), 86-97.

[8] Abdelmalik, K.W. (2016). Role of statistical remote sensing for Inland water quality parameters prediction. Egypt. J. Remote Sensing Space Sci. 21, 154.

[9] Avila, R., Horn, B., Moriarty, E., Hodson, R., and Moltchanova, E. (2018). Evaluating statistical model performance in water quality prediction. J. Environ. Manage. 206, 910.

[10] Bedri, Z., Corkery, A., O'Sullivan, J.J., Deering, L.A., and Demeter, K. (2016). Evaluating a microbial water quality prediction model for beach management under the revised EU Bathing Water Directive. J. Environ. Manage. 167, 49.

BIOGRAPHIES

Ambuj Verma is a student of Master of Technology, Environmental Engineering, Dept. of Civil Engineering, Institute of Engineering and Technology, Lucknow, Uttar Pradesh.

Dr. Anirudh Gupta is working asAssistant Professor, Dept., of Civil Engineering in Institute of Engineering and Technology, Lucknow, Uttar Pradesh.