.

Predictability Indicators of River Water Quality with AI-ML Modelling

Dilip Chate*

SMIT Enviro Solutions, Vantage Capital, Office 1016 & 1017, Pune-411057

*Email: admin@smitenvirosolutions.in

Received: 18.5.2025, Revised: 8.7.2025, Accepted: 13.7.2025

Abstract

This study presents the time-series analysis, frequency distributions, coefficients of determination and performance matrices for water quality predictability of distinctive Patalganga and Bhatsa rivers by processing eight water quality parameters database of recent two decades (2003 to 2023). The databases of water temperature, Dissolved Oxygen (DOs), Biochemical Oxygen Demand (BOD), pH, conductivity, nitrate + nitrite, fecal coliform and total coliform executions with General Regression and Extreme Gradient Boosting (XGBoost) modelling systems resulted in frequency distributions, coefficients of determinations, performance matrices with Root Mean Squared Error (RMSE), Mean Squared Error (MSE) and Mean Absolute Error (MAE). The results are discussed in a comparative analysis of time-elapsed evolutions with trend-line fits representing the variabilities in water quality over the period of recent 20 years. The BOD and total coliform shows better correlation with coefficient of 0.6 for Bhatsa and a very good correlation of 0.72 between fecal coliform and total coliform for Patalganga. Also, coefficients of determinations of 0.5> recommends fecal coliform, total coliform, BOD, conductivity, Nitrate + Nitrite best water quality predictors and XGBoost performers (RMSE, MSE, or MAE) of <1 supports the same. It means XGBoost model's average prediction error is very low (<1), which seems to be a sign of a good fit, especially the target variable's scale supports such smaller errors. The water quality data-driven analytical results of this case study corroborates the AI-ML forecasting system to be viewed as the balancing approach towards ongoing selective river water samplings with chemical analysis. This system can be the proactive decision-making tool in river water pollution management.

Keywords: water quality, predictability, AI-ML tools, data analysis, frequency distributions

1. Introduction

River pollution leads to significant loss to biodiversity and aquatic lives attributed to oxygen depletion by untreated sewage and industrial wastewater discharge. Agricultural lands in downstream suffer from contamination due to influx of polluted water, leading to crop damage and soil degradation. Local communities relying on the river water bodies for irrigation, fishing, and domestic usage experience the health hazards and economic losses[1]. The river water quality varies significantly across the Maharashtra. The Maharashtra Pollution Control Board (MPCB) monitors the river water quality, nevertheless, data availability and transparency as well as enforcement remains the challenge. According to MPCB, out of 251 monitored river locations state-wide (2007–2011), only about 10% consistently met all water quality standards[2]. Many rivers, especially those near urban and industrial zones, frequently exceeded permissible limits of the Biochemical Oxygen Demand (BOD), Dissolve Oxygen (DO), and coliforms[2]. While there are several studies on water pollution globally[3], there is limited research focusing specifically, on the water quality predictability of Maharashtra's rivers, which face unique challenges due to rapid urbanization and industrialization. The rapid industrial and urban growth has led to significant pollution in rivers through increased discharge of untreated industrial effluents, domestic sewage, and urban runoff. The lack of control over industrial effluents and inadequate waste management practices have resulted in alarmingly high pollution levels in water of the rivers. The degradation of water quality in rivers poses substantial environmental, health, and socio-economic risks.

The ongoing monitoring methods rely on physical sampling and laboratory analysis, which are time-consuming, costly, and spatially limited, making it difficult to provide real-time assessments and early warnings for pollution events. The data-driven AI-ML analytical techniques can be the complementary approach to traditional sampling and chemical analysis. The computational libraries offer data cleaning, data classification, visualizations of frequency distributions, generates heatmaps of coefficient of determinations from the river water quality variables. The best correlations among the water quality indicators specify the best predictors for simulations of water quality scenarios and predictability of water quality with performance metrics with the evaluation of AI-ML modelling system.

By identifying trends and predicting potential contamination events, traditional AI-machine learning (ML) models can help policymakers to implement targeted mitigation strategies, optimize wastewater treatment, and ensure drinking water supply safety. The AI-ML methods will be an added advantage over the existing water quality assessment methods. The AI-ML modelling systems can be trained on historical water quality datasets. The calibration and validation of these models should be performed using observational datasets that the AI-ML models have not previously encountered. Indeed, it must be cautioned that the accuracy of the AI-ML simulations has to be verified against the data gathered from regulatory grade water quality monitoring network.

This paper discusses the predictability of river water quality using AI-ML methods. Such fast and accurate analysis will be useful in water usage-related decision-making processes. The AI-ML methodology augmented with existing water quality analysis might be advantageous for environmental planning with respect to ongoing water quality management. In this study, we perform the general regression data analysis of recent 20-year water quality data (2003 to 2023) of Bhatsa and Patalganga rivers. Furthermore, we divide the long-term databases of eight variables for training and validating the Extreme Gradient Boosting (XGBoost) modelling system to determine the best data-driven indicators for predictability of water quality. Correlation matrices with model performing scores found apposite to policy-decision making water quality scenarios for river water pollution management with abatement action plans.

2. Experimental

2.1 Bhatsa and Patalganga Rivers

The Bhatsa river is a crucial freshwater source for Mumbai, supplies nearly half of the city's drinking water. The water quality is monitored regularly due to requirements of water supply for drinking, irrigation, and fisheries[4]. The Bhatsa's main challenges are the turbidity and coliforms exceedance during the monsoon, but its core water quality parameters generally remain within the limits for drinking water after the treatment[4]. The Patalganga river was historically a vital source of drinking and irrigation water for local communities and supported a rich aquatic ecosystem[7]. Over the past 50 years, the Patalganga River has experienced a steady and significant water quality degradation due to inadequate system for wastewater treatment with rapidly growing industrialization and urbanization, The Patalganga river, flowing through the Raigad district of Maharashtra, is among the most polluted rivers in the state. The water pollution is primarily attributed to untreated or partially treated industrial effluents from the Patalganga’s Maharashtra Industrial Development Corporation zone with dyeing, fertilizer, pesticide, steel, chemical, and other manufacturing industries[1]. Patalganga river often shows much higher levels of pollution, particularly with respect to organic load

(BOD) and microbial contamination as compared to that of Bhatsa[5,6]. Patalganga is classified as a Priority-III polluted river[7]. The Central Pollution Control Board (CPCB) has listed Patalganga among the 38 most polluted rivers in India which stresses the policy-induced water pollution abatement plans with implementation of innovative approach with AI-ML techniques.

2.2 Water Quality Databases

The water quality data over a period of 2003 to 2023 (20 years) of Bhatsa and Patalganga rivers are acquired from the CPCB (https://cpcb.nic.in/). The water quality datasets consist of 8 variables such as water temperature, Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), pH, conductivity, Nitrate + Nitrite, Fecal coliform and total coliform. The water quality databases of 8 variables seasonally averaged over annual scales available for the recent 20 years are assumed to be quality controlled and quality assured for comparative analysis of distinct rivers of Bhatsa and Patalganga.

2.3 General Regression AI-ML computational techniques

The databases of 8 water quality variables in times-series and trend-lines representing each of the water quality variables with linear fits analysis and regression equations superimposing on the same variable time-series distinguishes preventive measures scenarios from business as usual scenarios. Furthermore, these databases of Bhatsa and Patalganga executing with the inbuilt standard computational packages of general regression techniques on paid cloud computing workspace automatically cleans the datasets of 8 variables and generates data visualizations in terms of frequency distributions for water temperature, DO, BOD, pH, Conductivity, Nitrate + Nitrite, Fecal Coliform and total coliform.

2.4 Extreme Gradient Boosting (XGBoost) Modelling systems:

The Extreme Gradient Boosting (XGBoost) is a traditionally used gradient boosting method for regression analysis, known for its efficiency and scalability even with the smaller size databases[8]. It combines a loss function that quantifies prediction error with a regularization term that prevents overfitting by encouraging simpler models. The XGBoost objective function can be typically expressed in terms of following equation[8,9]:

$$\sum_{i=1}^{n}[L(yi + \widehat{yi})\ \psi(f)$$

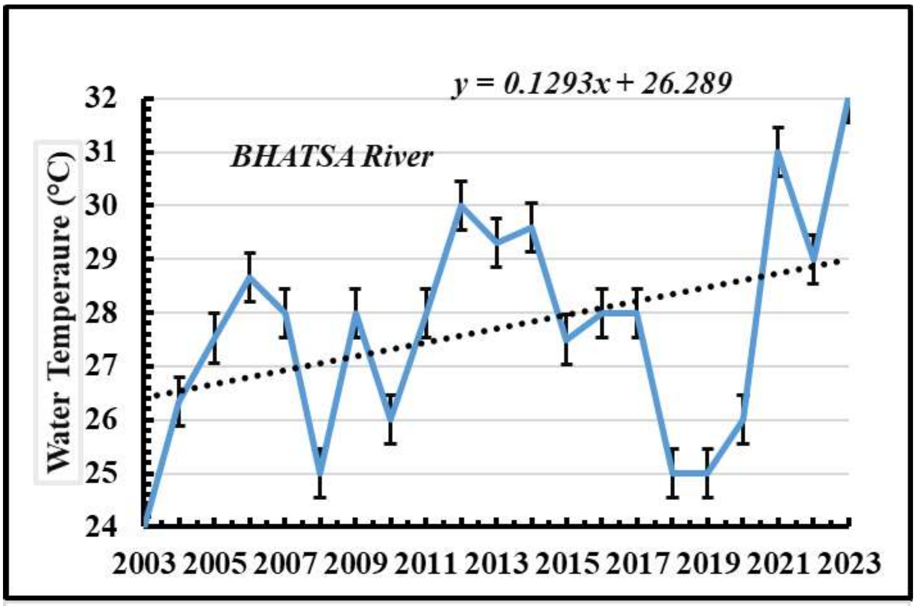
where $L(yi + \widehat{yi})$ represents the loss function that measures the difference between the true value yi and the predicted value $\widehat{yi}$, while $\psi(f)$ is the regularization term that penalizes model complexity encourages simpler models. The regularization term $\psi(f)$ ensures better generalization of the model by reducing overfitting.

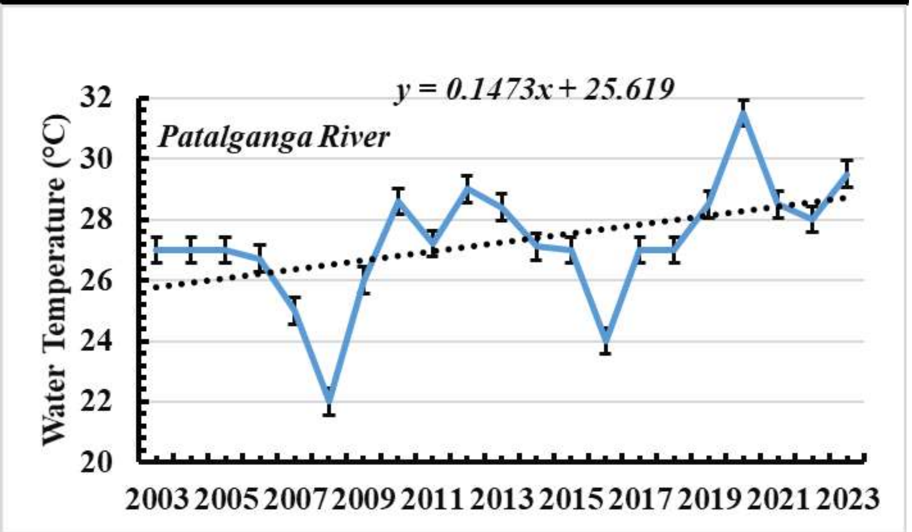
The quality controlled and quality assured water quality databases of Bhatsa and Pataganga rivers for the period of 2003 to 2023 are described elsewhere (https://cpcb.nic.in/). These datasets of 8 water quality variables are good for a first general regression and XGBoost models because all of the input variables are numeric and simple binary classification problem. The general regression and (XGBoost) classifies the eight variables in the water quality databases, recognises the patterns and précises the model outcomes with the distribution functions. The essential computing libraries are installed and configured on Virtual Machines (VMs) of cloud infrastructure to tune the XGBoost model hyper-parameters and heuristics for good parameter values, number and size of trees, the learning rate and finally the sampling rates in stochastic variations. The XGBoost optimises the decision tree algorithms to enhance water quality database processing[8,9]. In this method different terms of regularizations are selected to reduce the risk of overfitting. The training and testing of XGBoost models are performed on the paid resources (VMs) of Amazon cloud infrastructure (AWS). The water quality datasets are divided into 75% training and 25% testing datasets. The training dataset prepares the XGBoost model and the test dataset makes the predictions, from which we can evaluate the model performance. Furthermore, the inbuilt functions in the XGBoost modelling system executes and calculates automatically during the training phase of water quality dataset. The XGBoost modelling system consists of several decision trees, where, each tree considers the residuals of the previous tree and utilizes a gradient algorithm to construct new decision trees. As a result, the residual errors during the training process are minimized. The preceding decision trees are redeveloped in each iteration which further reduces the residual errors[9].

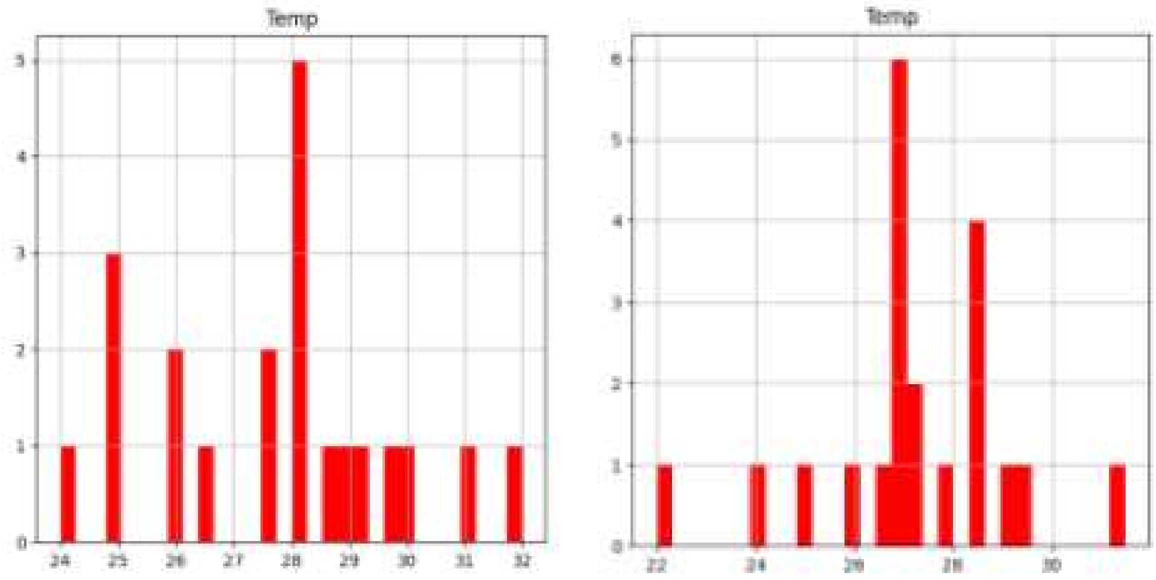
3. Results and Discussion

3.1 Time-series analysis and frequency distributions of water quality indicators:

For most of the rivers, when the water temperature exceeds 30 °C there is a significant strain on fish. Coldwater fisheries should be at water temperature < 18.9 °C and that of warm water fisheries at water temperature < 30 °C. Water temperature trend-lines with data spread and frequency distributions are shown in fig. 1 for Bhatsa and Patalganga. The time-series with error bars of water temperature trend-lines indicates the increasing trend over a period of 2003 to 2023 for both Bhatsa and Patalganga on account of industrialisations, rapid urbanisations, and changes in land use land covers. The maximum frequency distribution occurred at water temperature of 28°C for Bhatsa as compared to that of Patalganga at water temperature of 27°C over a span of 20 years (Fig. 1). Water temperature increasing trend with maximum frequency can be managed with restrictions of urbanizations within buffer zones of rivers.



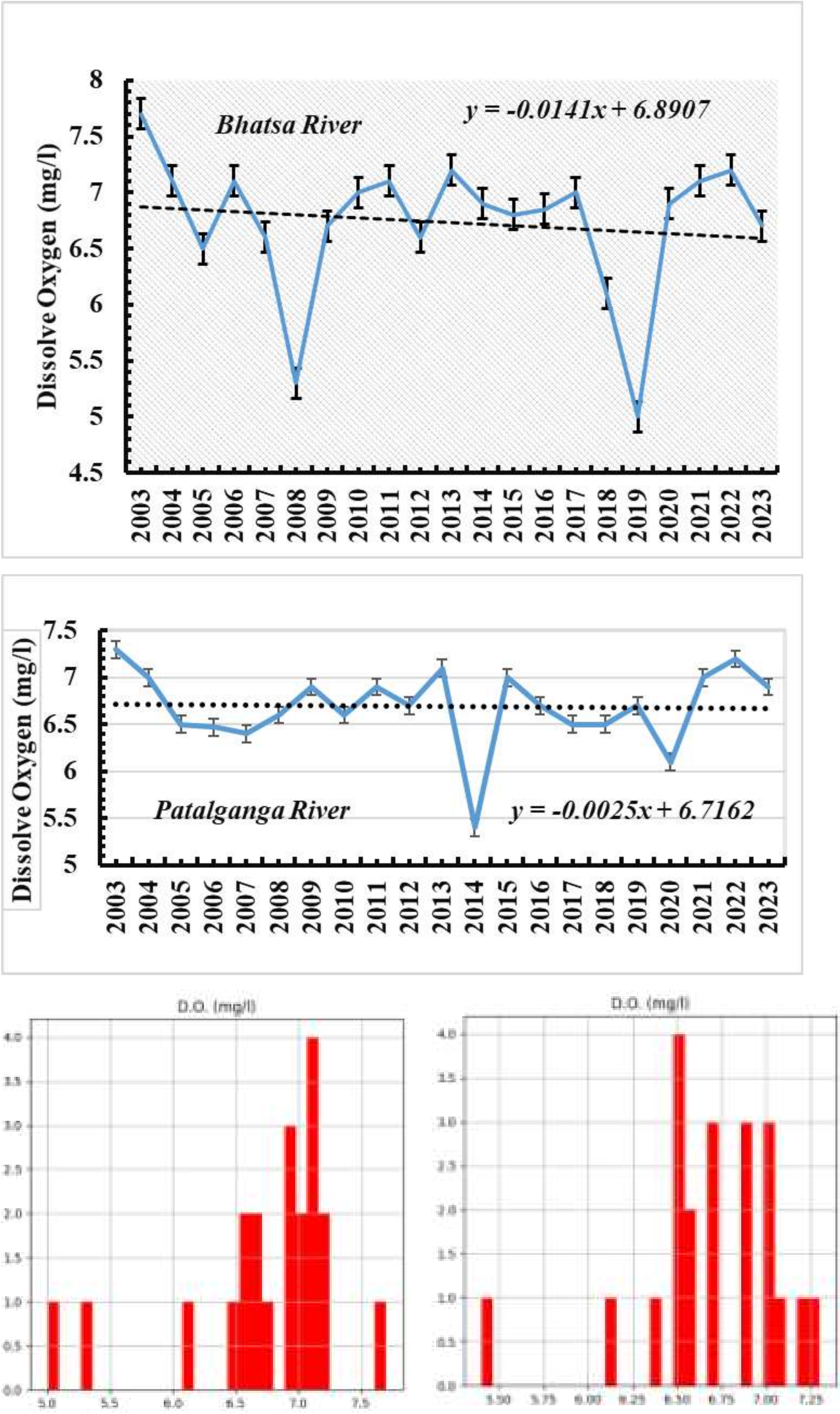




Bhatsa Patalganga

Fig. 1: Water temperature trends and frequency distributions

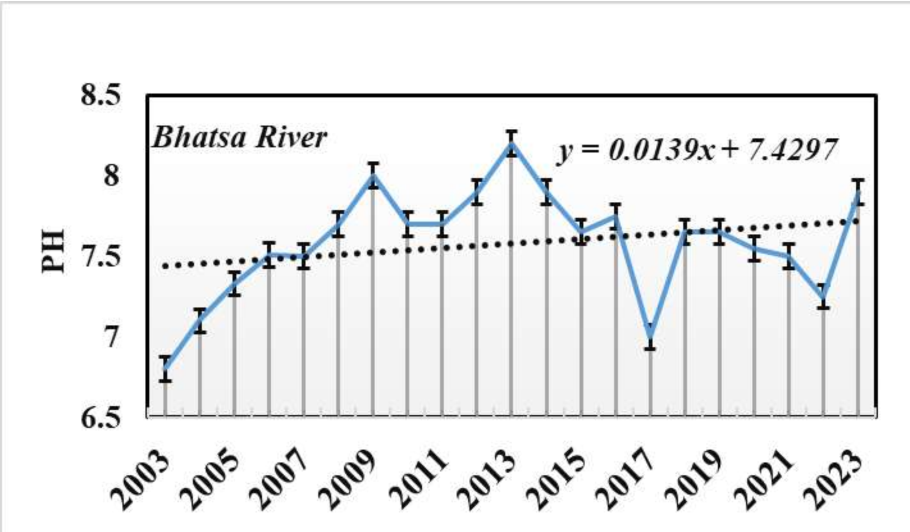
The maximum solubility of oxygen in river water depends on temperature, pressure, elevation, water salinity, degree of reaeration, and biochemical processes. In fig. 2, Dissolved Oxygen (DO) trend-lines and frequency distributions are shown for Bhatsa and Patalganga. The dip in DO (~5 mg/l) was found in 2019 and in 2008 for Bhatsa and in 2014 for Patalganga with respect to that of the DO of 7.5 mg/l in 2003. The maximum frequency distributions of DO occurred at 7.2 mg/l and 6.5 mg/l for Bhatsa and Patalganga respectively (Fig. 2). The authorities need to work towards reducing pollution sources that deplete DO of set standards for DO levels (e.g., 5 mg/l for rivers) for Patalganga in line with the measures taken for Bhatsa river.

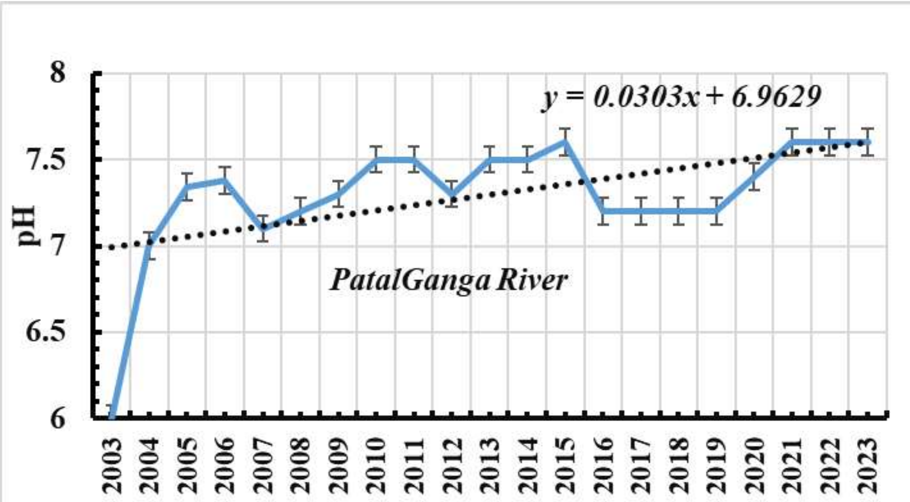


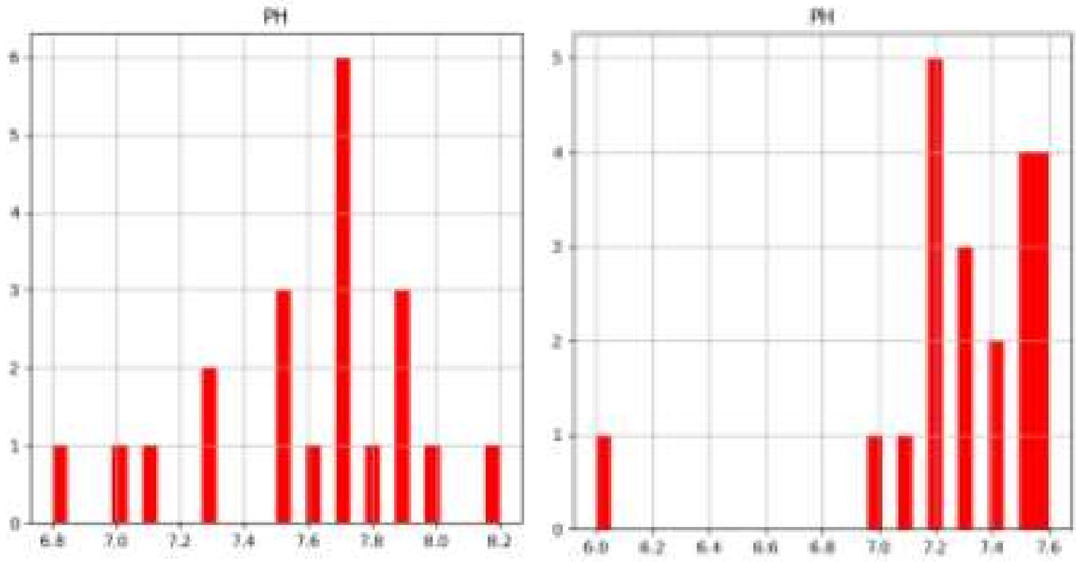
Bhatsa Patalganga

Fig. 2: Dissolved oxygen time-series and frequency distributions

The technical definition of river water pH can be a measure of the activity of the hydrogen ion (H^+). The time-series and frequency distributions of pH for Bhatsa and Patalganga rivers are shown in fig. 3. The minimum pH of 6.8 for Bhatsa and that of 6.0 for Patalganga in 2003 shown increasing pattern over a span of 20 years (2003 to 2023). The frequency distributions of pH showed maximum at 7.7 for Bhatsa and at 7.2 for Patalganga (Fig. 3). These pH levels are within the permissible pH range for river water in Maharashtra, as per the MPCB, which is generally between 6.5 and 8.5.



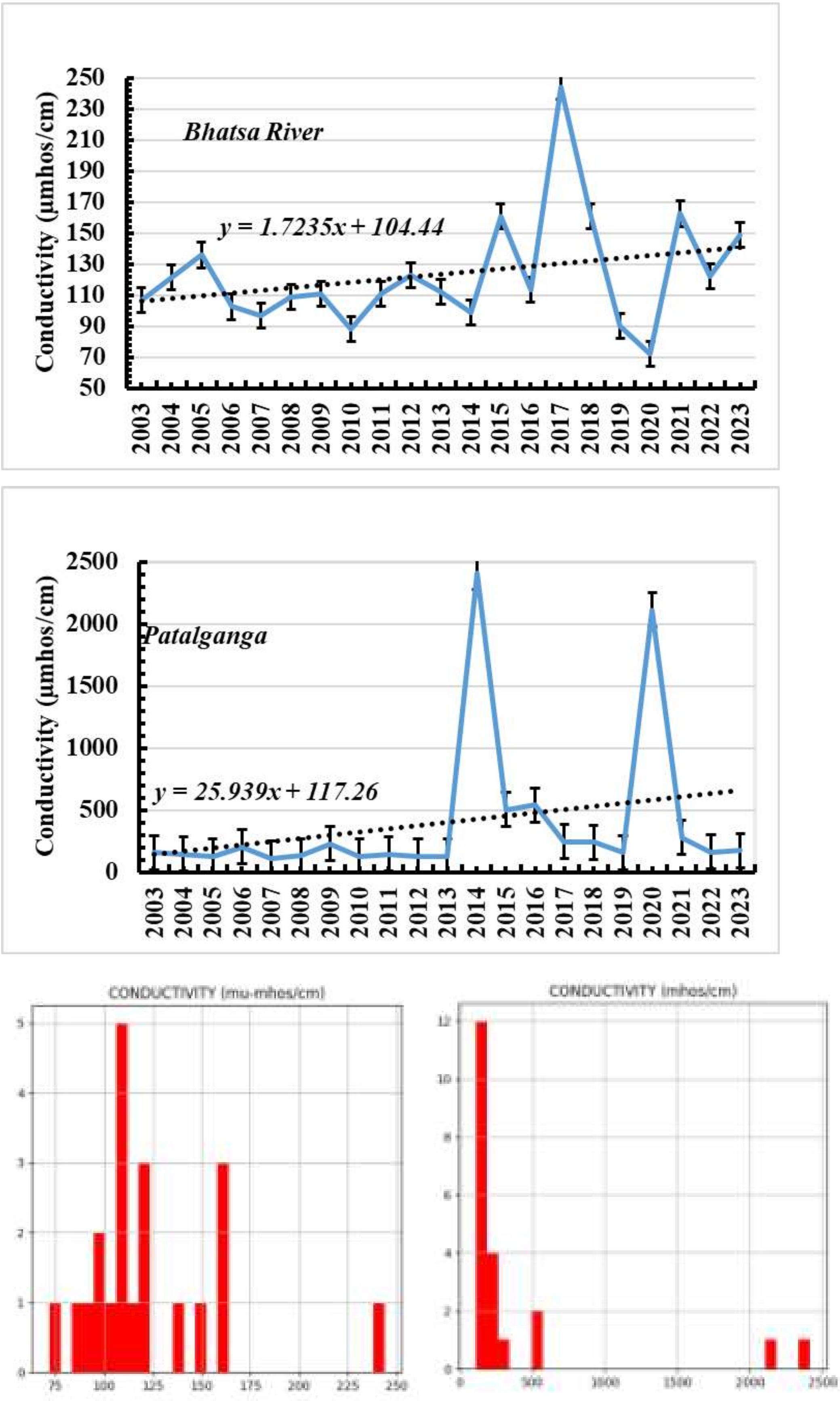




Bhatsa Patalganga

Fig. 3: Trends in pH values and frequency distributions

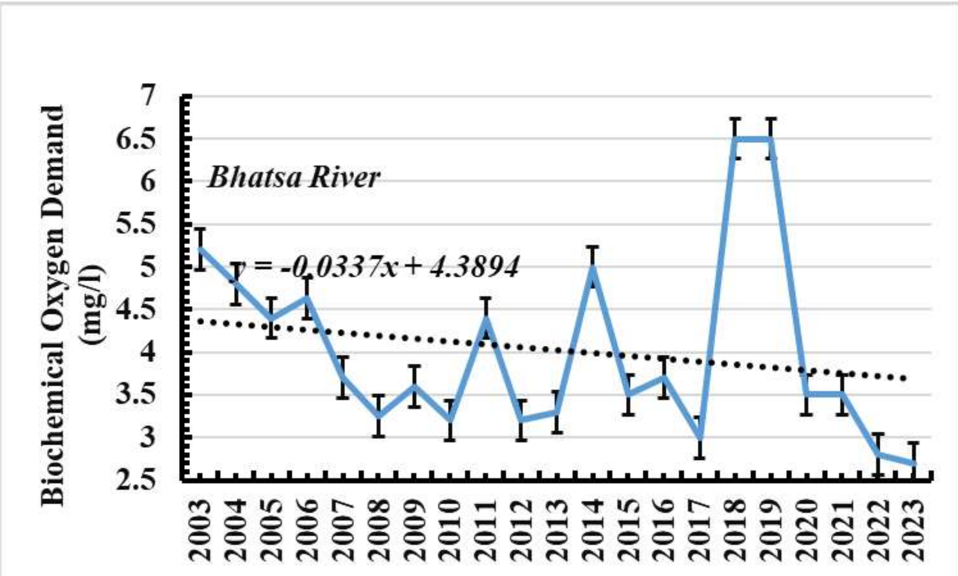
Nevertheless, the pH alone is not a measures the strength of acidic or alkaline water and that solely cannot be the characteristics or limitations of the potable water supply. Indeed, the water conductivity is generally about 100 times the total cations or anions expressed as their equivalents. Fig. 4 shows the water conductivity trends and frequency distributions for Batsa and Patalganga. The water conductivity of Bhatsa and Patalganga steadily increased over a period from 2003 to 2013 with peak (~250 µmhos/cm) in 2017 for Bhatsa and that of ~2500 µmhos/cm (tenfold to that of Bhatsa) for Patalganga in 2014. The frequency distribution of water conductivities shows highest (~110 µmhos/cm) for both Bhatsa and Patalganga (Fig. 4). There is no uniform conductivity range for river water set by the regulatory authorities, instead their main on ensuring judicious, equitable, and sustainable management, allocation, and utilization of water resources, including setting criteria for water distribution and usage

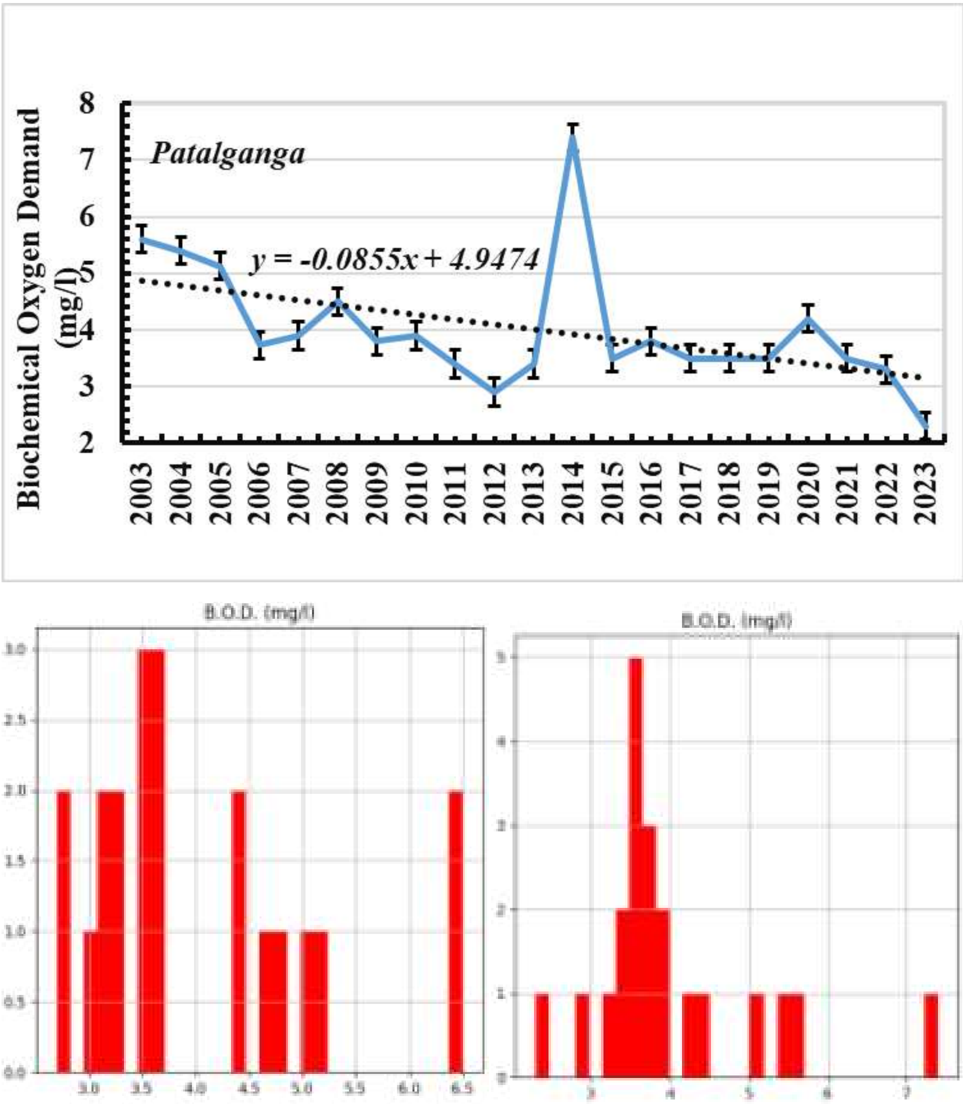


Bhatsa Patalganga

Fig. 4: Conductivity of river water trends and frequency distributions

Biochemical oxygen demand (BOD) is dissolved oxygen amount needed by aerobic biological organisms in river water to breakdown organic compound present in the water sample to CO_2 as a function of the biotic population, temperature, time, and constituents in the water. The time-series analysis and frequency distributions for BOD of water are shown in fig. 5 for Bhatsa and Patalganga over a period of 20 years (2003 to 2023). It is worth to note that BOD values of Bhatsa and Patalganga showed decreasing trends over span of 2003 to 2023, except a peak for Bhatsa at 6.5 mg/l during 2018 to 2019 and that of BOD at 7.5 mg/l in 2014 for Patalganga (Fig. 5). The maximum frequency distribution of BOD is identified at 3.5 mg/l. The BOD relates to the readily available amount of decomposable organic compound which indicates values within the permissible range (Fig. 5). In the controlled conditions, BOD is the oxygen demand over a 5-day period when the sample is incubated at 20°C by excluding nitrification which is the next variable of the discussion.

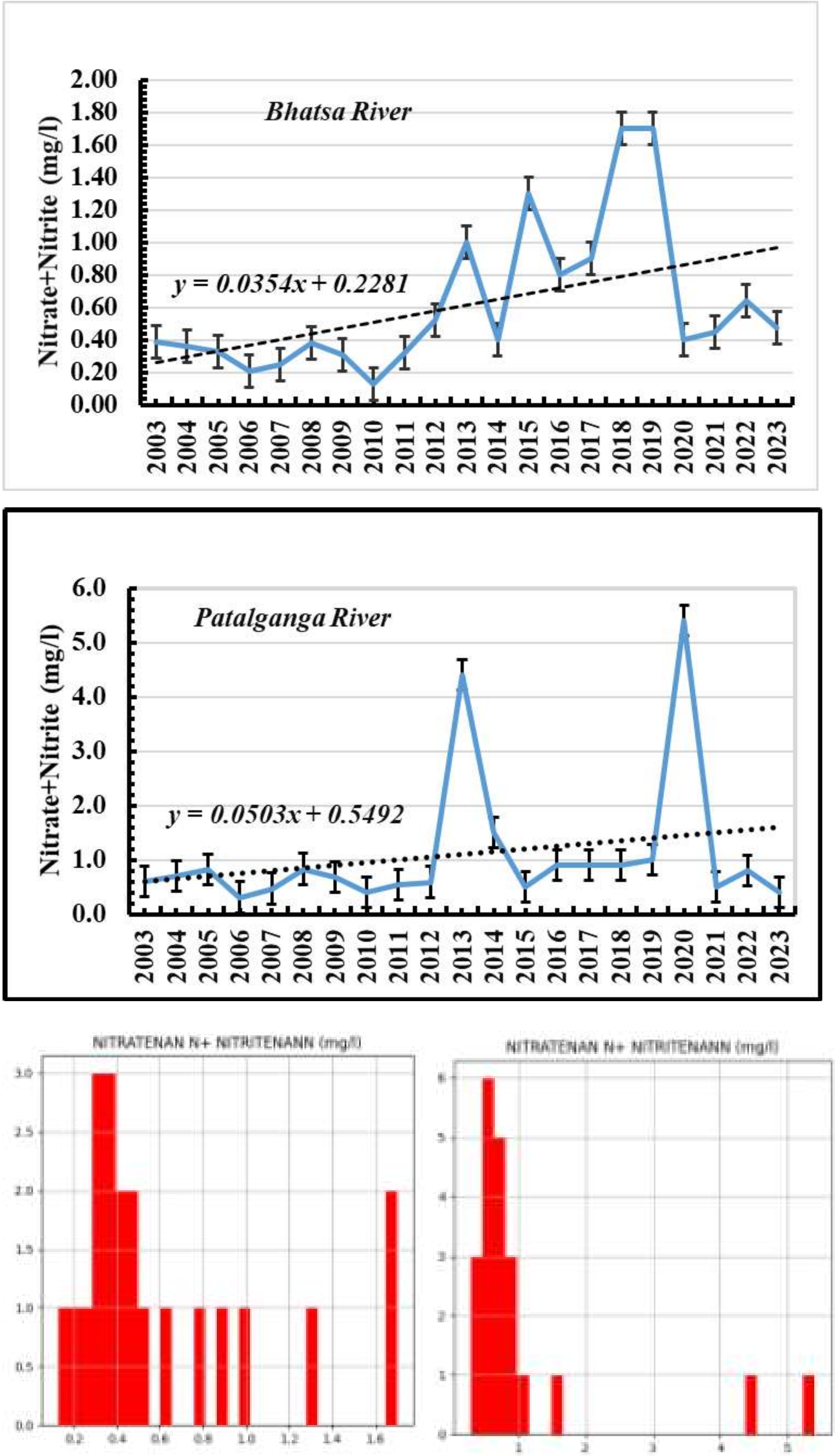




Bhatsa Patalganga

Fig. 5: Biochemical Oxygen Demand trends and frequency distributions

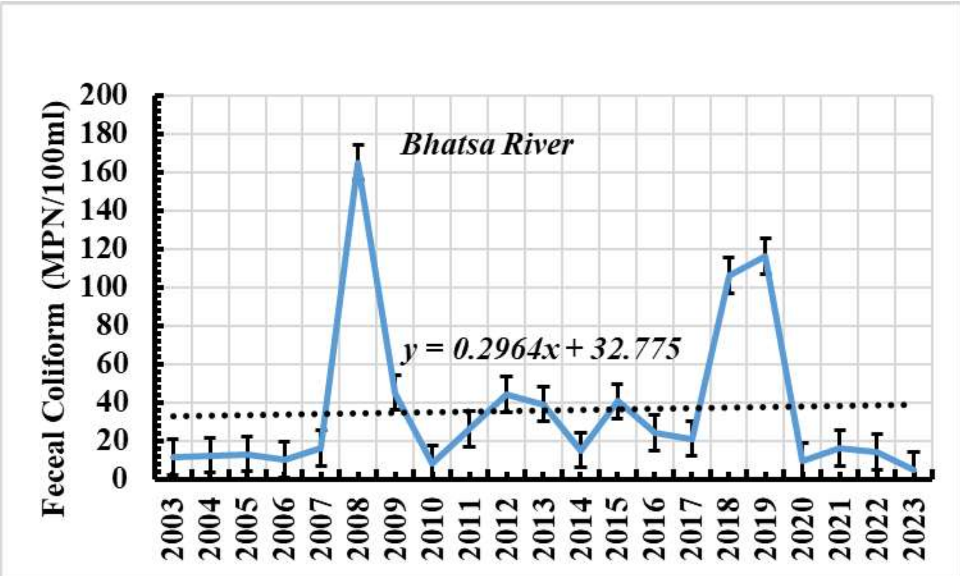
In order to know the nitrification of Bhatsa and Patalganga water, time-series trends and frequency distributions of nitrate + nitrite are shown in fig. 6. The nitrate +nitrite values remained within 1 mg/l during 2003 to 2012 with first peak at 1.7 mg/l during 2018-2019 and next peak at 4.5 mg/l in 2013 for Bhatsa and for Patalganga at.5 mg/l in 2020. The maximum frequency distributions of nitrate + nitrite respectively occurred at 0.35 and at 0.5 mg/l within the range of permissible limit for Bhatsa and Patalganga (Fig. 6).

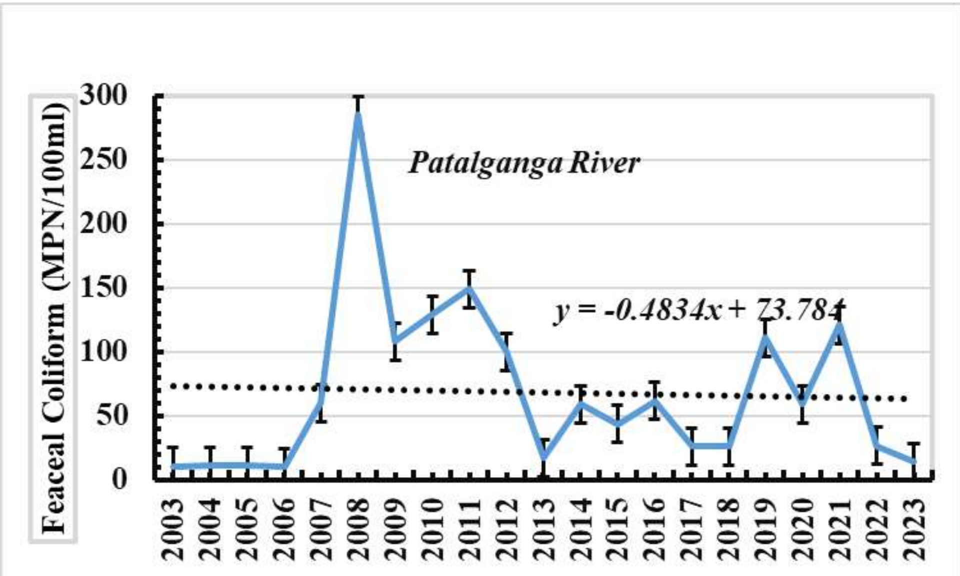


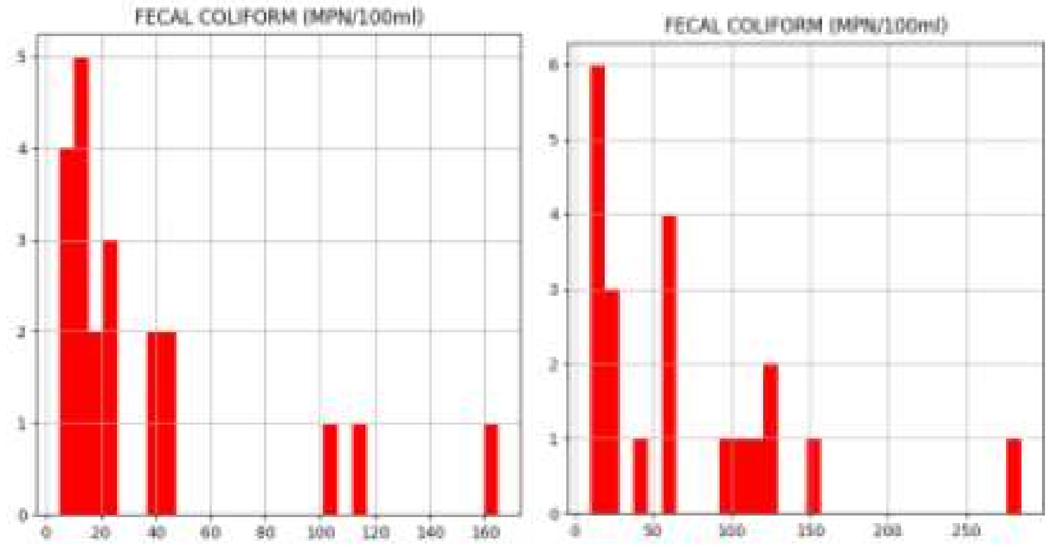
Bhatsa Patalganga

Fig. 6: Nitrate + Nitrite trends and frequency distributions

The fecal coliform group includes all of the rod-shaped bacteria that are non-spore forming, *Gram-Negative,* lactose-fermenting in 24 hours at 44.5 °C (controlled conditions). The fecal coliform time-series and frequency distributions are shown in fig. 7 for Bhatsa and Patalganga. The annual variability of fecal coliform showed steady values within 50 MPN/100ml. The noticeable peak at 165 MPN/100ml seen in 2008 and next peak at 106-116 MPN/100ml during 2018-2019 for Bhatsa. For Patalganga, first peak is realized at 285 MPN/100ml in 2008 and second one at 150MPN/100ml in 2011. The frequency distribution of fecal coliform is maximised respectively at 10 MPN/100ml and at 20 MPN/100ml for Bhatsa and Patalganga (Fig. 7). Generally, fecal coliform can grow with or without oxygen and are typically associated with warm-blooded animals. Total coliform needs to be considered next.



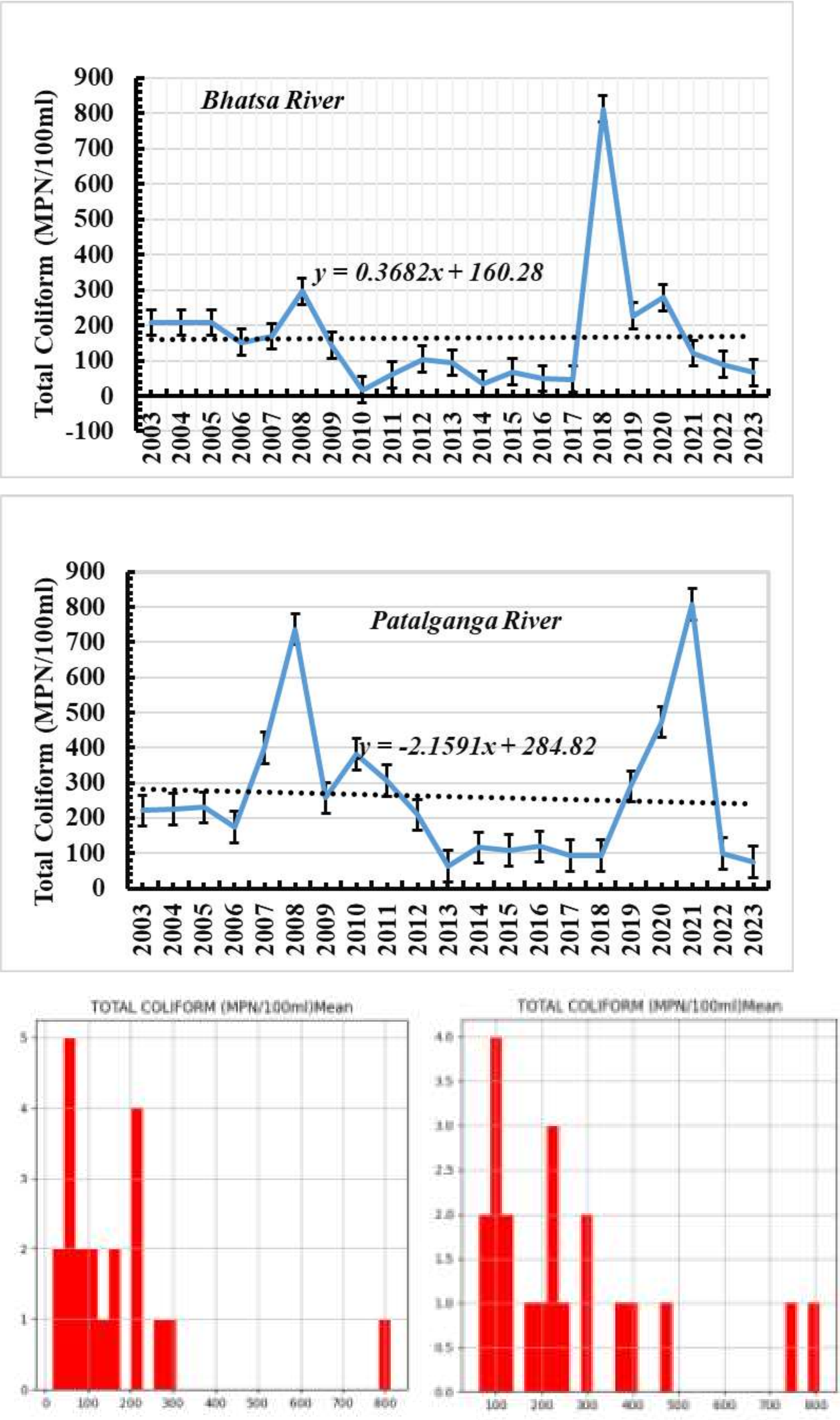




Bhatsa Patalganga

Fig. 7: Fecal Coliform trends and frequency distributions

The total coliforms including bacteria established in the environment, soil, and insects and the bacteria can ferment the lactose at 37 °C. The total coliform patterns and its frequency distributions are plotted in fig. 8 for Bhatsa and Patalganga. For Bhatsa, total coliform remained at its steady level (100-200 MPN/100ml) during 2003 to 2023 with first peak at 300 MPN/100ml in 2008 and second highest concentration at 812 MPN/100ml during 2020-21. Similarly, for Patalganga, total coliform showed slightly declining trend during 2003 to 2023 with a first peak at 737 MPN/100ml in 2008 and second highest one at 808 MPN/100ml in 2021. The maximum frequency distribution of total coliform found to be at 80 MPN/100ml and at 100 MPN/100ml for Bhatsa and Patalganga respectively (Fig. 8). The total coliform range for potable water in rivers is generally less than 500 MPN/100 ml for desirable quality.



Bhatsa Patalganga

Fig. 8: Total Coliform trends and frequency distributions

To highlight aforementioned results further, trend-line representing variabilities in water quality datasets over the period of 20 years (2003-23) for water temperature, Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), pH, conductivity, Nitrate + Nitrite, Fecal coliform and total coliform are tabulated in Table 1 for Bhatsa and Patalganga rivers. These expressions are useful for variables to be anticipated in short range for water quality decision making.

Table 1: Time-series trend-line equations for variability of water quality datasets

S. N.	Parameters	Bhatsa river	Patalganga river
1	Water Temperature	$y = 0.13x + 26.3$	$y = 0.15x + 25.6$
2	Dissolved Oxygen (DO)	$y = -0.014x + 6.9$	$y = -0.0025x + 6.7$
3	Biochemical Oxygen Demand (BOD)	$y= -0.0034x+4.4$	$y =-0.085x+4.95$
4	pH	$y = 0.014x + 7.43$	$y = 0.03x + 7.0$
5	Conductivity	$y = 1.7x + 104$	$y = 25.9x + 117$
6	Nitrate + Nitrite	$y = -0.034x + 4.4$	$y =- 0.085x + 4.9$
7	Fecal coliform	$y = 0.3x + 32.8$	$y = -0.5x + 73.8$
8	Total coliform	$y = 0.37x + 160.3$	$y = -2.16x + 284.8$

3.2 Correlation matrices and AI-ML model performers:

The correlation matrices identify the best coefficients of determination among the water quality indicators. The heatmaps of correlation matrices of water quality parameters in fig. 9 and fig. 10 are respectively for Bhatsa and Patalganga. The coefficients of determination >0.5 among the water quality parameters generally considered to be best predictors for water quality of these distinctive rivers. As visualized in heatmap of Fig. 9, BOD and total coliform shows better correlation with coefficient of 0.6 for Bhatsa river water. The fecal coliform and total coliform indicates relatively good correlation of 0.54 than that of 0.51 between Nitrate + Nitrite and fecal coliform for Bhatsa. Based on the coefficients of determinations, BOD, total coliform, fecal coliform and Nitrate + Nitrite are found the best water quality predictors for Bhatsa (Fig. 9). As seen in heatmap of fig. 10, a very good correlation of 0.72 is found between fecal coliform and total coliform for Patalganga and of 0.55 between BOD and conductivity. Furthermore, a good correlation of 0.54 shows between conductivity and Nitrate + Nitrite for Patalganga (Fig. 10). These results suggest that fecal coliform, total coliform, BOD, conductivity, Nitrate + Nitrite are the best water quality indicators for predictability of these river water quality.

Fig. 9: Correlation Matrix with water quality parameters database of Bhatsa

Fig. 10: Correlation Matrix with water quality parameters database of Patalganga

3.3 XGBoost modelling system performers:

The Root Mean Squared Error (RMSE), Mean Squared Error (MSE) and Mean Absolute Error (MAE) are vital evaluation metrics for regression models, each determining unique insights into prediction accuracy. The RMSE, MSE and MAE statistical scores of XGBoost Modelling system for Bhatsa and Patalganga water quality databases are listed in Table 2.

Table 2: AI-ML Model system performers with water quality databases

S. N.	Model Performers	Bhastsa	Patalganga
1	RMSE	0.736	0.185
2	MSE	0.5415951609611511	0.03424854949116707
3	MAE	0.4995839595794678	0.14479629695415497

The RMSE is the square root of MSE and has the same units as the target variable. The MSE is the average of squared differences. As seen from Table 2, since errors are very small (< 1), MSE is less than 1. The MAE is the average of absolute differences between predicted and actual values. The average prediction error is less than 1 unit of the target variable leads to MAE less than 1 (Table 2). Having RMSE, MSE, or MAE less than 1 means XGBoost model's average prediction error <1 is generally a sign of a good fit, especially the target variable's scale supports such smaller errors < 1 (Table 2).

3.4 Case Study Limitations

Implementation of such AI-ML based technique for real-time monitoring and forecasting of river water quality well in advance is subject to the calibration and validation of modelling system with the datasets of regulatory grade monitoring network beyond 2023 to till date. Although MPCB monitors the river water quality, non-availability of water quality data on seasonal or monthly scales till date from regulatory grade monitoring network constraints the innovative approach like training and testing of relevant AI-ML modelling systems which can provide ensemble real-time monitoring and forecasted information. Such forecast information is crucial for water quality management and policy-induced preventive measures to be enforced for potable water scarcity regions.

Conclusions

The water quality databases over a period of 20 years (2003-2023) for distinctive river Bhatsa and Patalganga are obtained from the central pollution control board. The water temperature, DO, BOD, pH, conductivity, nitrate + nitrite, fecal coliform and total coliform trend and their frequency distributions leading to better understandings of the variability in water quality datasets of Bhatsa and Patalganga during 2003 to 2023. The water temperature showed increasing trends for Bhatsa and Patalganga on account of rapid urbanizations with industrial growth in Maharashtra Industrial Development Corporation zone. The DO, BOD, pH, conductivity, nitrate + nitrite, fecal coliform and total coliform variability followed the water pollution trends and frequency distributions indicate the decision making water quality management by data-driven case study of Bhatsa and Patalganga. The processing of the temporal water quality databases with proven AI-ML modelling system (XGBoost) with outcomes as frequency distributions for water quality parameters (temperature, DO, BOD, pH, conductivity, nitrate + nitrite, fecal coliform and total coliform), coefficients of determination and model performers for Bhatsa and Patalganga can be a complementary approach in line with the ongoing national innovative water quality management program.

The best coefficients of determination among the fecal coliform, total coliform, BOD, conductivity, Nitrate + Nitrite are indicators for the predictability of the water quality. The AI-ML modelling tools to be used for the water quality management of major and minor rivers all over India. The dynamic water quality predictions of river using AI-ML modelling with the availability of high temporal water quality databases may be future scope of research. The main challenge is dealing with water-related data as they are problematic to handle owing to their nonlinearity, nonstationary feature and vague properties due to the unpredictable natural changes, interdependent relationship, human interference and complexity. Traditional, deep learning and hybrid modelling systems can handle such data owing to their higher accuracy to deal with non-linear data, robustness, reliability, cost-effectiveness, problem-solving capability, decision-making capability, efficiency and effectiveness. These models are the perfect tools for river water quality monitoring, management, sustainability and policymaking.

Authors' Contributions:

Concept, data compiling, illustrations, writing, etc.

Conflict of Interest

No conflicts of interest

Acknowledgement:

Author thanks the CPCB for water quality parameters data. Thanks to Aditya, CTO, SMIT Enviro Solutions (https://www.smitenvirosolutions.in) for developing computing infrastructures (codes, computing libraries, running the regressions and XGBoost) in cloud environments (paid services). Author acknowledges Director, SMIT Enviro Solutions for allowing the usage of cloud computing resources and facilities within its office premises.

References

1. S. Jagtap and M. R Tikam, Water quality assessment of river PATALGANAGA Intn. J. Adv. Res. & Innov. Ideas in Edu., 5 (1), 203-210, 2019.
2. R. Kumar, I. Gupta, S. Sawant, K. Shah, V. Singh, V.P. Chaudhari, J. K. Siddhu, R. Irani, Water Quality Status of Water Bodies of Maharashtra with Recourse to Analytical/Statistical Tools (2007-2011), MPCB, CSIR- National Environmental Engineering Research Institute (NEERI), pp 749, 2014.
3. A. Haarstrick, L. Sharma Urban river pollution control. Managing urban rivers. Elsevier, Amsterdam, pp 131–159, 2024.
4. A. Bhalerao and S. Tawde, Baseline study of water and sediment of Bhatsa river (Vasind, MS), Journal of Emerging Technologies and Innovative Research, 6(5), 132-137, 2019
5. *P. Pazare and R. Dani,* Patalganga River: Plight of The Pollution, EARTH5R (https://earth5r.org/patalganga-river-plight-of-the-pollution/), accessed on 12th May 2025.
6. A. Munde, P. Jankar, K. Patil, H. Patil, P. Patil, H. Dalvi and S. Patil, Int. Sci. J. Eng. Mgmt., 4(4), 1-4, 2025.
7. Polluted River Stretch Priority-III (Patalganga River, State –Maharashtra (https://www.mpcb.gov.in/sites/default/files/water-quality/River_action_plans/priority_III_PATALGANGA_2019_08112019.pdf), accessed on 13th May 2025.
8. T. Chen, C. Guestrin,, XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August, 2016, pp. 785–794, 2016.
9. J. Dong, W. Zeng, L.Wu, J. Huang, T. Gaiser, A. K. Srivastava, Enhancing Short-Term Forecasting of Daily Precipitation Using Numerical Weather Prediction Bias Correcting with XGBoost in Different Regions of China. Eng. Appl. Artif. Intell., 117, 105579, 2023