DOI: 10.24906/isc/2023/v37/i4/43716

Search And Survey

Investigating Significance of Biochemical Parameters on Surface Water Acidity/Basicity in Agra, India

Soumyadeep Poddar[1] and Nilabhra Rohan Das[2]

Abstract

The study aims to investigate the significancy of different biochemical factors which affect surface water quality of Agra, India. The study reveals how various factors like pH, dissolved oxygen (DO), and chemical composition affect surface water quality. Data from 2007-21 was analysed using a multiple linear regression model to understand how these different factors relate to pH. The analysis revealed that Chloride, Fluoride, Dissolved Oxygen, Bicarbonates, Sulphates, Magnesium, and Sodium make significant contributions to determining pH of surface water of Agra. The existence of a non-linear relationship may be explored in future studies.

Keywords: Agra, Biochemical components, Multiple linear regression, pH, Surface water quality.

Introduction

Water stands as a highly invaluable natural resource on a national scale, serving as an essential foundation for human civilization, the sustenance of living organisms, and the nurturing of natural habitats. Water plays a pivotal role in every facet of evolutionary progress, propelling economic advancement, bolstering the well-being of ecosystems, and constituting a vital, fundamental element of existence. Among various purposes, water serves as a crucial resource for drinking, sanitation, agriculture, industrial activities, recreational pursuits, livestock husbandry, and electricity generation. Nevertheless, it's worth noting that merely a fraction of the Earth's total water resources, less than 3%, is freshwater, with a scant 2% of that being surface water. Consequently, only a minuscule portion of freshwater is actually accessible for human use [1].

Water represents a renewable resource, and India boasts a wealth of water bodies, including rivers, lakes, ponds, reservoirs, and more. Despite its initial status as a water-rich nation, India is gradually transitioning into a state of water scarcity. This shift is primarily attributed to the mounting pressure from a growing population and the rapid pace of urbanization.Approximately 18% of the global population is confined within the country's borders, while the available water resources within those borders account for only around 4% of the world's total water resources. Rivers have traditionally served as the lifeblood

[1]30/6/A1, Barikpara Road, Kolkata 700034, soumyadeeppoddar72@gmail.com
ORCID: Soumyadeep Poddar: https://orcid.org/0009-0005-7258-3182
[2]19 Raj Krishna Pal Lane, Kolkata 700075, nr.das@yahoo.com
ORCID: Nilabhra Rohan Das: https://orcid.org/0000-0001-8187-0080

of the nation's development and culture.Among more than 400 rivers [2], 12 are categorized as major, collectively covering an expansive catchment area of approximately 253 million hectares. Additionally, there are 46 medium-sized rivers, with a combined catchment area of around 24.6 million hectares [3]. The majority of India's river systems, along with their tributaries, flow perennially, while others are seasonal. The Ganga-Brahmaputra-Meghna River basin stands as the most extensive river system in India, encompassing a significant 43% of the total catchment area of all major river systems. Notable among the other major river systems are the Indus, Sabarmati, Mahi, Narmada, Tapi, Brahmani, Mahanadi, Godavari, Krishna, Pennar, and Cauvery.Within the medium-classified river systems, the Subarnarekha, boasting a catchment area of 1.9 million hectares, holds the distinction of being the largest.

Given such an abundance of surface water bodies, the question of usability arises. Many do not understand that the health problems caused by drinking water can be attributed to the pH of the water. Many individuals may not recognize that their health issues could be linked to the pH level of the water they consume. The pH of water is determined by the concentration of hydrogen ions within it, subsequently dictating its acidity. Unregulated acidity can lead to various complications. A low pH can transform water into a harmful solution, making it more prone to attracting heavy metals within the human body compared to neutral water. This study considers this pH or potential of hydrogen as an acclaimed markerof water related health problems [4].

In recent years, pHhas become an important factor that comes under analysis when considering the usability ofa water body. Water is regarded as a neutral liquid (pH 7), neither acidic nor basic. However, the presence of organic matter and inorganic impurities makes this quite impossible.As an example, water discharged from an abandoned coal mine may exhibit an extremely low pH level of 2, signifying a highly acidic nature. Such acidity would undoubtedly have adverse effects on any fish attempting to inhabit it [5]. Traditional literature states that the acidity or basicity of water is dependent on the presence or absence of inorganic and organic compounds, microorganisms, etc. among others. If water becomes too acidic or too basic, it becomes unfit for use universally unless treated properly. Extremely low or high pH values can be damaging to the use of water. Elevated pH levels in drinking water can impart a bitter taste to it [6][7][8]. Water supplies, pipelines, and various appliances can accumulate deposits when exposed to high pH levels. Furthermore, a high pH can impede the effectiveness of chlorine as a disinfectant. Elevated pH levels not only pose direct harm but also contribute to the corrosion and dissolution of metals in water, which can elevate toxicity levels. Conversely, low pH levels can encourage the dissolution of heavy metals in water, leading to an increase in water toxicity as the concentration of heavy metals rises [9]. An estimated 70% of India's surface water is presently unsuitable for consumption. Each day, nearly 40 million liters of wastewater are discharged into rivers and other water bodies, with only a minimal portion undergoing proper treatment. Other than wastewater, natural occurrences of limestone quarries, coal fields, and acid rain among others, contribute significant amounts of sulphates, carbonates, and bicarbonates to water bodies.Before classifying such chemical compounds as pollutants, assessment of present conditions must be carried out with due diligence. In such a situation, a reinvestigation of the traditional knowledge is required to properly access the present impact of different biochemical parameters on the acidity/basicity of surface water. The objective of this study is to reestablish the claims of historical literature or establish new relationships between the biochemical parameters under analysis and their effect on pH [10].

Among the many metropolitans under development in India, Agra is notable for its internationality. The Taj Mahal is one of the seven wonders of the world and attracts tourists from all over the world. As the city develops into an urban metropolitanwith a thriving tourism industry has led to the swift development of Agra and its suburbs, its water bodies are getting toxic at a very fast pace providing a practical site for carrying out the analysisthrough collected data [11].

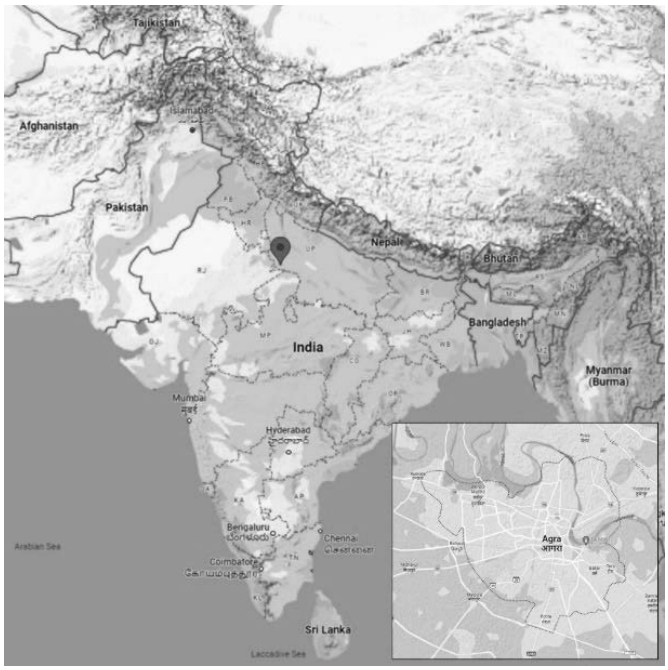

Figure 1: Map of Agra, India.

Methods

The variables (parameters) being considered for the investigation are power of hydrogen (pH), total alkalinity, biochemical oxygen demand (BOD), Chloride, Fluoride, dissolved oxygen (DO), carbonates, bicarbonates, sulphates, boron, calcium, iron, potassium, magnesium, and sodium (Table 1).

Linear regression is a method used to model the connection between one or more independent variables and a single numerical outcome in a linear manner (commonly known as independent and dependent variables respectively). When dealing with a single explanatory variable, it is referred to as simple linear regression. However, when there are multiple independent variables involved, the technique is known as multiple linear regression.

Multiple linear regression (MLR), often simply referred to as multiple regression, is a supervised machine learning method that leverages multiple explanatory variables (represented as x_i) to forecast the outcome of a response variable (y). The primary objective of multiple linear regression is to establish a linear model that characterizes the relationship between the independent (explanatory) variables and the dependent (response) variable. For this investigation, pH is the response or dependent variable, and all the other parameters are the explanatory or independent variables.In essence, multiple regression represents an extension of ordinary least-squares (OLS) regression because it encompasses the use of more than one explanatory variable [12].

Table 1. Biochemical parameters under consideration.

Variable	Parameter
y	pH
x_1	Total Alkalinity
x_2	Biochemical Oxygen Demand (BOD)
x_3	Chloride
x_4	Fluoride
x_5	Dissolved Oxygen (DO)
x_6	Carbonates
x_7	Bicarbonates
x_8	Sulphates
x_9	Boron
x_{10}	Calcium
x_{11}	Iron
x_{12}	Potassium
x_{13}	Magnesium
x_{14}	Sodium
x_{15}	Nitrogen

For this study, the multiple regression model is,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{15} x_{15} + \varepsilon,$$

y = *dependent (response) variable,*

x_i = *independent (explanatory) variables,* i =1(1)15,

β_0 = *intercept (constant),*

β_i = *slope coefficients,* i = 1(1)15,

ε = *error term (residuals),*

and Multiple linear regression relies on several assumptions, including the presence of a linear relationship between the dependent and independent variables, the absence of excessive correlation among the independent variables, the random and independent selection of y observations from the population, and the

requirement that the residuals (the differences between predicted and actual values) exhibit a normal distribution with a mean of 0. These assumptions form the foundation of the multiple linear regression model.

The coefficient of determination (R^2) is a statistical measure employed to assess the extent to which the fluctuations in the dependent variable can be accounted for by the fluctuations in the independent variables. R^2 consistently rises when additional predictors are incorporated into the multiple linear regression (MLR) model, even though some of these predictors might not necessarily be associated with the outcome variable. This is a characteristic of R^2 as it evaluates the overall explanatory power of the model. So, R^2 cannot be used to identify the inclusion or exclusion of predictors in a model all by itself.Every time a new predictor variable is added to the model, R^2 is almost always guaranteed to increase even when that variable is useless. The adjusted R^2 is a modified version, that accounts for the number of predictors in a regression model and is defined as $1-\frac{(1-R^2)(n-1)}{n-k-1}$ where n is the number of observations and k is the number of predictor variables. Since R^2 always increases as more predictors are added to a model, adjusted R^2 can serve as a metric that expresses how useful a model is, adjusted for the number of predictors in a model, that is, the percentage of variation explained by only the independent variables that actually affect the dependent variable. Similar to R^2, the value of adjusted R^2 lies between 0 and 1 with high values closer to 1 indicating that all the predictor variables in the model have significant effects. For this investigation, n is 485 and k is 15. So, the above mathematical expression may be re-written as $1-\frac{(1-R^2)\times 484}{469}$ for this particular study.

A drawback of the MLR model is the occurrence of multicollinearity. Multicollinearity is a statistical term that arises when two or more independent (predictor) variables within a model exhibit strong correlations, meaning that one of these independent variables $x_1, x_2, x_3, \ldots, x_{15}$ can be mathematically represented as a linear combination of the others.

Considering our linear model as stated above, multicollinearity exists if,

$$x_1 = \lambda_0 + \lambda_1 x_2 + \lambda_2 x_3 + \cdots + \lambda_{14} x_{15} + \varepsilon, \textit{or}$$

$$x_2 = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_3 + \cdots + \lambda_{14} x_{15} + \varepsilon, \textit{or}$$

$$x_3 = \lambda_0 + \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_{14} x_{15} + \varepsilon,$$

and so on.

Multicollinearity does not impact the accuracy of the predictors or the goodness-of-fit statistic (R^2). However, it becomes problematic when attempting to make inferences about the independent variables, as is the case here. Due to multicollinearity, the coefficients in the multiple linear regression (MLR) model lose their meaningful interpretation. The model is incapable of discerning the individual influence of each predictor (x_i) variable. Additionally, the p-values, which signify the statistical significance of the independent variables, may become less reliable.

The method used for testing multicollinearity is by looking at the Variance Inflation Factor (VIF) values. The Variance Inflation Factor (VIF) quantifies the increase in the coefficient of an independent variable caused by collinearity with other independent variables. A VIF of 1 signifies that the regression coefficient remains unaffected by the presence of other predictors, indicating the absence of multicollinearity in the supervised learning model.As a general guideline, when VIF values surpass 5, it is advisable to conduct a more thorough examination of the multiple linear regression (MLR) model. At this point, multicollinearity is typically present but may not necessarily demand immediate attention. However, when VIFs exceed 10, this indicates a severe degree of multicollinearity, and the coefficient estimates and p-values in the regression results are likely to be unreliable. In an optimal scenario, VIF values are generally kept below 3.

Multicollinearity in a MLR model needs to be dealt with if the significance of the regression coefficients are objects of interest. For this study, the variables exhibiting multicollinearity will be removed from the model to get rid of the multicollinearity in the model.

Lastly, in the realm of multiple linear regression (MLR), added variable plots, often

as a visual interface, serve to explore the connection between a dependent variable and a single independent variable. Importantly, this exploration takes place while keeping other independent variables in the model constant. Such plots enable us to scrutinize the relationship between each specific independent variable and the dependent variable within the context of the other independent variables' influence.

Results

The purpose of this investigation is to test the hypothesis $H_0: \beta_i = 0$ *vs* $H_1: \beta_i \neq 0$, $(i = 1,2,3,...,15)$. The null hypothesis posits that there is no connection between the dependent variable and the independent variables, implying that all coefficients in the model are zero. Conversely, the alternative hypothesis asserts the presence of a relationship, signifying that at least one coefficient is non-zero. In simpler terms, the null hypothesis assumes no impact, while the alternative hypothesis suggests that there is an impact. In practical terms, when the p-value falls below 0.05, it typically indicates the existence of at least one non-zero coefficient in the model.

Table 2. Summary of the multiple linear regression model.

Variable	Estimate	p-value
Intercept	7.448	<0.01
Total Alkalinity (x_1)	-7.32e-05	0.76
Biochemical Oxygen Demand (BOD) (x_2)	-4.14e-04	0.78
Chloride (x_3)	1.11e-03	0.03
Fluoride (x_4)	1.7e-01	0.096
Dissolved Oxygen (DO) (x_5)	1.99e-02	0.005
Carbonates (x_6)	1.56e-03	0.11
Bicarbonates (x_7)	-1.96e-03	<0.001
Sulphates (x_8)	2.5e-03	0.04
Boron (x_9)	-1.66e-02	0.34
Calcium (x_{10})	1.55e-03	0.47
Iron (x_{11})	1.08e-01	0.32
Potassium (x_{12})	6.63e-04	0.55
Magnesium (x_{13})	1.02e-02	0.001
Sodium (x_{14})	1.09e-03	0.03
Nitrogen (x_{15})	-6.37e-03	0.08

From the analysis, the formulated linear model is,

$$\begin{aligned} y = 7.448 &+ (-7.317e-05)x_1 + (-4.146e-04)x_2 + (1.113e-03)x_3 \\ &+ (1.699e-01)x_4 + (1.989e-02)x_5 + (1.555e-02)x_6 \\ &+ (-1.963e-03)x_7 + (2.503e-03)x_8 + (-1.662e-04)x_9 \\ &+ (1.546e-03)x_{10} + (1.082e-01)x_{11} + (6.634e-04)x_{12} \\ &+ (1.020e-02)x_{13} + (1.089e-02)x_{14} + (-6.365e-03)x_{15} \end{aligned}$$

The model may be re-written as,

$$\begin{aligned} pH = 7.448 &+ (-7.317e-05)\ Total\ Alkalinity + (-4.146e-04)BOD \\ &+ (1.113e-03)\ Chloride + (1.699e-01)\ Fluoride + (1.989e-02)DO \\ &+ (1.555e-03)\ Carbonates + (-1.963e-03)\ Bicarbonates \\ &+ (2.503e-03)\ Sulphates + (-1.662e-02)\ Boron \\ &+ (1.546e-03)\ Calcium + (1.082e-01)\ Iron \\ &+ (6.634e-04)\ Potassium + (1.020e-02)\ Magmesium \\ &+ (1.089e-03)\ Sodium + (-6.365e-03)\ Nitrogen \end{aligned}$$

From the analysis and calculations, based on the data, there is enough evidence against the null hypothesis H_0 for the variables $x_3, x_5, x_7, x_8, x_{13}$ *and* x_{14} while for the rest of the variables, there is not enough evidence (Table 2).

Table 3. VIF values of the parameters.

Variable	Parameter	VIF
x_1	Total Alkalinity	1.75405
x_2	Biochemical Oxygen Demand (BOD)	1.103432
x_3	Chloride	2.932253

Variable	Parameter	VIF
x_4	Fluoride	1.140776
x_5	Dissolved Oxygen (DO)	1.100339
x_6	Carbonates	1.312032
x_7	Bicarbonates	3.112188
x_8	Sulphates	2.023856
x_9	Boron	1.031561
x_{10}	Calcium	1.816975
x_{11}	Iron	1.057045
x_{12}	Potassium	1.113096
x_{13}	Magnesium	2.799547
x_{14}	Sodium	2.201614
x_{15}	Nitrogen	1.097929

Furthermore, the dataset does not contain any multicollinearity between the independent variables as all the VIF values are quite small (Table 3).

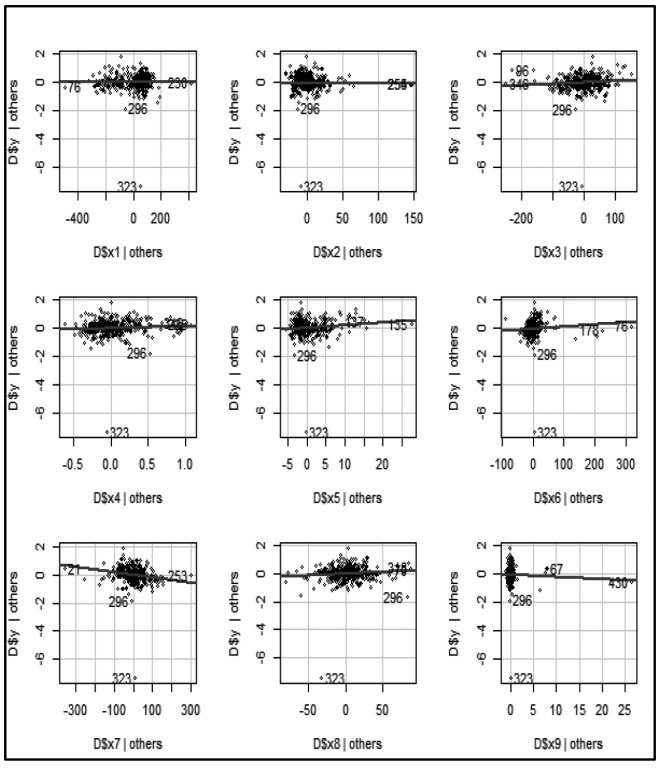

Figure 2. avPlots of the variables (1/2).

Sometimes, even when variables in a multiple regression model are statistically significant, they may or may not be practically significant. This is where the practicality of added variable plots comes into play. Partial regression plots, also known as added variable plots, are graphical representations of the regression of pH on individual parameters keeping the other predictor variables under control. Visually, relationships between pH and $x_3, x_5, x_7, x_8, x_{13}$ *and* x_{14} appear to be linear and significant (Fig. 2 and Fig. 3).

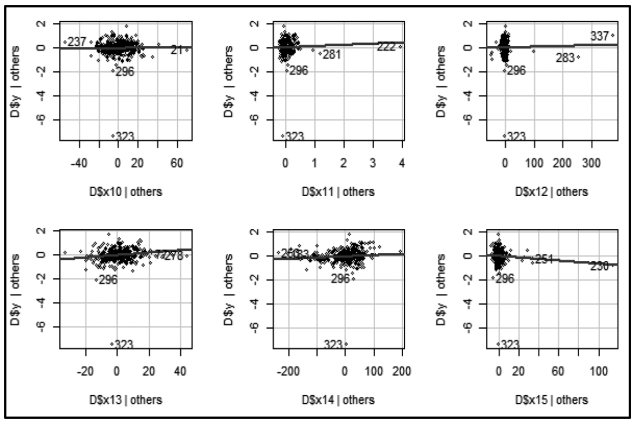

Figure 3: av Plots of the variables (2/2).

Discussion

The result obtained is both quite surprising as well as logical. Historical literature along with common chemical sense says that both inorganic and organic compounds should affect pH. However, elements like Calcium, Iron, Potassium, and Nitrogen, that readily form oxides and hydroxides, having insignificant effects on pH comes off as a surprise [13]. Boron, although water soluble, is fairly non-reactive. Furthermore, Carbonates are regarded as moderately strong bases [14]. So, along with total alkalinity, it having no significant effect on the pH of surface water in Agra is unprecedented unlike the case with BOD. BOD is a measure of the organic quality of water [15], and understandably should have little to no effect on the pH. Again, among the significant parameters, DO should not affect the pH of a solution as there is no physical-chemical connection between the two. Further investigation is required in this regard to establish a theory behind this outcome. The conclusions drawn clearly deviate from previous knowledge, but theyaredefinitely notinaccurate. The parameters that have been labelled as insignificantdo in fact affect the pH, just not strongly enough to become significant. It is also worth notingthat the analysis and the results adhere tothe Agra

region, and the significant parameters might change for any other region where some other parameter is present in abundance in place of the ones analyzed in this study.

Conclusion

The analyses results point towards the conclusion that Chloride, Fluoride, Dissolved Oxygen, Bicarbonates, Sulphates, Magnesium, and Sodium arestatistically significant in predicting the pH levels of surface water in Agra, Uttar Pradesh (Total Alkalinity, Biochemical Oxygen Demand, Carbonates, Boron, Calcium, Iron, Potassium, and Nitrogen being insignificant).

Further research may be conducted in investigating a probable non-linear relationship between the predicting variables and the pH.

References

1. 1st Census Report of Water Bodies Volume 1, Department of Water Resources, River Development and Ganga Rejuvenation, Minor Irrigation (Statistics) Wing, Ministry of Jal Shakti, Government of India, page 1-200, 2023.
2. A Banerjee, India's misunderstood rivers, HT Media Limited, 2015.
3. R Kumar, R D Singh, K D Sharma, Water Resources of India, Current Science,Vol 89, page 794-811, 2005.
4. H Yehia, S Said, Drinking Water Treatment: pH Adjustment Using Natural Physical Field, Journal of Biosciences and Medicines, Vol 9, page 55-66, 2021. https://doi.org/10.4236/jbm.2021.96005
5. pH and Water, Water Science School, United States Geological Survey, 2019.
6. C H Zalvan, S Hu, B Greenberg, J Geliebter, A Comparison of Alkaline Water and Mediterranean Diet vs Proton Pump Inhibition for Treatment of Laryngopharyngeal Reflux, JAMA Otolaryngology: Head & Neck Surgery, Vol 143, page 1023-1029, 2017.https://doi.org/10.1001/jamaoto.2017.1454
7. S J Li, W Y Lv, H Du, Y J Li, W B Zhang, G W Che, L X Liu, Albumin-to-Alkaline Phosphatase Ratio as a Novel Prognostic Indicator for Patients Undergoing Minimally Invasive Lung Cancer Surgery: Propensity Score Matching Analysis Using a Prospective Database, International Journal of Surgery, Vol 69, page 32-42, 2019. https://doi.org/10.1016/j.ijsu.2019.07.008
8. Y Tan, Y Zhu, Y Zhao, L Wen, T Meng, X Liu, F Hu, Mitochondrial Alkaline pH-Responsive Drug Release Mediated by Celastrol Loaded Glycolipid-Like Micelles for Cancer Therapy, Biomaterials, Vol 154, page 169-181, 2018.https://doi.org/10.1016/j.biomaterials.2017.07.036
9. Y Sun, S Hou, S Song, B Zhang, C Ai, X Chen, X N Liu, Impact of Acidic, Water and Alkaline Extraction on Structural Features, Antioxidant Activities of Laminaria Japonica Polysaccharides, International Journal of Physical Macromolecules, Vol 112, page 985-995, 2018. https://doi.org/10.1016/j.ijbiomac.2018.02.066
10. G B Martins, F Tarouco, C E Rosa, & R B Robaldo, The Utilization of Sodium Bicarbonate, Calcium Carbonate or Hydroxide in Biofloc System: Water Quality, Growth Performance and Oxidative Stress of Nile Tilapia (Oreochromis niloticus), Aquaculture, 468, page 10-17, 2017.
11. Surface Water Quality Uttar Pradesh 2007-2021, Department of Water Resources, River Development and Ganga Rejuvenation, Ministry of Jal Shakti, Government of India, 2022
12. A K Kadam, V M Wagh, A A Muley, B N Umrikar, R NSankhua, Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India, Modeling Earth Systems and Environment, Vol 5, page 951-962, 2019.
13. T Ogino, T Suzuki, K Sawada, The formation and transformation mechanism of calcium carbonate in water, Geochimica et Cosmochimica Acta, Vol 51, No 10, page 2757-2767, 1987. https://doi.org/10.1016/0016-7037(87)90155-4
14. A O Alghamdi, M O Abu-Al-Saud, M B Al-Otaibi, S C Ayirala, A Alyousef, Electro-kinetic induced wettability alteration in carbonates: Tailored water chemistry and alkali effects, Colloids and Surfaces A: Physicochemical and Engineering Aspects, 583, 2019. https://doi.org/10.1016/j.colsurfa.2019.123887
15. N R Das, Nonparametric Tests for Potability of Damodar River from Small Sample, Indian Science Cruiser, Vol 37, No 2, page 30-34, 2023. https://doi.org/10.24906/isc/2023/v37/i2/223483