A Comparative Assessment and its Characterization of the Integrated Novel Water Pollution Index and its Statistical Approach for the Evaluation of Spatial Variations Using Factor Analysis: A Geospatial Approach in Mahanadi River, Odisha

Abhijeet Das[1]

[1]Research Scholar, Department of Civil Engineering, C.V. Raman Global University (CGU), Bhubaneswar, Odisha, India; ORCID ID: https://orcid.org/0000-0003-4599-5462

Abstract: Knowledge on water quality and its assessment, is necessary for both human health and environmental benefit. To account for spatial distribution, surface water quality parameters were analysed using integrated interpolation, geographical information systems (GIS) and multivariate analysis. A total of 19 locations and 13 water quality indicators were analysed, for a duration of six years (2018-2024). The study's main objective was to assess the seasonal and regional variations in the water quality index (WQI) of Mahanadi River in Odisha using $(N)_{pi}$, $(S)_{pi}$, $(O)_{pi}$, $(C)_{pi}$, $(E)_y$-WQI, Int_w-WQI and Multivariate Statistical tools namely Factor Analysis (F_a). However, in the current investigation, pH, HCO_3^-, Na^+, K^+ and Mg^{2+} were within the permissible limits as per WHO standards. According to this study, the order of prevalence of ion concentrations is signified as follows: $Mg^{2+} > Ca^{2+} > K^+ > Na^+$ for cations and $HCO_3^- > Cl^- > SO_4^{2-}$ for anions. The analysis of $(N)_{pi}$ indicated that about 15.79% of the sampled area, is affected by turbidity content, which is highly unsuitable for consumption. However, the remaining area (84.21%) is within the safe category of water. Classification of water based on $(S)_{pi}$ represents most of water samples falls between good water quality. Three unsuitable samples is noted as a result of excessive TDS and EC. In case of $(O)_{pi}$, over 84.21% of the samples fell into categories of excellent, indicating the suitability for human activities. Using surface water quality results from $(C)_{pi}$ model, that reflects that out of 19 samples, 16 were suitable for drinking. Whereas 2 were polluted and 1 is seriously polluted, thus promotes unsuitability. Although there are several established techniques for calculating the WQI, the current study uses the quality index to consider a variety of water quality concerns in a cohesive manner. Meanwhile, in case of $(E)_y$-WQI, 84.30% were excellent whereas 10% and 5% were poor and high polluted category. Over 42.11% of the samples fell into the categories of poor/very poor/not suitable, using the Int_w-WQI diagram. Therefore, using these six approaches resembles a precise and comprehensive method to comprehend water quality in relation to pollution for human usage. In later stage, a factor analysis (F_a) can be applied to lessen the subjectivity and dimension of water quality characteristics. It reveals that the first five principal components explain almost 95.61% of dataset variation. This method removes the aggregation problems, weighting, opacity, and biases seen in traditional water quality evaluation techniques. The results of F_a suggested that turbidity, TKN, Ca^{2+} and Cl^-, were the primary determinants of the water's quality. The amount of organic pollution that was released into the river was influenced by anthropogenic activity in the vicinity of the river. In addition, the traditional dense habitation next to the river and the manufacturing waste that is transported from upstream to downstream are the sources of the high amount of TKN in urine and faeces. Therefore, given the high spatial distribution of geogenic turbidity and TKN occurrence, the study's findings minimize uncertain causes and offer insights into surface water pollution regimes. They will also be useful to policy makers in helping to better plan, allocate resources, and manage the area's potable water supply.

Keywords: WQI, Mahanadi, $(N)_{pi}$, $(S)_{pi}$, $(O)_{pi}$, $(C)_{pi}$, $(E)_y$-WQI, Int_w-WQI, Multivariate.

1. Introduction

The demand for water is rising globally, making efficient management of water resources a major problem, especially in areas with erratic and unpredictable precipitation patterns [1]. Likewise, the quality of the water on the planet has been steadily declining for decades due to the combined effects of manmade and natural factors. Meanwhile, the rate of pollution in the environment has increased due to rapid industrialization, urbanization, and human habitation development. However, surface water quality is essential to preserving the aquatic ecosystem's natural state. The availability of water is always vital to the local population since it sustains life and the global economy [2]. However, because of the world's population growth at a rapid pace and the declining capacity of existing water resources, the relevance of both national and global water resources has increased. Further, numerous studies have documented variations in surface water, river flow, and the temporal and spatial distribution of water as a result of variations in temperature, evaporated state, and precipitation rates. Because of the lowering water table, unfavourable effects from agricultural use, mining, growing urbanization, and climate change, limited surface water supplies are under increased stress [3]. Increased water usage for irrigation and domestic uses as a result of the growth of agriculture and urbanization has directly affected changes in land use and cover, water scarcity, flood hazards, and soil erosion, all of which have an impact on living conditions. Also, human-induced climate change and land-use change scenarios have sparked worries about the sustainability of water resources all throughout the planet. Water resources worldwide may be negatively impacted by the combined effects of land-use change and climate change, which may also have a severe influence on water quantity and quality [4]. In order to properly manage water resources, it is possible to analyze issues with water pollution, identify possible sources of contamination, and give information assistance and reference through the use of dynamic spatiotemporal evaluation of water quality. Thus, before water is used for human consumption, it is imperative to assess its purity [5]. Thus, comprehending these complex relationships makes a substantial contribution to the discipline since it holds the key to sustainable methods that might lessen adverse effects on ecosystems, water quality, and the long-term stability of the landscape. For thorough research and successful prevention and management of surface water contamination, dependable data on water quality are essential [6]. Consequently, improvement of surface water has thus been given priority in research areas related to water resource management. A long-term monitoring program should be created to precisely evaluate the physicochemical and biological properties of surface water, taking into account the spatiotemporal variance in those features [7]. Hence, comprehensive information on the heterogeneity and connectedness of water quality, as well as the main mechanisms governing surface water flow and quality, can be obtained by carefully analyzing the chemistry of surface water and its geographic variation [8]. Incorporating water quality index (WQI), consistently serves as a helpful technique for assessing the shift and direction of the water environment quality by combining several original parameters into a single index. Using WQI to assess aquatic ecosystems has several benefits, one of which is the easy reproducibility of presenting vast amounts of previously available data on water quality in a single index value [9]. Hence, when it comes to classifying the water quality for public and sustainable water resource development, our designed index is comparatively user-friendly. In a similar vein, the specified water quality standards' allowable limits are taken into account while determining its index value. Since it functions like a single index value, that is connected to a possible classification scheme for water use, that aids in determining the aquatic resource's ecological potential. The single index value and the classification scheme it is associated with are essential components of the framework for determining the most important course of action for sustainable and efficient management [10]. Again, the WQI model makes it easier to make decisions and plan for the management of water resources at the district, regional, and national levels because it is more adept at clearly identifying the overall condition of an aquatic ecosystem using a single index than other more thorough analytical methods like traditional water quality assessment programs [11]. In addition, numerous adjustments were produced in compliance with the guidelines, and a large number of extra indexes were made. The difficulty is figuring out whether the variations in water quality are more likely to be caused by biogeochemical changes in natural processes or by human activity contaminating rivers [12]. Pollution Index (PI) models like such as Numerow Pollution Index (N_{pi}), Synthetic Pollution Index (S_{pi}), Overall Index of Pollution (O_{pi}) and Comprehensive Pollution Index (C_{pi}) has been presented in the current investigation. Unlike the WQ models, together, these methods assess the water quality by quantifying the extent of surface water body contamination. As a result, these models frequently exhibit lengthy persistence as well as bioaccumulation and biomagnification in aquatic compartments [13]. For this, the most used indexing technique for assessing surface water quality for a variety of applications is the (N_{pi}). This model frequently determine the portion of heavy metal pollution in water bodies that is caused by humans [14]. As a method for evaluating the water quality, in order to describe the overall water quality condition and grade, it integrates several factors into a dimensionless variable and assigns a relative weight to each parameter based on its significance that helps in protecting the aquatic environment [15]. Since, it uses or creates a single index value as a measure of pollution, adopting the same methodology as WQI models, so applying (S_{pi}) value is frequently evaluated in comparison to ideal values and standards in order to be categorized as highly or lowly polluting water [16]. On the other hand, because it offers a straightforward, reliable, and verifiable unit of measurement and disseminates information about water quality to decision-makers and concerned citizens, it serves as a useful instrument for expressing water quality. However, because of the interactions between experts, this model aims to remove some of the biasing effects [17]. To estimate the level of contamination throughout the river basin, (O_{pi}) is applied, and it uses an integrated method by converting parameters into a single value index that allows for the recognition of even minute changes in the amount of any input parameter to define water quality. As an adequate indicator, it provides an easy-to-read, fact-based overview of water quality and potential changes [18]. Moreover, using this for several objectives, it could take a lot of time and be ineffective for a single study, whereas (C_{pi}) can offer a conceptual representation of the water quality based on any input parameters and be used, with the suggested standards of the parameters taken into consideration, for a number of specific purposes [19]. These indexes are considered as useful tools for quickly interpreting the quality of water. They combine several physical, chemical, and microbiological factors to produce a numerical value that permits a given degree of pollution and to show the condition of the water in rivers and other bodies of water at the moment [20]. So, it functions as a very useful and efficient method of assessing water quality and sharing data regarding general water quality. Nevertheless, the majority of these indices do not take into account the realities of each zone; rather, they are typically chosen, given a subjective weight, and then converted to a scale. One of the major studies by [21], that offers a thorough analysis of the WQI's development throughout time, identifies significant constraints in the index-building process, and offers suggestions for overcoming those restrictions [22]. The sensitivity of the index can be impacted by significantly weighted characteristics, indicating that weighing of the parameters is quite essential, as per the way, on which the water is utilized for irrigation, drinking, or domestic uses [23]. Therefore, a slight adjustment to the weight distribution will have an impact on how water quality is interpreted overall. To get around the inaccurate parameter weighting, Entropy (E_y) and Integrated-Weighted (Int $_w$)-WQI, are frequently employed techniques for calculating relative weights of parameters. The concept of information entropy (E_y), was initially put forth by [24] and is a useful method for supplying probability and uncertainty. After then, the variations brought on by human factors can be avoided since this novel method establishes the index weight based on the degree of fluctuation of each index value. In the latter scenario, the higher the entropy score, the larger the fraction generated within each pollutant group or within the component contribution, and the lower the quality of the water [25]. However, entropy weights possess drawbacks as well, such as instability traits brought on by variations in sample size and value and the failure to take into account index conflicts. To make up for the deficiencies, use of (Int $_w$) approach, has been utilized to compare and evaluate the various possibilities in order to choose the best solution [26]. Under unidentified circumstances, this method helps to choose the best and most appropriate answer by offering a systematic approach. As a result, the combination of these approaches enables a more objective and nuanced assessment of many criteria and sub-criteria [27]. The novelty includes the methodical organization of the decision hierarchy, the exacting pairwise comparisons used to establish the weights of the criteria, and the use of these indexing WQ procedures for spatial analysis afterward. So, this complicates the procedure, even if the computations are straightforward [28]. Thus, previous studies have demonstrated that simulations based on deterministic water quality models frequently produce false findings because these models are unable to handle the flaws and ambiguity in the underlying data or adequately depict the complexity of water quality, which depends on numerous factors [29]. To achieve the goals, we have employed a thorough modeling strategy that comprises of multivariate statistical tools (MST) such as Factor Analysis (F_a). In an extensive assessment of the quality of the water, the challenge is to ascertain if the biogeochemical alterations in natural processes or human activity contamination of rivers should be blamed for the changes in water quality [30]. Thus, F_a has been extensively used to assess the quality of water, has the ability to remove unnecessary information without sacrificing important information, and can simplify data dimensions from complicated water quality data matrices [31]. Hence, with the least amount of information lost, this method extracts a limited number of components that capture the key patterns of complex multivariate data and visualizes the outcome in a low-dimensional space. Consequently, it lessens arbitrary assumptions and raises the index's accuracy [32]. Among the various MS techniques, F_a is the dimensionality reduction approach that is most frequently used in the assessment and gathering of statistical data to clarify and explain relationships between various variables that are included in the dataset [33]. Hence, this application helps in the examination of intricate data matrices in order to clearly comprehend the water quality and identify potential influences on the water chemistry in an ecosystem [34]. In later stage, spatial analysis is the practice of using data analysis to find trends and comprehend how geographical features and occurrences relate to one another spatially. The spatial distribution of surface water quality groups, was further investigated by plotting the geographic membership values of each group on a map, which offered a helpful perspective on the hydrogeochemical and quality attributes of surface water [35]. It necessitates the use of a Geographical Information System (GIS), to produce maps and do geographic feature analysis in order to visualize and work with spatial data. The GIS methodology, when combined with the IDW interpolation method, has been used in recent years to regularly examine and monitor the quality of surface water. This technique has shown to be an effective means of studying and assessing spatial data related to water resources [36]. Thus, when combined with spatial statistical models, it could prove to be a useful instrument for urban and regional planning. The IDW interpolation technique employed in this investigation, is currently a useful method for surface water quality parameter spatial interpolation, resulting in the creation of spatial distribution maps [37]. Several studies have effectively identified possible outcomes using GIS, and they have spatially modeled urban social components [38]. Because of this, GIS and MS techniques can be used to provide tools for systematic overlaying of many quantitative and qualitative criteria on a single platform. This has proven to be highly helpful in agricultural management studies [39]. Given that the Mahanadi River provides water in the State of Odisha, for its use in agriculture, energy production, fishing, consumption by people, industry, and aquatic life habitat. However, the increasing urbanization of the area, intense farming practices, and the discharge of household and industrial waste water have all had an impact on the waterways [40]. So, this investigation yields a thorough knowledge of the water quality in terms of pollution status, quality in comparison to surface water criteria, and ecosystem viability. Therefore, in order to identify a clear relationship between spatiotemporal oscillations and the effects of water pollution, 13 water quality measures were kept under observation during a 6-year period (2018–2024), at nineteen distinct sites, that produces an enormous data matrix. But none of the researchers concentrated on defining the river with a clear knowledge of the overall condition of the river water through the combination of multivariate statistical analysis and WQI, which would have greatly informed future underlying research needs and strategies for efficient and sustainable management. Therefore, the main objectives for the study are (1) to ascertain the water quality characteristics through the adoption of four indices with support of E_y and Int $_w$ method, (2), to analyze the parameters concentration of the rivers spatiotemporally through ArcGIS interpolated maps and finally, (3) using Factor Analysis (F_a), it assists in extracting data from monitoring stations, identifying the key classification variables for spatial variations in water quality, analyzing the potential influences on water quality in various regions with varying levels of contamination, and investigating potential sources of contamination. Additionally, this will assist local stakeholders, governmental organizations, and managers of water resources in putting into practice sustainable surface water resource management techniques or updating techniques from previous decades for efficient management.

2. Materials and Methods

2.1 Study Area

The study area comprised the Mahanadi River System **[41]** and is situated in the region between the longitudes 80°30' and 86°50' E and latitudes 19°20' and 23°35'N, which is bordered by the Maikalahill range to the west, the East Gates to the south and east, and the Central Indian hills to the north. This basin also, scatters around 141, 589 Km^2, which thereby, represents around 4.3% that was of the country's land area. In this specific case, the catchment area for the region gets distinguished in larger portions in Orissa, that reveals about 65,600 Km^2 and Chhattisgarh State, that approximately detailed as 74970 Km^2, while the smaller parts of Jharkhand, contribute of 650 Km^2, Maharashtra as 250 Km^2, and finally, Madhya Pradesh represents about 130 Km^2. Additionally, it is noteworthy that one of the biggest peninsular rivers in eastern India, it spans 851 km from its source to its confluence in the Bay of Bengal, with 357 km falling within the states of Chhattisgarh and Odisha. In addition, the main source of crops that are grown are: rice, sugarcane, and oilseeds, that listed are the principal crops grown in this river. Generally, the Mahanadi River's average yearly discharge is around 1895 m^3/sec, reaching its highest point of about 6352 m^3/sec, when the summer monsoon arrives, while at least 759 m^3/sec, contributes towards the dry season i.e., from October to May. For this purpose **[42]**, this river has a tropical monsoon environment, with records for the highest July temperature reaching up to 45 °C and in the winter, the temperature stays at about 25°C, respectively. However, the most of the rainfall, roughly 75% occurs in the summer, from mid-June to mid-September, during the Southwest monsoon. Additionally, it receives about 140 cm of precipitation on average each year **[43]**. Likewise, the interchange of these crops through the nearby, well-developed towns or cities in the state of Cuttack is facilitated by a respectable transportation network. The Hirakud reservoir is employed in the production of hydropower, irrigation, and flood control, and the reservoir's installed capacity is fixed at 347.5 MW, coupled with a culturable command domain, that reflects about 235,477 ha. Consequently, the dam's design flood is termed to be 42,450 m^3/sec of its maximum spillway capacity, that signifies in the present study area. The local farmers who cultivate crops for the sale in and around cities or towns as their primary source of income benefit from this exchange **[44]**. The section of the river used for this investigation, consisted of three sections and nineteen sampling locations, such as upstream, midstream, and downstream (Figure 1).

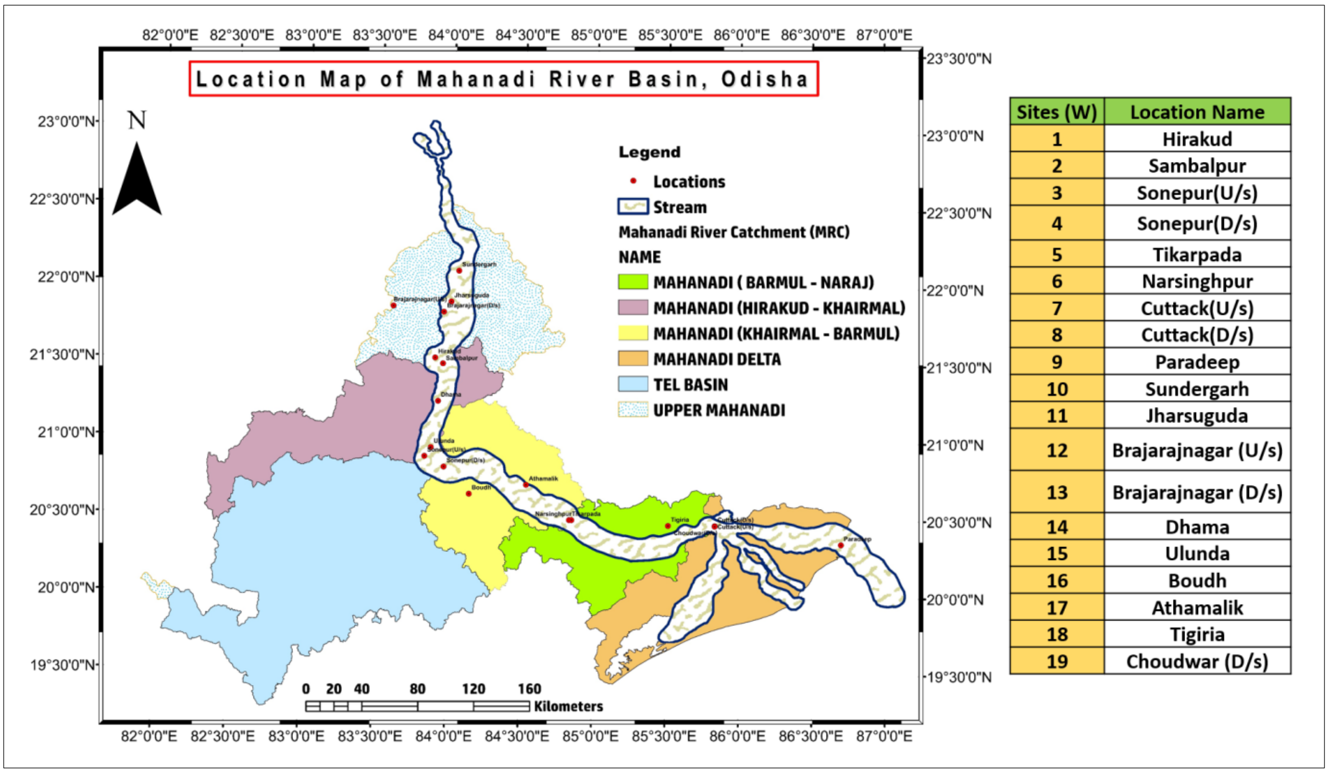


Figure 1. Location of Mahanadi River, Odisha, and water sampling sites in the studied area

2.2 Sample Analysis

The river basin's field verification maps and well inventory survey, were completed with the aid of toposheets, that has been obtained from the Survey of India (SOI). For determination of locations, Global Positioning System (GPS) is adopted to record the location and its altitudes. In the ongoing work, water samples were gathered at 19 sample locations, during the timeframe of 2018-2024 in pre-monsoon season. When selecting the parameters to assess the water quality, it is crucial to carefully take into account a number of factors. These samples were gathered in 1-liter high density polyethylene plastic bottles prior to testing, and they were then utilized in a lab setting for additional physicochemical investigation. First, reviewing earlier studies on assessments of water quality in similar rivers or locations may provide information on features that are often assessed [45]. Also, it is crucial to understand the local environment, topography hydrology, and the impact of humanity. The reported 13 water quality criteria were assessed and chosen based on World Health regulations. The State Pollution Control Board (SPCB) carried out the water sample analysis, in accordance with the standard operating procedure, as directed by the American Public Health Association (APHA) standards and has adopted adequate sample preservation guidelines, as stated by WHO [46], and every sample was examined using several criteria. Further, the samples were moved into the lab of C.V. Raman College of Engineering, Odisha, that were cross-checked by which frequent runs of established standards and blanks were performed in between samples to prevent errors and preserve the instrument's quality, accuracy, and precision [47]. The chemical analysis of water was carried out using the duplicate analyses method. The 13 water quality parameters and salts include: Temperature, Electrical conductivity (EC) and the concentration of hydrogen ion (pH) were examined using a digital thermometer, EC meter and pH meter. In addition, Total dissolved solids (TDS) are acquired by multiplying the EC by a factor (0.55-0.75), according to the equivalent amounts of ions. The bicarbonate (HCO_3^-) ion is estimated by HCl titrimetric method, while the calcium (Ca^{2+}) is examined by EDTA volumetric method. Further, sodium (Na^+) and potassium (K^+) were measured by flame photometer while the Chloride (Cl^-) was measured using $AgNO_3^-$ titrimetric method. Total kjeldahl nitrogen (TKN) was determined using the spectrophotometer, while the turbidity is estimated using turbidity meter. Remaining parameters like magnesium (Mg^{2+}) is measured by standard methods, as implemented according to APHA guidelines. The chemical parameters' units, in this ongoing investigation, were indicated by milligrams per liter (mg/L), except for parameters like pH (non-dimensional) and also, EC (µS/cm at 25°C). The concentration of cations (Ca^{2+}, Mg^{2+}, Na^+ and K^+) and anions (HCO_3^-, Cl^-, and SO_4^{2-}) are expressed in mg/L of each surface water sample were computed for checking the accuracy of the

 https://doi.org/10.1051/matecconf/202440002007

analyzed results, using charge balance error (CBE). This is expressed in %. The computed CBE of the water samples is <5%, which is within the limits of acceptability, indicates the reliability of the analyzed results. Moreover, plotting different thematic layers was done using the ArcGIS software of version 10.8 spatial analysis module. Using GIS, the IDW approach was utilized to create spatial variation plots of the WQIs categories. The study intends to pinpoint an area that may potentially have surface water availability in the Mahanadi Watershed in Odisha. Modern, sophisticated computational techniques have been applied in a number of research lately to lower model uncertainty. Seven models were created in this study using statistical techniques and indexing algorithms. However, the specific data relevant for the study was obtained from State Pollution Board, Odisha, while the toposheets were gathered from Survey of India (SOI). These thematic layers are essential to comprehending the hydrological circumstances that affect the existence of surface water. They provide a strong foundation for a thorough assessment of the surface water potential of a given area. The specifics of the study's methodology are provided in Figure 2.

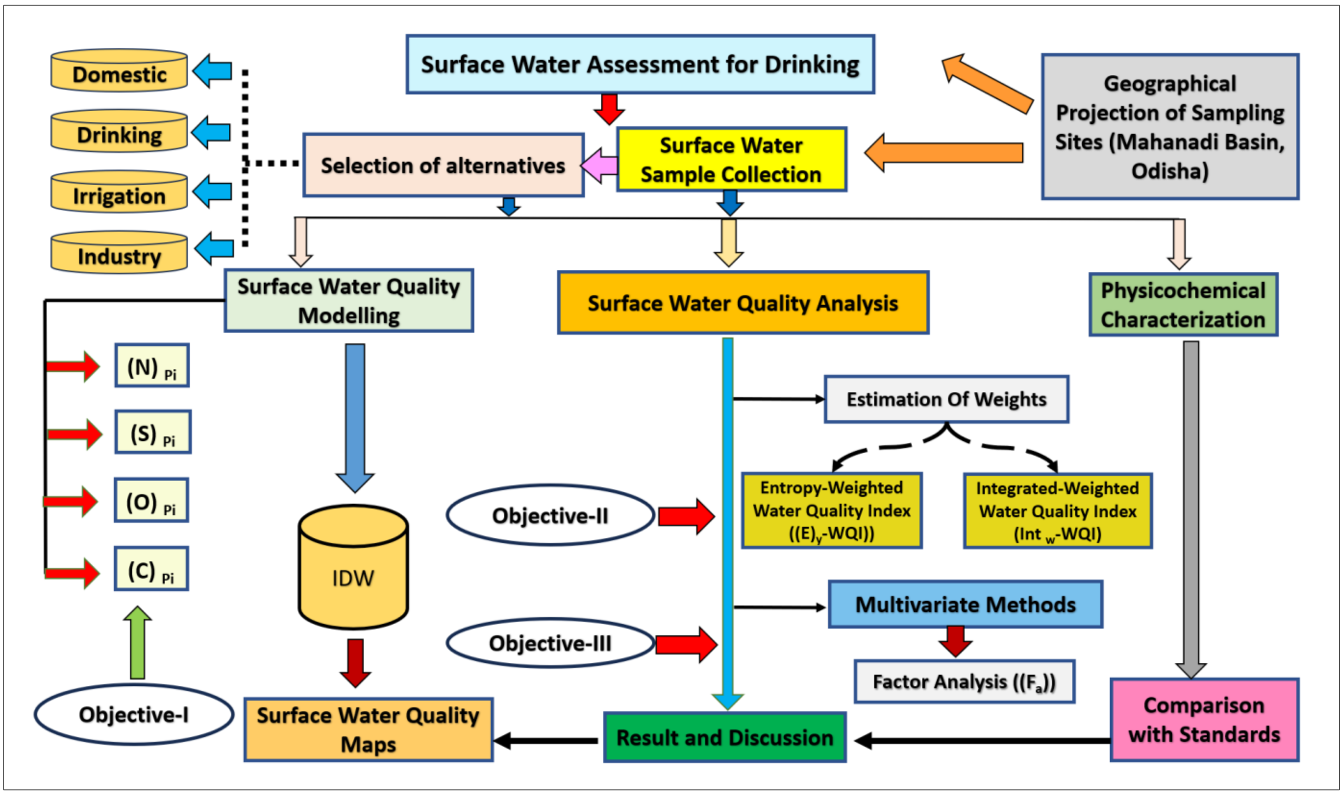


Figure 2. Conceptual framework for selecting robust combination of water quality indicators to develop a novel water quality index model

2.3 Numerow Pollution Index ((N) pi)

The current study used $(N)_{pi}$, that acts as an overall pollutant indicator to account the combined influence of several pollutants for a given application **[48]**.

As per [49], it is calculated using the following eq. (1) as given by

$$D_k = (P)_k / (I)_k$$

$$(N)_{pi} = \Sigma D_k \qquad (1)$$

where $(D)_k$ is the nth parameters pollution index. The term $(P)_k$, describes the observed value while $(I)_k$ is interpreted as k^{th} parameter's permissible value. In this model, four primary types of water have been distinguished, based on $(N)_{pi}$ values. Between 10 and 20, it indicates excellent; good water has a range of 20 to 30; very poor has in between 30-40 and ultimately, higher than 40, indicates about unsuitability for drinking.

2.4 Synthetic Pollution Index ((S) pi)

As addressed by [50], this technique developed a surface water pollution index to evaluate the relative influence of each chemical parameter on the overall quality of the surface water suitable for drinking requirements. However, this serves as a helpful tool for managers, decision-makers, and the general public to view the water quality data **[51]**. The value of $(S)_{pi}$ was determined using the subsequent equation Eq. (2):

$$(S)_{pi} = \sum_{i=1}^{k} \left(\frac{Tp}{Up}\right) * Xp \quad (2)$$

where (T)p represents the analytical value of the water quality parameters, $(X)_p$ is the weight of each physicochemical parameter and $(U)_p$ stands for the standard value of the water quality parameters recommended by WHO. By designating (T)p of the water quality indicators, as the suggested baseline, beyond which, the water is not suitable for consumption, and this model's critical value was evaluated. Consequently, the value represents the total effects of numerous water quality criteria that are thought to be harmful to residential water. The rating scale is suggested as 0-0.25, that signified as excellent, 0.25-0.50 is classified as Good; between 0.50 and 0.70, grouped as Poor water; value of 0.75-1, as Very poor and finally, worst water quality is being obtained when its index value >1.

2.5 Overall Index of Pollution ($(O)_{pi}$)

The applicability of the $\mathbf{(O)_{pi}}$ index reflects the water's health quality in Indian circumstances and is employed to determine the contamination level for each water sample, which is used to evaluate the quality of the water **[52]**. The ultimate value was calculated using the equation Eq. (3), that follows as:

$$\mathbf{(O)_{pi}} = (1/u) * \sum_{x=1}^{u} (Z)x \quad (3)$$

where $(Z)_x$ reveals about pollution index value for the x th parameter, 'u' signifies as the number of parameters, $(Z)_x$ = ($(R)_u$: observed value of parameter))/ ($(R)_s$: Standard value of parameter). According to its categorization [53], pollution index is classified into 5 classes as excellent (OIP<1.9, class C_1), acceptable (OIP<3.9, class C_2), slightly polluted (OIP<7.9, class C_3), polluted (OIP<15.9, class C_4) and heavily polluted (OIP>16).

2.6 Comprehensive Pollution Index ($(C)_{pi}$)

This index Measure the concentrations of each specific drinking water quality standard and assess the overall quality of the water **[54]**. Based on the equation below, the results of water quality were determined as given in Eq. (4):

$$\mathbf{(C)_{pi}} = 1/d * \sum_{p=1}^{n} (H)i \quad (4)$$

where d represents number of monitoring parameters; $(H)_p$ explained the pollution index number. As a result, $(H)_p$ is calculated as per Eq. (5):

$$(H)_p = (B)_p / (M)_p \quad (5)$$

where $(B)_p$ corresponds to the measured concentration of parameter's number in water and $(M)_p$ refers to the permitted limitation of parameter number according to environmental standard **[55]**. Based on this, division of different classes is carried out as per $\mathbf{(C)_{pi}}$ score. If its value is between 0 and 0.20, surface water is in clean category; 0.21 to 0.40, holds a property, present in sub-clean category; 0.40-1.0 is represented by slightly polluted water; value among 1.01 to 2.0, signifying polluted water in the study area and finally, higher than 2.01, surface water stands as heavily polluted category.

2.7 Entropy (E_y)-Water Quality Index (WQI)

This technique has evolved as a helpful method that replaces traditional WQIs, which are typically based on Delphi, and the analytical hierarchy method, which typically depend on assigning weights to factors based on expert and subjective judgment [56]. To get beyond this, it uses information entropy to give water quality criteria weights. Furthermore, this creative approach illustrates how to assess the spatiotemporal variability of a water source's water quality by creating a straightforward WQI calculation procedure that requires less work and produces more accurate results for surface and subsurface water quality assessments **[57]**. Hence, this model demonstrates the practicality and reliability of applying entropy theory, which has been extensively used to allocate air pollution and riverine pollution sources, to the allocation of water pollution sources by evaluating the contributions of potential pollution sources **[58]**. So, efficacy of the entropy approach's recommended procedures, depend on the standard vocabulary for distinguishing and addressing complex water concerns, as well as the conceptual basis used for evaluation processes [59]. This $(E)_y$-WQI model is categorized into five classes: <50 (excellent zone), 50-100 (Good class), 101-150 (medium zone), 150-200 (poor zone), and >200 (denotes the water to be "extremely poor" or 'unsuitable' to use). The formulas for establishing and prioritizing the selection criteria, calculating the revised weight, and estimating the coefficient of each criterion and finally, the relative weight calculation $(W)_j$, are advisable by **[60]. A flow chart illustrating the modelling steps is shown in Figure 3.** Lastly, the combinations of weights and a Quality rating scale, that is expressed as $((L)_n)$, is given by Eq. (6):

$$(E)_y\text{-WQI} = \sum (W)_j * (L)_n \tag{6}$$

Where, $(L)_n$ for each parameter is given as the ratio of the monitored value of the n^{th} parameter $(V)_n$ to its standard value $(S)_n$, is given in Eq. (7):

$$(L)_n = (V_n/S_n) * 100 \tag{7}$$

Hence, several Artificial Intelligence (AI) disciplines, such as data analysis, pattern recognition, control systems, and decision-making, employ this reasoning. Thus, it is especially helpful when there is little or conflicting information available, as traditional reasoning might not be able to offer unambiguous answers [61].

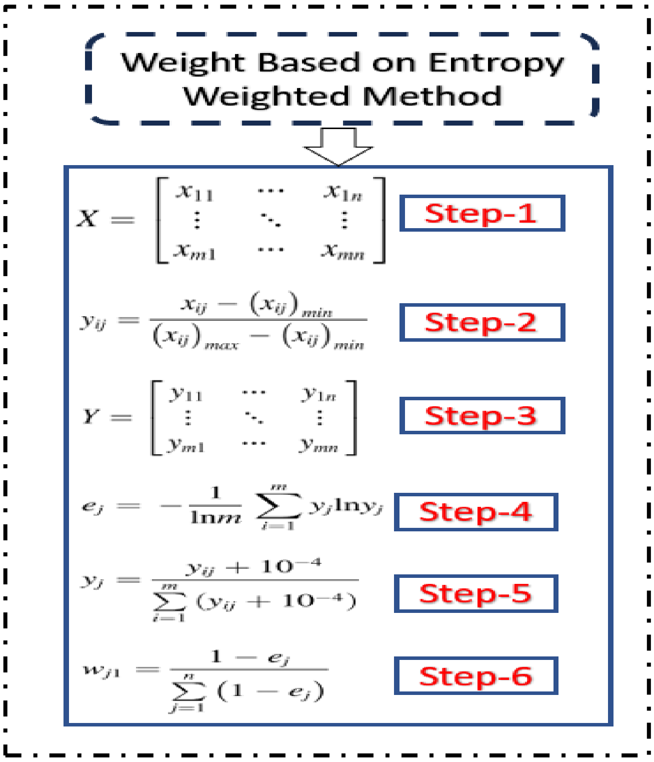


Figure 3. Flowchart of $(E)_y$-WQI, showing the steps taken to preprocess and model concentrations of physiochemical parameters across the sampling locations

2.8 Integrated-Weight (Int_w)-Water Quality Index (WQI)

This new approach has been defined by the adoption of two concepts, entropy, and CRITIC, as the primary metrics for surface water quality classification based on the application of standard parameters during surface water characterization [62]. However, the prevalent flaws in the current WQ models are ambiguity and eclipsing. This approach provides a new framework for evaluating models by objectively reflecting the relevant information through the incorporation of Shannon's entropy [63]. The assessment of surface water quality was advanced by using the CRITIC technique once more as an objective weighing method to determine relative weights of variables and circumvent the drawbacks of traditional information entropy [64]. Consequently, the use of this technique made use of the method to compute the ambiguity and model eclipsing issues. However, ambiguity issues are typically brought about by improper sub-index findings, as well as over- and under-estimation of aggregate results [65]. Hence, it serves as a generally accepted method for measuring any type of uncertainty that exists in a model or system. Regarding I $_w$-WQI, the value varied from 0 and 300, and there were no dimensions. So, it explains the general surface water quality that is suitable for human consumption. Depending on a number of water characteristics, the findings aid in categorizing surface water as: excellent water, that termed as <50; Good as 50-100; Poor quality of water contributes about 100-200; Very Poor as value between 200 and 300 and finally, index score>300, is considered unsuitable for drinking activities at a specific site. An overview of this technique may be found in Figure 4.

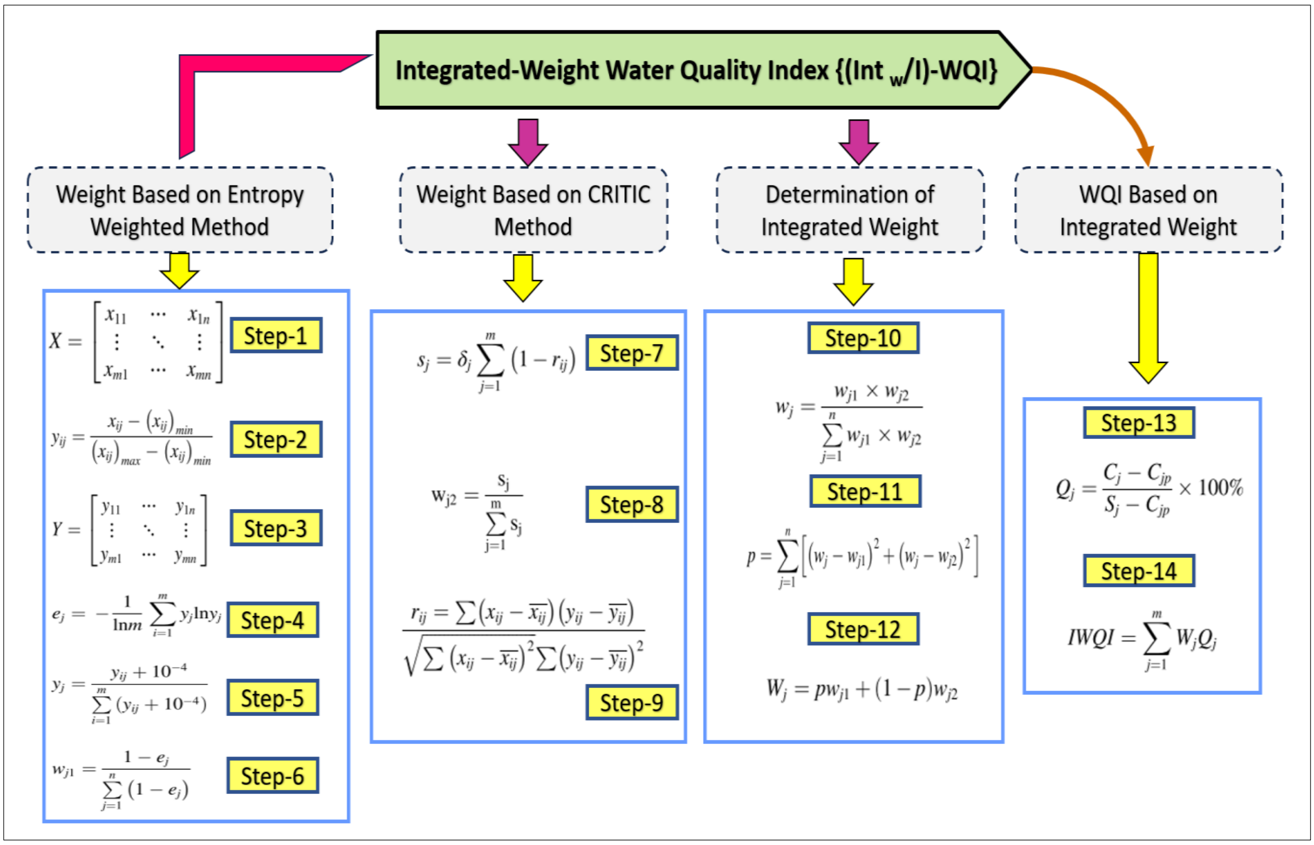


Figure 4. Flowchart of Int $_w$-WQI, showing the steps taken to preprocess and model concentrations of physiochemical parameters across the sampling locations

2.9 Factor Analysis (FA/F$_a$)

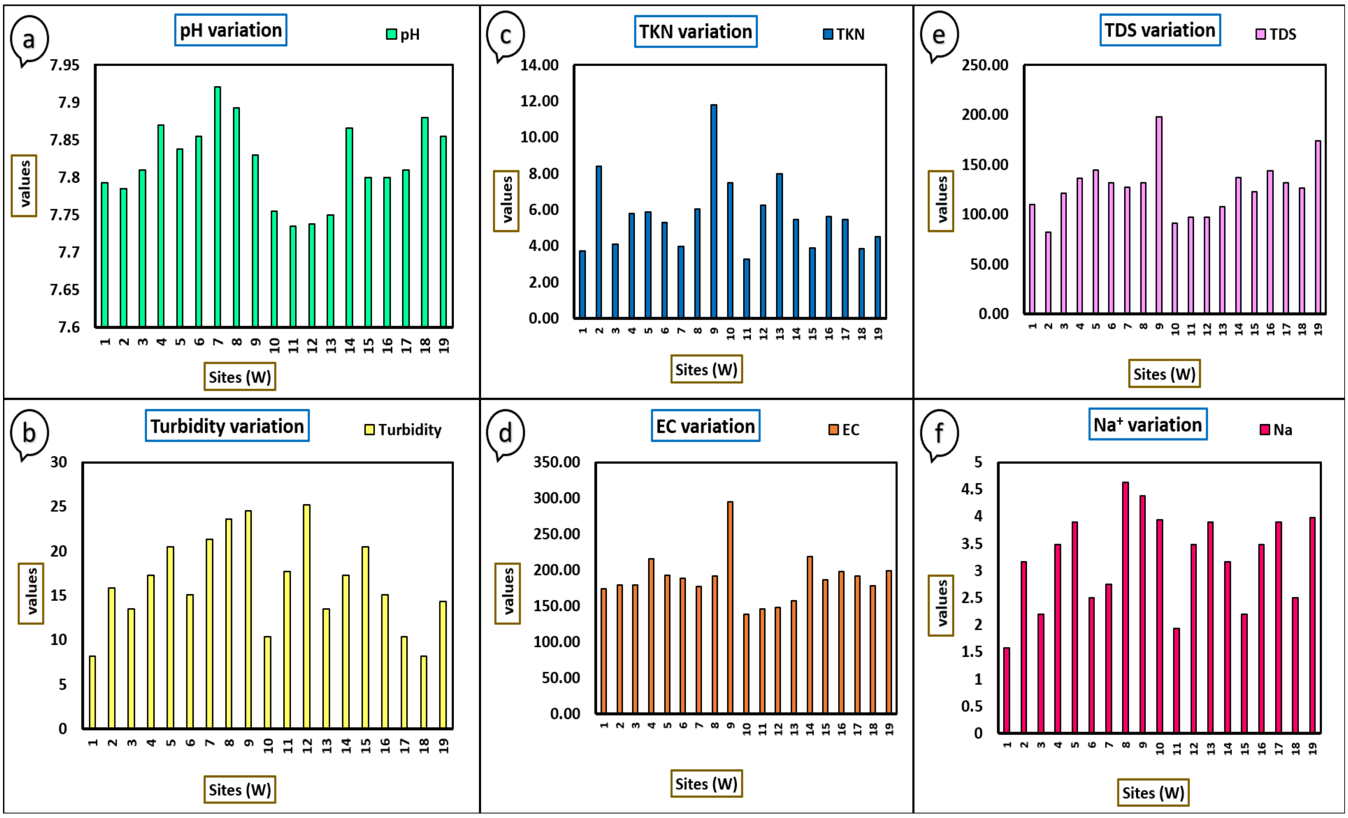
This is a technique that uses a linear combination of prospective hypothetical and random influence variables to represent the original parameters while capturing common factors from variable groups **[66]**. However, multiple indexes are converted into several comprehensive indexes made of a linear combination of the original variables under the assumption of minimizing loss; these indexes are referred to as components **[67]**. When using mathematical computation to calculate PCs, eigen values and eigen vectors are acquired by the use of covariance or other cross-product matrices, which show how the initial variables and the numerous observed parameters are distributed [68]. Hence, through this approach, eigen values from the original set are extracted to create new principal components (PCs) that are orthogonal to one another and hence unconnected. Therefore, it was a statistical technique that made use of the concept of dimensionality reduction. By examining sets of variables and groupings that share comparable attributes, it might make it easier for us to explain observations by allowing us to find patterns or structure in the midst of confusing or chaotic data **[69]**. This approach's unique feature is its ability to extract components that make up a linear combination of all variables, which may be used to represent the dataset's overall variance. The largest residual variability is explained by the remaining components **[70]**. In addition, the majority of the variability in the original data set is included in the PCs with an eigen value bigger than unity, which are often taken into consideration in this method. Since the initial loadings are typically rotated until a simple structure is attained, they may not be easily interpreted **[71]**, that implies each variable's factor loadings are extremely high, as high as 1, on a single one of the PCs and extremely low factor loadings on the other PCs, as low as zero. Therefore, the Kaiser Rule and Scree plots were employed as the only threshold for determining the total number of components. With eigen values less than 1.0, all components are excluded under the Kaiser rule. The scree plot assisted in identifying key elements and understanding the information's fundamental structure. The PC loading factors which follows the order: >0.75, 0.50-0.75 and 0.30-0.50, are categorized as strong, moderate, and weak factor, respectively [72]. Subsequently, the advantage of the F_a receptor model, which was also applied to the data matrix, is that it adds a non-negativity constraint to the data and weights the uncertainty of each data point, guaranteeing that the source contributions are always positive.

3. Results and Discussions

Water quality assessment is mostly done with a variety of biological, chemical, and physical properties. It was important to conduct a stretch wise assessment in order to more accurately depict and understand the consequences of channelization. Therefore, determining the degree of pollution in river water is an important and quickly expanding topic of research that necessitates the collection, processing, and interpretation of data. The amount of water quality parameters may vary, nevertheless, due to the various methodology utilized in the indexing operations. Because comparing the results of multiple models might increase the validity of a conclusion, source allocation using multiple models is therefore essential. In this scenario, the WHO Standards are compared with the statistical information gathered from surface water samples taken during the pre-monsoon season. Additionally, the numerical spatial distribution of the parameters was constructed using the GIS environment and the analytical data, and maps of the spatial distribution of the parameters relevant to water quality were created using the IDW approach.

3.1 Physicochemical characteristics

Temperature is an additional metric that is typically represented as and measured in the field as °C. It has been observed that the direct discharge of wastes from industrial processes, agricultural practices, and communities into river systems containing high-temperature liquid waste poses a significant threat to aquatic life. The amount of DO in the water is generally influenced by the temperature of the water, the flow of the water body, and the amount of organic matter present. The value of Temperature, in the current work ranged from 25.72 to 30.25°C over the different sampling stations on the river, while the optimum temperature range for phytoplankton growth in the waters is around 20-30 °C. This showed that the temperature in the studied area was mostly regulated by differences in altitude. The pH of any solution provides information about the solution's strength and its acidity or alkalinity. The primary governing body is the dissolution of H_2CO_3 (H^+) and by the hydrolysis of bicarbonate (OH^-). The maximum acceptable pH level for drinking water is estimated to be 8.5. Results obtained for pH varied between 7.74 and 7.92. The relatively alkaline pH recorded during the study period may have led to the dissolution of CO_3^{2-} and OH^- mineral in the sediments and an increased desorption of metal cations, hence to enrichment of Fe^{2+} in surface water. Still, the pH value in the study area is within allowable limits for surface water. The alkaline value seen during the monitoring period could be ascribed to the use of chemical detergents, the dumping of industrial waste, or pollution from household garbage. In addition, Turbidity of water is defined as the cloudiness or haziness of a fluid caused by a large number of isolated particles that, like smoke in the air, are frequently unseen to the human eye. However, both organic and inorganic particles, tiny organisms, granular suspended matter, and the muddy or turbid appearance of water are the main causes of cloudiness or suspended solids in water. While observing the standards for the quality of drinking water, the maximum amount of turbidity, that is permitted is about 5 NTU, which is the maximum allowed value. The fluctuation of turbidity value in the research work varies between 8.5-25.20 NTU. The presence of clay materials, small inorganic and organic particles, dyes, algae, and plankton from home sewage discharge and agricultural runoff is thought to be the cause of the elevated values seen at all sites. At nineteen locations, however, the concentration of surface water often exceeds 5 NTU. Consequently, TKN is a measurement of the water's organic nitrogen and ammonia-N content. Elevated amounts of nitrogen in freshwater can lead to a rise in the quantity of algae and chlorophyll, although overall, it is not as concerning as other signs. Synthetic fertilizers, residential and commercial wastewater, inflammatory processes that remove natural deposits, and microbiological activation of organic substances containing nitrogen are the sources of contamination in surface water. In this work, between 3.28 and 11.80, TKN was found. Although all the samples exceed the permissible limit (5 mg/L), which may result from inappropriate manure management techniques, hygienic landfills, leachates from waste disposal, or excessive utilization of inorganic nitrate fertilizer. In later stage, EC of water determines the water body's suitability for irrigation and combating fires by measuring the concentration of ions in it. Temperature, concentration, and kind of ions present all play a role. In addition, the parameter EC in water samples, may result from saline source mixing, aquifer material leaching or dissolving, or from a combination of these activities. However, for drinking purposes, EC should not exceed 250 μ-S/cm. The research area's surface water had EC values ranging from 138.08 to 295. The results highlighted that water samples in all the locations were found within the recommended limits of 250 according to WHO (2011) and can be used for drinking water without any further treatment except at one location. High conductivity at W-(9), can result from both human and natural deterioration. Additionally, the presence, aggregate concentration, transportation, valence, and temperature of the ions all affect this ability. However, TDS referred as a crucial indicator of water quality that is frequently used to evaluate whether water is suitable for irrigation and drinking. Another important characteristic of any body of water is its oxygen level, which is necessary for all biological life forms and has a significant impact on the solubility of metals. However, man-made sources including home sewage, septic tanks, and agricultural practices may also be to blame for the high concentration of this substance in water. It is demonstrated that there is a positive association between the value of TDS and electrical conductivity, with surface water often having higher TDS than subsurface water. Based on the investigation, it was determined that the TDS varied between 82-198 mg/L. In contrast, around 100 mg/L of total dissolved salts are allowed in drinking water. The high percentage of dissolved solids in the surface water at W-(9) indicates that there is a high concentration of suspended materials in the runoff and significant anthropogenic activity along the river's course. Since sodium is one of the most plentiful elements on earth and its salts are extremely soluble in water, sodium may be found in all natural waterways. Also, salinity variations have an impact on the quality of water used for drinking and irrigation, change in Na^+ and $MgSO_4$, is often leads to a gastrointestinal effect. This is because the reaction between Na^+ and soil causes particle blockage, which lowers permeability. Furthermore, an alkaline soil created by high Na^+ concentrations has an excess of Na^+ in the water, which alters the properties of the soil and causes crust formation, water logging, decreased aeration, decreased infiltration rate, and decreased permeability. The WHO recommends a 200 mg/L maximum for Na^+ in drinking water. Applying this index to the samples indicates that 100% of the surface water belongs to suitable category with Na^+ ranging from 1.58 to 4.63. However, no location falling under doubtful or unsuitable category. This important indicator's occurrence namely, (K^+) in all water bodies, is roughly 1/10th to 1/100th that of Na+, which might be caused by its weak capacity for migration and resistance to weather-induced breakdown. The K^+ value (2-10.2) obtained is within the tolerable limits (12 mg/L). The minimum and maximum value are W-(1) and W-(8). The mean concentration value is around 6.88 mg/L, which implies that because this parameter is actively absorbed by plants, it is rarely present in waterways at high concentrations. High levels of it can occasionally be a sign of pollution and are mostly to blame for eutrophic conditions. Additionally, it is believed that runoff from fertilizers, industrial effluents, and household wastewaters, especially those that include detergents, that contributes to surface water elevation. The reported higher value is noticed at W-(8), in which the permeability and soil composition could be negatively impacted, because of the elevated K^+ concentration, which causes alkaline soils to occur. On the other hand, Ca^{2+} is quite prevalent in the natural world. The main purpose of this mineral salt is to help the bone mineralize by forming calcium phosphate salts. It implies that it is thought to be the fifth most common element in nature, dissolved from rocks and soils, and the primary cause of water hardness. Rocks and minerals like calcite are its most frequent suppliers in water. Additionally, geological sources, agricultural waste, and landfill trash in drinking water could be the main causes. Insufficient calcium intake can lead to osteoporosis, tooth defects, nephrolithiasis namely, kidney stones, rickets, hypertension, and stroke. However, the value of calcium content varied between 9.43 and 28.18. Since the desired level of 75 mg/L for calcium fails to be reached in this study region, there was no calcium concentration at any of the locations in the several analyzed sites. In addition, the main agents like sewage, industrial wastes, and a variety of rocks are the main sources of Mg^{2+} in drinking water. The most frequent source is the weathering of minerals like magnesite and rocks. Elevated magnesium levels have been linked to both high water hardness and cardiovascular disease. This is due to the fact that Mg^{2+} uptake at cation exchange sites in the soil, particularly in highly saline and extremely high Na^+ waters, causes the dispersion of soil aggregates and loss of soil structure. The officially allowable level for drinking purposes is 30 mg/l. The reported Mg^{2+} vales ranged from 14.83-28.72 mg/l. The bulk of all the samples is below the allowable limit. . Compared to the St-(1), the samples' hardness varies more in W-(2) to (3) and also, W-(14) to (15), respectively. However, the rise in the concentration of Mg^{2+} in water may be attributed to the fact that Mg^{2+} has less biological activity than Ca^{2+} and because $MgSO_4$ and hydro-carbonate are more soluble than Ca^{2+} equivalent molecules. However, the sites W-(3), (7), (9), and (15) exceeded the suggested drinking limit by a small amount, as stated by WHO, and for irrigation, the recorded value is estimated as 25 mg/L. A high score is displayed at 4 sites, that clarifies the additional impact of fertilizers. Mainly, the chief source of HCO_3^- depends on home sewage, the partial pressure of CO_2, and the existence of carbonate minerals in rocks and soil. But the creation of CO_2 that raises the concentration of HCO_3^- accompanies denitrification, which uses organic carbon as the electron donor. Meanwhile, the CO_2 occurs when minerals found in the soil combine with the rock to generate bicarbonate, which causes the surface water to become alkaline. The reported score in the current investigation, lies in a range from 41.92-87.55 mg/l. The maximum permitted value is 100 mg/l, according to the WHO. It is created when water and carbon dioxide react with carbonate rocks. Generally, The main ions present in water, are mostly produced by atmospheric deposition, chemical weathering, and human waste from sewage and industry that ends up in rivers. On the other hand, chloride is a widely dispersed element that may be found in all kinds of rocks. It has a strong affinity for salt. Because of this, the concentration is higher in ground waters with higher temperatures and less precipitation. Water acquires an unpleasant and bitter taste and distribution network oxidation is possible, when Cl^- concentration exceeds 200 mg/L. Fortunately it was discovered that the samples' Cl-content, which ranges from 9.66 to 271 mg/L, was well below the 250 mg/L acceptable limits. This season's high hardness value, which indicates the existence of hazardous related cations released from corroded pipes, is the cause of the high value of Cl^- ions at W-(9). Additionally, this rise was a result of farmers using sanitizers excessively. Likewise, sulphate (SO_4^{2-}) exists as SO_4^{2-} spontaneously in surface water. The amount of sulphate in water rises as a result of household and industrial waste. Meanwhile, some possible sources include biological processes, human activity, Sulphur minerals, and heavy metal sulfides, that are commonly found in igneous and metamorphic rocks. In addition to the natural sources, fertilizers can be used to add it. The current study's results varied within a frame of 4.97-217 mg/l. Nevertheless, the average sulfate concentration is within the acceptable range of 200 mg/L at about 26.43 mg/L. With an overabundance of Mg^{2+} in the surface water, the sulphate concentration becomes unstable and has a laxative impact on the human system, when it surpasses the maximum allowed limit at W-(9). Additionally, too much of it at W-(9) results in unpleasant color, taste, odor, discoloration, and corrosion. Here, some parameters shows an increasing trend. As obtained from the median values of chemical datasets from samples of surface water, Mg^{2+} and HCO_3^{1-} are the most prevalent anion and cation, respectively, in samples of surface water. Consequently, the variable that governs the dominance order of cations inside the research region is denoted as $Mg^{2+} > Ca^{2+} > K^+ > Na^+$, whereas the anions is found as: $HCO_3^- > Cl^- > SO_4^{2-}$. When the river flows downstream, the discharges from the different cities along its length mix with the river water, raising the pollution levels. As a result, some of the downstream variables show a declining trend. Ultimately, the analytical findings were used to produce the numerical distribution of the study's parameters, as was evident from the Figure 5a-l.



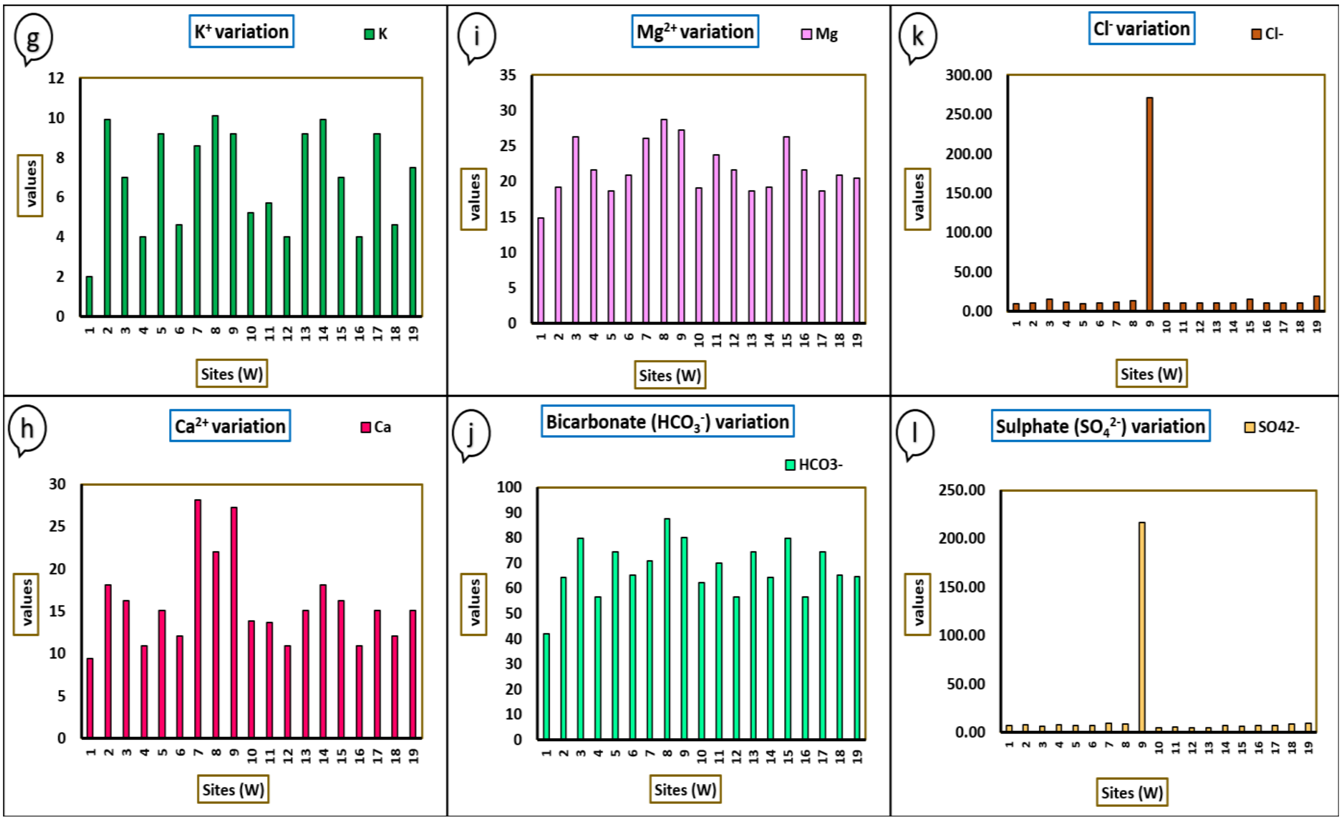


Figure 5. Abundance in WQ parameter's amount over selected stations found in surface water for (a) pH, (b) Turbidity, (c) TKN, (d) EC, (e) TDS, (f) Na^+, (g) K^+, (h) Ca^{2+}, (i) Mg^{2+}, (j) HCO_3^-, (k) Cl^-, and (l) SO_4^{2-}

As a result, it is significant to remember that this study used the six models to assess the river water quality and that factor analysis was used to identify the sources of contamination. This allowed recommendations to be made regarding the necessary treatment measures to ensure that the water is safe for inhabitants to drink. Geospatial techniques, like IDW, use nearby known data to estimate an unobserved position while interpolating the spatial variability of point attributes. Therefore, utilizing normally distributed data, it serves as the most effective and broadly applicable method for predicting unknown values. However, the chemical characteristics and indices could be used to differentiate between the acceptability of water quality for different uses, such as irrigation or drinking. Using the methods below, managers and administrative organizations can rate water according to its quality and usage while taking into account the effects of specific parameters. This information can be useful when making decisions. As a result, the new research employing the indices below will lay the groundwork for future initiatives to regulate river growth and save coastal ecosystems.

3.2 Numerow Pollution Index ($(N)_{pi}$)

The $(N)_{pi}$ has been employed to evaluate the general quality of the water when comparing sites over time and space. It has been shown that the IDW interpolation technique, along with GIS-based spatial analysis, is a highly effective tool for representing the distribution of significant ions in the research area. For the 12 water quality parameters, it was computed using the water quality metrics that were reported. When the result is determined by means of using $(N)_{pi}$, absolute index scores, which range from 0 to 40 and have values between 0 and 10 to indicate the best overall state in terms of safeguarding aquatic life for human use. In this work (Figure 6a), the 5.8 to 69.21 obtained range denotes a good to unsuitable category. It is revealed that sixteen samples fall under excellent category, two samples appears in poor category, and the final one located is based on the results, that is categorized as unsuitable. Here in Figure 7a, around three stations had water quality values higher than 40, indicating poor or inappropriate water quality, although the majority of the stations had excellent water. Thus, this may be the outcome of increasing levels of anthropogenic and diverse pollutants gradually building up in the river water. Nonetheless, this index evaluates the element of water quality related to ecological projects and offers a useful way to summarize intricate water quality information for architects, managers, lawmakers, and wholesalers.

3.3 Synthetic Pollution Index ($(S)_{pi}$)

According to the results of grading based on $(S)_{pi}$ after applying, these findings could vary extensively. Thus, to draw water quality of stations under study, IDW zoning in Arc GIS environment were utilized. In this current area (Figure 6b), indicating a good to extremely poor category, the calculated range is 0.25 to 0.99. However, such changes could normally be seen in 17 sampling points, and hence, could be classified under good quality prioritization. Remaining 2 locations comes under poor/very poor category. The study found that the persistence of declining water quality was caused by a comparatively high quantity of contaminants exceeding potability criteria. Therefore, the poor water quality in this part of the river at W-(19) and (9) in Figure 7b, was a result of the inadequate measures put in place for appropriate disposal prior to the city's urbanization. Additionally, careless waste dumping in marketplaces, drainage systems, and urban rubbish all significantly worsen river pollution. Indeed, 16.70% (2 indicators i.e., TC and turbidity) do not adhere to water quality regulations, which could have a detrimental impact on the health of the populace and aquatic life. Yet, researchers also came to the conclusion that an essential mineral and organic load with anthropogenic origins from metropolitan areas represented the deterioration of the water quality.

3.4 Overall Index of Pollution ($(O)_{pi}$)

In case of this approach, it leads us to conclude that no surface water body's problematic situation is indicated by analysis, monitoring, or index development alone. As a result, while conducting additional study, it will be simple to determine the sampling locations' relative pollution levels in relation to the drinking water quality standard using $\mathbf{(O)_{pi}}$ approach. The obtained range is around 0.29 to 15, demonstrating a classification of excellent to contaminated (Figure 6c). About sixteen examples are included in the "excellent" zone, and the remaining three samples are a result of "polluted" category, as obtained from the results. The upstream areas close to the sampling sites show evidence of improving water quality. Taking this into consideration, the outcome can be attributed to increased rainfall during the study period, which caused water to seep into the ground and dilute the surface water system. With regard to the duration of the research, at three sites namely W-(8), (19) and (9), it is noticed from the **Figure 7c**, that high-polluted areas with the highest nutrient concentrations were discovered close to industrial and built-up areas. Analysis and surveys of land use and land cover revealed that runoff from agricultural areas has had the greatest impact, followed by wastewater from homes and businesses.

3.5 Comprehensive Pollution Index ($(C)_{pi}$)

As detailed in $\mathbf{(C)_{pi}}$ section in methodology part, this approach often depends on a number of variables, including the geolocation, the physicochemical characteristics of the water to be assessed, eutrophication, health issues, oxygen accessibility, suspended constituents, the planned use of the water, ecological significance, and data accessibility. As a result, it also involves a methodical approach to standardize and formalize the parameter selection procedure for widespread use. The value of $\mathbf{(C)_{pi}}$ ranged between 0.29-3.46, reflecting that the water quality fall into 'sub-clean to-seriously polluted' category. The results (**Figure 6d**) show that higher value was recorded at W-(7), which reveals that water has organic pollution, where there are crowd industrial parks and busy transport and fishing ports. On the basis of its ratings, out of 19 samples, 16 are categorized as sub-clean, by the score of $\mathbf{(C)_{pi}}$**,** 2 as polluted, and 1 is under the seriously polluted zone of water, in the current investigation. This furthermore, revealed a possible danger of nitrogen contamination, TDS, Cl^- and Ca^{2+}, that are connected to human activity and urban discharges. A high positive loading of Ca^{2+} and EC is ascribed to a number of natural processes, including the weathering of rocks that contain calcium carbonate (Calcite) and limestone, as well as other ion-exchange activities occurring in the surface water system. It is observed from the spatial diagram as given in **Figure 7d**, that unsuitable sites have been identified at 3 places. This primarily be connected to wastewater, excreta, and also, fluoride in drinking water, gets influenced by human activity, farm fertilizers near survey stations, and waste from households.

3.6 Entropy (E_y)-Water Quality Index (WQI)

To increase the measurement accuracy, it is suggested that the weights for the models mentioned above be changed based on the model application. As a result, when the same model is used for marine and river water bodies, the parameter weights should be different. Therefore, entropy is a single value that can be used to describe the water quality of a source, reducing the multitude of factors to a single, straightforward face that makes water quality observation data easier to grasp. River sediment contamination zones can be categorized using GIS and the spatial interpolation approach, allowing for improved river and catchment management. In carrying out a further study (Figure 6e), the E_y-WQI ranged from 14.60 to 277, that depicts excellent to extremely poor categories.

Approximately, 84.2% of the sample locations exhibit excellent conditions, 10.52% of the sample locations exhibited poor conditions and 5.26% of the sample locations (1 location) indicated extremely poor water conditions. It is possible that the elevated levels of Ca^{2+}, SO_4^{2-}, EC, and TDS in the surface water were the primary causes of these elevated readings, indicating poor quality drinking water and high salinity. The presence of TKN in water is often indicative of a partially completed organic matter degradation process and is a reliable marker of river pollution caused by urban effluents. Based on the findings, 16 samples are classified as excellent, and 3 samples are classified as poor or extremely poor. According to the interpolated results, which are shown in Figure 7e, the demarcations showed the spatial extent of the affected area which was more exposed to pollutants. However, sites at W-(7)-(8) and (9), and W-(18) to W-(19), revealed a maximum change in the entropy value, suggesting that runoff is not a major factor in the river pollutant's dilution in this area. The WQI value changed far less in the remaining 16 regions, therefore after the monsoon, the quality of the water was still being evaluated in these places. The reason for the preservation of the equilibrium between input and output ions could be attributed to the increased release of sewage and industrial waste into the water during monsoon season [78]. Ultimately, the parameter's variation clearly indicates that three sites lies in the unsafe category. Being a seaside region at W-(9), one of the main causes of the decline in the quality of the water that is fit for human consumption in some areas is the incursion of saltwater.

3.7 Integrated-Weight (Int_w)-Water Quality Index (WQI)

In terms of Int_w approach, weights are obtained by two approaches, namely, entropy and the CRITIC Method. To calculate WQIs, this method could be applied to determine the optimal parameters. When there are a lot of output variables and a high degree of collinearity and noise in the input variable data, this is a useful way to handle the data. As a result, both the quantity and cost of the components utilized in the chemical analysis to determine WQIs are decreased. One popular method for visualizing a linear relationship between large independent and dependent parameters is termed as the Integrated-weight approach. Hence, it can be investigated as a different method of predicting indices in this research area because it is quick, easy to use, and only requires a few steps to compute, especially for the WQIs. The values for the present study ranged between 39 and 310, with a mean of 101.37, based on which Int_w-WQI map was prepared as shown in Figure 6f. It is apparent from the figure that out of 19 samples, 52.63% of the samples falls under 'excellent (n=10)', 5.26% samples belong to class 'good (n=1)' and 42.11% of locations, comes under 'poor/very poor/not suitable (n=8)' category, indicating their adverse effects on human consumption. The river receives the discharge of domestic sewage from these locations since treatment plants cannot handle the volume of sewage generated in cities. Untreated sewage wastes have reportedly been dumped into the river, which may be the main source of pollution in the area around it. Subsequently, it is occasionally observed that in low-quality sites, variations in the water system's turbidity might be caused by organic matter decomposition, carbonate degradation, temperature and salinity changes, and the mixing or dilution of effluents by aquatic streams. However, SO_4^{2-} and TKN indicator are closely related to human activity and urban discharges. Its two main sources are home effluent (mostly detergents) and agricultural operations, in addition to the reductive hydrolysis of minerals that include phosphate. Based on the spatial analysis (Figure 7f), Contaminated zones within the study area were classified and located. Moreover, Na^+, Cl^- and Mg^{2+} indicates that anthropogenic inputs and industrial activity have a combined effect on aquatic systems that cannot be disregarded. Furthermore, because the research location is a low-lying flood plain, the neighbouring hilly areas' sediment erosion, transport, and deposition may have an impact on the indicator's concentration discharged into the riverine system over a wide range of geochronological scales. Thus, to distinguish the region based on pollutants, around 42.11% of the research area must be maintained before to use.

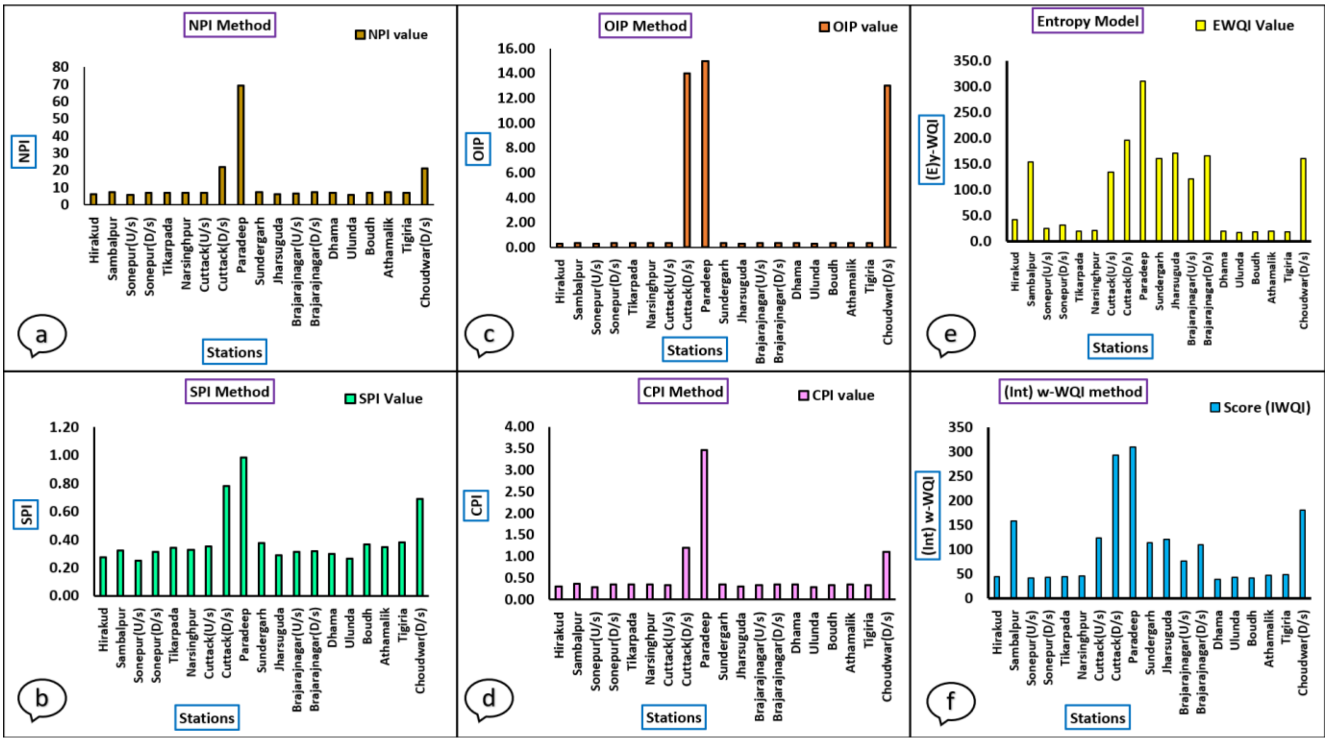


Figure 6. Layout of site's variation results for (a) NPI, (b) SPI, (c) OIP, (d) CPI, (e) E $_{y}$-WQI and (f) Int $_{w}$-WQI

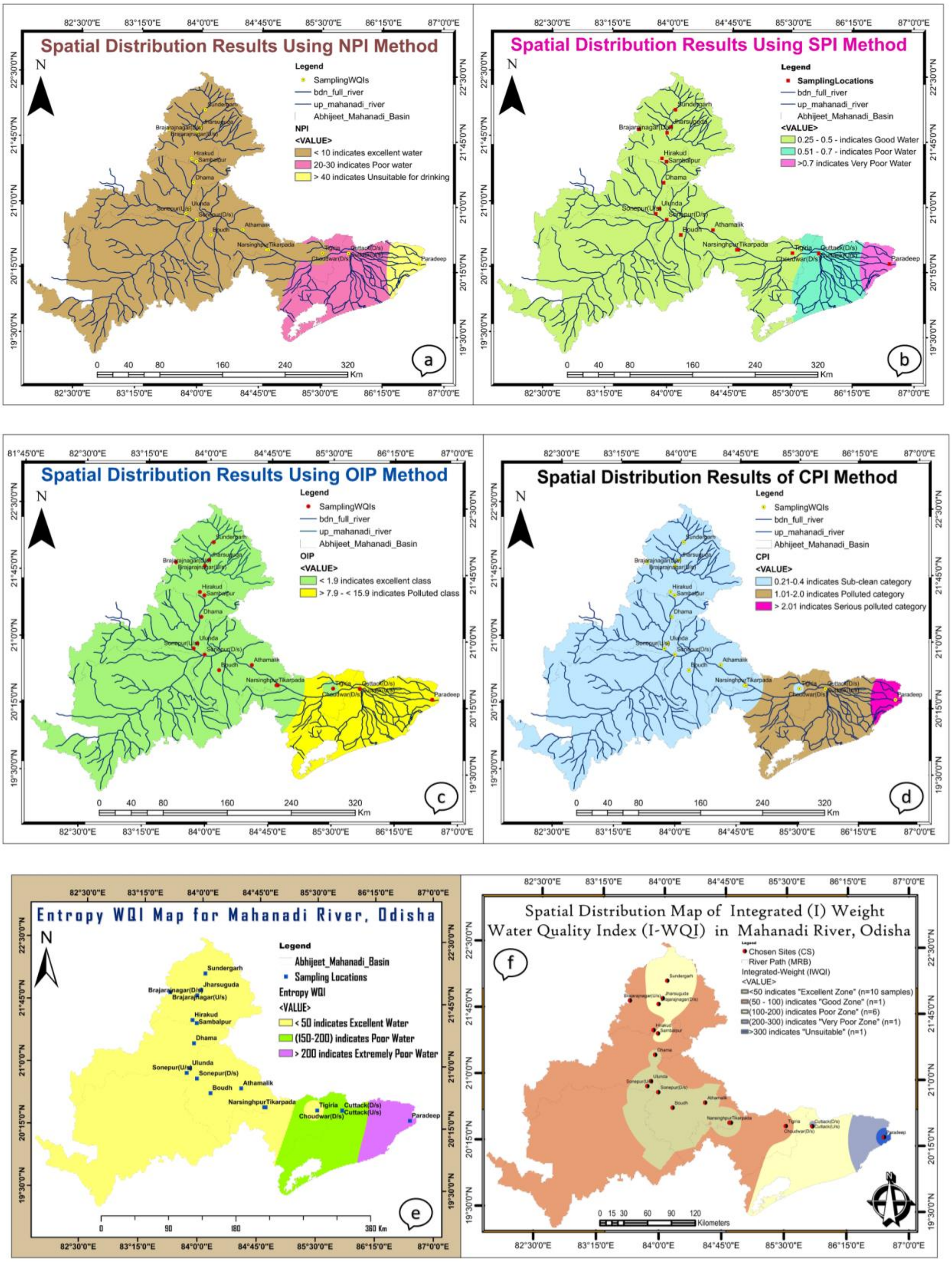


Figure 7. Layout of interpolated results for (a) NPI, (b) SPI, (c) OIP, (d) CPI, (e) E $_y$-WQI and (f) Int $_w$-WQI

3.8 Factor analysis (F_a)

Since, the water quality index approach along with pollution indices methodology has some limitations and uncertainty. So, the researchers claim that utilizing improper aggregating functions or sub-indexing methods that improperly distort the genuine relative influences of the parameters are the root causes of this issue. Therefore,

adopting factor analysis (F_a) ought to be made in order to reduce the eclipse issue and the uncertainty of the WQ model. However, the conventional six WQ approaches have the capacity to visualize comparisons, but in addition **F_a** gives semi-quantitative data on how WQI factors are distributed within the sampling space using a variable or components. However, by emulating the ambiguity and uncertainty brought on by the many criteria, it excels at managing complicated choice scenarios. However, the data were subjected to factored rotation. Consequently, the recorded value based on Kaiser Meyer Olkin (KMO) principle, for water in the current study, was estimated to be as 0.364 and the significant value of Bartlett's sphericity was found as 0.01, indicating **F_a** was profitable in the studies. Therefore, by using this technique, the original datasets were changed to create a new, uncorrelated collection that resembled a linear combination of the original sets. Additionally, it aids in the information extraction process by removing less significant factors with the least amount of original data lost. To determine how many PCs are needed to understand the underlying data structure, a scree plot has been created. For this purpose, the eigen value is included that illustrates the amount of variance that the factored component may account for, and as previously stated, an eigen value > 1 means that one factor component (C), could provide many explanations for a variable. Therefore, we considers only components with an eigen value greater than 1. In this research, the scree plot graph (**Figure 8**) demonstrates the substantial decrease in slope that occurs after the fifth eigen value, indicating that only five components remain. From the **Table 1**, it is noticed that first five components (C1-C5), based on the eigen value, together account for 95.61% of the total variance in the dataset, in which the first component (C1) exhibit 40.32% of the total variance, the second component (C2) explains 20.52%, the third component (C3) exhibit 15.76%, the fourth (C4) and fifth component (C5) is around 11.56% and 7.45% of the total variance respectively. The most fascinating factor configurations were for C1, where a strong correlation exits between turbidity, TKN, Ca^{2+} and Cl^-, were displayed from the dataset. However, hardness consists of Ca^{2+}, Mg^{2+}, CO_3^{2-}, and HCO_3^- and is diluted in water from frequent washing, the use of detergents, domestic sewage, and industrial effluent discharge. Furthermore, a number of variables, including the use of fertilizers, manures, pesticides, herbicides, leaking septic tanks and sewage lines, wastewater discharge from various sources, and accumulation of livestock waste, contribute to the buildup of TKN in surface water. Here, Cl-may be caused by the local geology, agricultural runoff, industrial and residential effluent, and agricultural runoff. Notably, these high levels of these parameters can be associated with commercial and electroplating sectors, as well as industrial processes and agro-fertilizer application. Those who lived close to the river and used its water for domestic purposes during the time reported an outbreak of cholera in the water, which was further evidence of nitrogen pollution. In 2nd component, this factor has high loading on pH, EC, and Na^+. The pH expresses a strong relationship indicating that nitrification and wastewater recharging are the crucial processes. Here, nitrification's acidic protons lower pH and cause mineral breakdown. Usually, carbonate dissolution in the zone neutralizes an acidic proton instead of silicates. Additionally, there is a connection between the change in EC and the organic matter content, which during the dry season is linked to discharges of home and commercial wastewater. As this is going on, the high concentration of algae encourages the generation of toxins by raising the turbidity, Na^+, and DO levels of the water. These circumstances limit the utilization of this resource for public supply and are detrimental to aquatic life. Again, in 3rd Component (C3) includes TDS and K^+. On the other hand, 4th component (C4) had strong positive loadings on Mg^{2+} and HCO_3^- and finally, component (C5) exhibits a load with SO_4^{2-}. Here, TDS perhaps linked to surface runoff, bank erosion, and erosion from upland regions. The primary non-point sources of K^+ in this area are runoff from regions used for agricultural fertilization and storage for animal waste. Mg^{2+} validates the silicate minerals' dissolution, raising the surface water's hardness. It is well knowledge that activities like leaching and overuse of fertilizers will have an impact on the quality of surface water when HCO_3^- values are elevated. Water samples could therefore be impacted by related agricultural practices and man-made activities [73]. Additionally, SO_4^{2-} demonstrates how human activity affects the drinking system. Consequently, it also involves the breakdown of carbonate minerals and mixing of seawater. Therefore, these five elements can be understood as the surface water's mineral component, a combination of natural and man-made sources of contamination (soil erosion or weathering, air deposition, biofertilizers, runoff from ore producers, etc.), and saline water invasion, which is frequently observed in coastal aquifers due to excessive surface water extraction brought on by rapid population growth, commercial activity, and other factors. The process of determining the factor's representation's zone and spatial variation is done in order to calculate factor scores. The two techniques most commonly used to obtain them are the regression approach and the weighted least squares method [74]. The factor score (FS) plots (**Figure 9**) explains that urban wastewater has a more noticeable effect on sampling sites in the downstream course that are situated on the basin. The nutrient group of pollutants, which includes anthropogenic and non-anthropogenic point sources like municipal effluents and non-point sources like erosion and agricultural runoff, that made a major impact [75]. The most influenced are sampling locations located on the downstream course of the river i.e., W-(8), (9), and W-(19), followed by the tributary in the middle course.

Finally, the study's findings demonstrate that F_a, a powerful multivariate statistical tool, allows for the evaluation of correlations between variables while reducing a dataset's dimensionality while preserving as much of its variability as feasible [76]. This result suggests that the parameters in the river have comparable origins or are the result of human-induced activities. Conclusively, the results of this F_a analysis are consistent with the previously indicated conversations regarding the assessment of water pollution and the clarification of the sources of hazardous factors present in the areas of the investigated watershed [77].

Table 1. Extracted factor loadings of the studied parameters of the surface water

Parameters	**C1**	**C2**	**C3**	**C4**	**C5**
pH	0.56	0.88	0.44	0.23	0.34
Turbidity	0.87	0.45	0.32	0.43	0.15
TKN	0.93	0.37	0.63	0.51	0.59
EC	0.72	0.85	0.68	0.31	0.47
TDS	0.66	0.71	0.79	0.47	0.61
Na^+	0.43	0.77	0.65	0.67	0.669
K^+	0.56	0.68	0.83	0.59	0.54
Ca^{2+}	0.95	0.57	0.34	0.66	0.43
Mg^{2+}	0.61	0.44	0.32	0.85	0.67
HCO_3^-	0.55	0.66	0.45	0.91	0.43
Cl-	0.88	0.51	0.28	0.44	0.37
SO_4^{2-}	0.41	0.61	0.17	0.53	0.84
Eigen value	8.61	5.67	3.28	2.54	1.74
% variance	40.32	20.52	15.76	11.56	7.45
Cumulative %	40.32	60.84	76.6	88.16	95.61

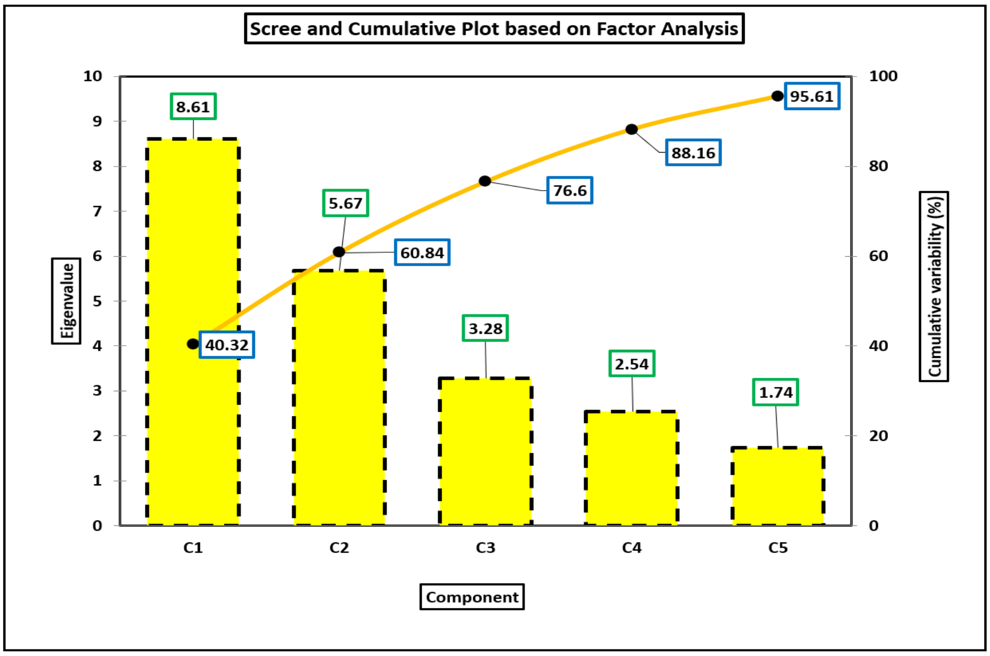
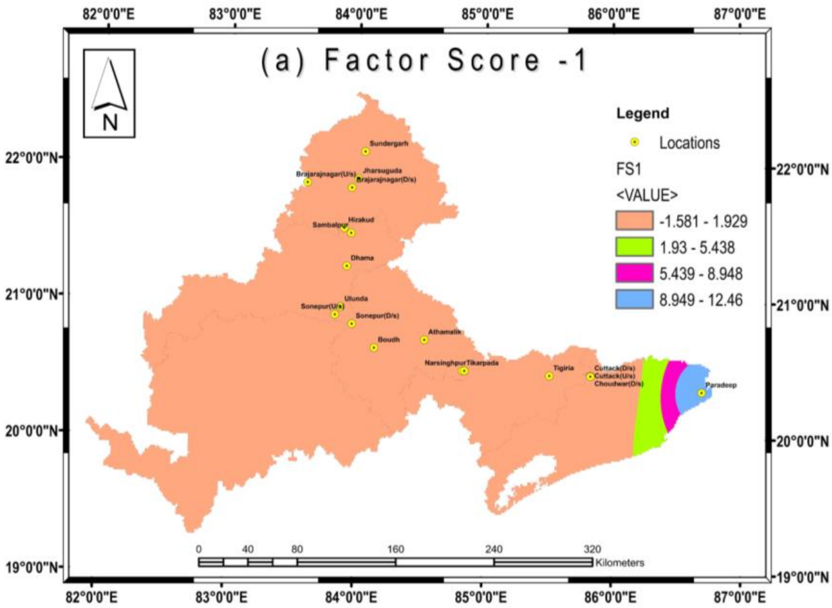
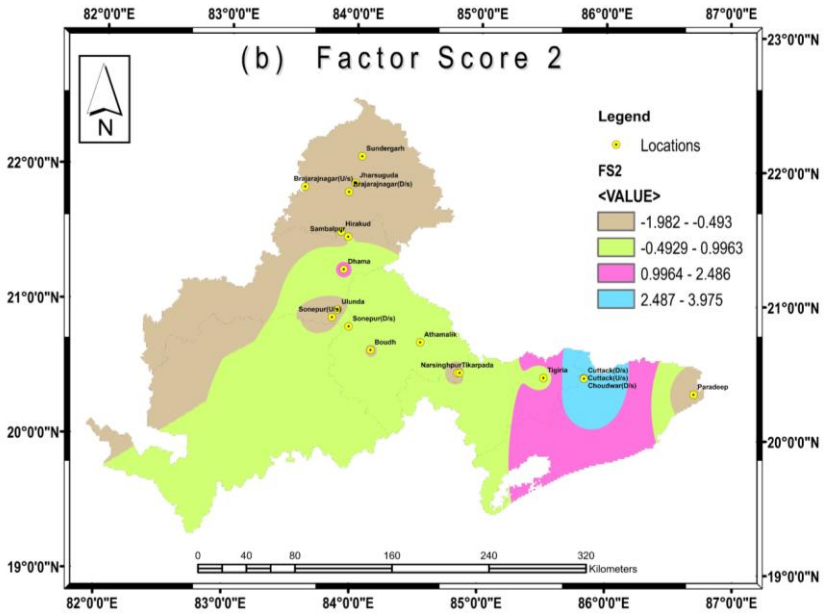
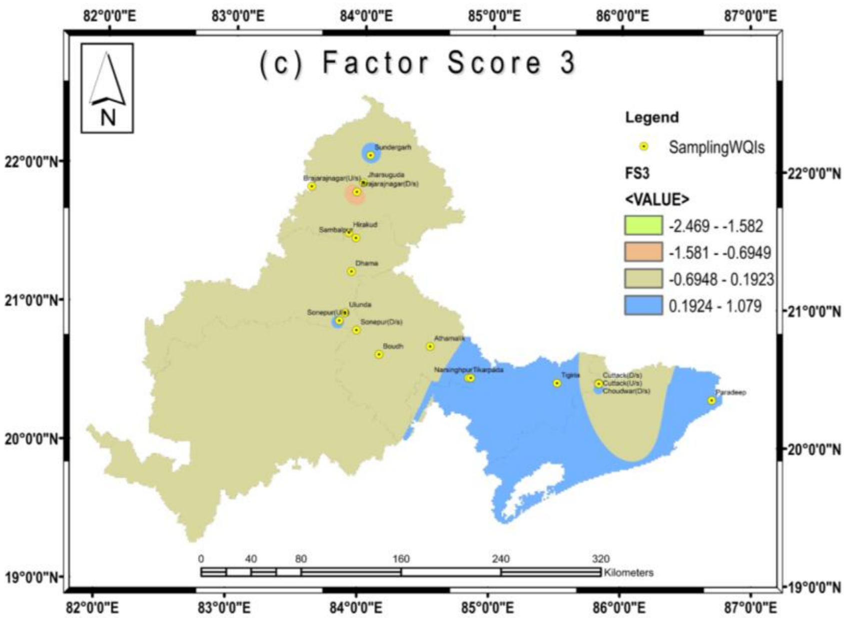


Figure 8. Scree plot of variance of factored components




(b) Factor Score 2
Legend
Locations
FS2
<VALUE>
-1.982 - -0.493
-0.4929 - 0.9963
0.9964 - 2.486
2.487 - 3.975
N
Sundergarh
Brajarajnagar(U/s)
Jharsuguda
Brajarajnagar(D/s)
Sambalpur
Hirakud
Dhama
Sonepur(U/s)
Ulunda
Sonepur(D/s)
Boudh
Athamalik
Narsinghpur
Tikarpada
Tigiria
Cuttack(D/s)
Cuttack(U/s)
Choudwar(D/s)
Paradeep
0 40 80 160 240 320
Kilometers
82°0'0"E 83°0'0"E 84°0'0"E 85°0'0"E 86°0'0"E 87°0'0"E
23°0'0"N 22°0'0"N 21°0'0"N 20°0'0"N 19°0'0"N


(c) Factor Score 3
Legend
SamplingWQIs
FS3
<VALUE>
-2.469 - -1.582
-1.581 - -0.6949
-0.6948 - 0.1923
0.1924 - 1.079
N
Sundergarh
Brajarajnagar(U/s)
Jharsuguda
Brajarajnagar(D/s)
Sambalpur
Hirakud
Dhama
Ulunda
Sonepur(U/s)
Sonepur(D/s)
Boudh
Athamalik
Narsinghpur
Tikarpada
Tigiria
Cuttack(D/s)
Cuttack(U/s)
Choudwar(D/s)
Paradeep
0 40 80 160 240 320
Kilometers
82°0'0"E 83°0'0"E 84°0'0"E 85°0'0"E 86°0'0"E 87°0'0"E
22°0'0"N 21°0'0"N 20°0'0"N 19°0'0"N

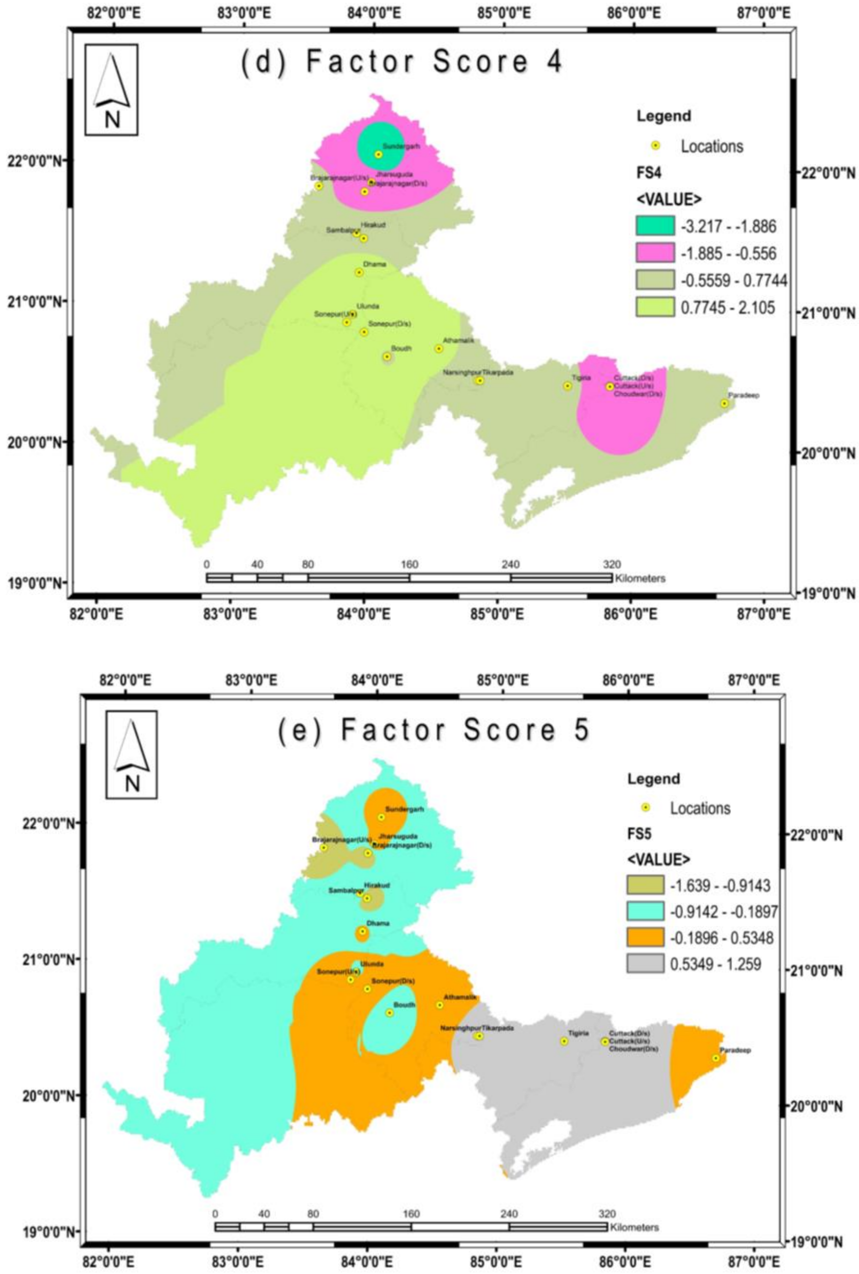


Figure 9. Spatial distribution of interpolated results of F_a scores for sampling site's upstream to downstream (a) FS1, (b) FS2, (c) FS3, (d) FS4 and (e) FS5

4. Conclusions

In the present work, a study has been taken up to assess the performance of NPI, SPI, OIP, CPI, $(E)_y$-WQI, Int $_w$-WQI and Multivariate algorithms for predicting the surface water quality. A GIS is an effective tool for interpreting spatial data and producing visual representations that yield outcomes. It is related to the multivariate approach used in this work to determine the factor loadings for the parametric layers. The comparison of these techniques has been made and the following conclusions are drawn from this work, are as given below:

1. pH results show that the river water is slightly alkaline, buffered by a HCO_3^{-}–CO_3^{2-} system.
2. The turbidity and EC values were high in the majority of the places, making the water unfit for irrigation and drinking unless the soil has sufficient drainage, a particular management strategy for controlling salinity is implemented, and salt-tolerant plants are applied.

3. Results further demonstrate that the river flows in the category of fresh (TDS<100 mg/L) except two sites (W-(9) and (19)), though may be categorized as either sensitive or non-sensitive depending on the river's alkalinity levels.
4. According to Na^+ values, the water can be used for drinking, with little risk of dangerous amount of exchangeable salt developing (low sodicity).
5. HCO_3^- and K^+ values show that the river might be regarded as having not received any major or extremely severe pollution emissions. The river may not be able to sustain aquatic life due to pollution, as indicated by the turbidity and TKN levels.
6. The river exhibits an overall pattern of ionic dominance, based on the calculated means: $Mg^{2+} > Ca^{2+} > K^+ > Na^+$ and $HCO_3^- > Cl^- > SO_4^{2-}$.
7. The pollution indices of 19 samples in the present study ranged from 5.8 to 69.21, indicating good to unsuitable category ($(N)_{pi}$), 0.25 to 0.99, signifying good to poor water ($(S)_{pi}$), 0.29 to 15, displaying a classification of excellent to contaminated ($(O)_{pi}$) and 0.29-3.46, reflecting that the water quality fall into 'sub-clean to-seriously polluted' category, in case of $(C)_{pi}$. Among the total samples, only three samples indicated poor water quality in case of $(N)_{pi}$. Around 16 samples come in the "excellent" zone, and the rest three samples belongs to "polluted" category, according to the results of ($(O)_{pi}$). The following 2 samples exceeding the norms as per ($(S)_{pi}$), indicating water to be unfit for consumption. The results of $(C)_{pi}$ highlights that higher value was recorded at W-(7), which exhibits the organic contamination in water. The discharge of sewage into rivers is the cause of the elevated levels of turbidity, TKN, and Ca^{2+} in the water that lower drinking water quality, as indicated by the following pollution indices.
8. To fully understand the river's contamination levels, a $(E)_y$-WQI was computed utilizing twelve physicochemical factors. It is worth mentioning that at sixteen places, it was determined to be of outstanding quality, since the water there is in an operating condition with no obstacles to its flow and the ability to clean itself, whereas at downstream sites (n=3), because of the area's dense population and the wastewater that the river receives from drains, the water quality was found to be poor.
9. Over 42.11% of the samples fell into the categories of poor/very poor/not suitable, using the Int_w-WQI diagram, establishing that the surface water is unfit for human consumption.
10. The advantages of this pollution approaches is that it creates a mathematical equation that represents the ecological status of the water by combining information from multiple surface water quality metrics. It also demonstrates the importance of each metric in assessing and managing the water quality and helps determine whether a surface water source is suitable for human use.
11. F_a is used to identify and extract the variables causing variations in the quality of the water, and it reveals that the first five principal components explain almost 95.61% of dataset variation. After the fifth eigen value, the slope begins to decline slightly. The first component takes 40.32% of total variance, indicating that C1 alone can statistically represent almost half of the group. The factors that had the biggest effects on the quality of the water were found to be turbidity, TKN, Ca^{2+} and Cl^-, that made it possible to formulate the parts made of the five eigen values. The second component (C2) contains three parameters and explains 20.52% of the total variance. This component has proposed that their primary source was a combination of geogenic and anthropogenic causes caused by humans. The third and fourth component (C3, and C4), which each has two parameters, explains 27.32% of total variance. Finally, the fifth component (C5) has a single variable that accounts for 7.45% of the set, giving it a sufficient representation. However, the presence of nutrients is influenced by three other factors, though particulate matter, which is also in charge of the river's natural mineralization processes; hardness and alkalinity processes, as well as their connection to the processes involving algal biomass; the dependence of Mg^{2+} and TKN concentrations on pH; and the interactions between oxygen, nutrients, and water transparency. These places are all considered ideal for artificial surface water recharge. It is also reflected that a low surface water potential location is one with significant runoff but poor infiltration due to geo-environmental variables.
12. It is therefore, advised that suitable monitoring plans be started by the pertinent organizations, including Water Resources Commission (WRC), which is the organization in charge of managing the river basin in order to maintain the river's ecosystem. In order to maintain the sustainable management of surface water resources, the current inquiry fervently advocates for the effective implementation and administration of effluent treatment facilities and restricts any direct and indirect cause of contamination.
13. The surface water potential zones were validated using the approach performed by F_a and the analysis reflected a very good corelation. The current method's reduced labour requirements, lower costs, and increased effectiveness over previous research methods are major advantages. The present study results

are beneficial for managing surface water resources, creating an artificial recharge structure, and fostering surface water sustainability in the current study region.

Acknowledgements

The author Abhijeet Das has been conceived of the presented idea and decided the structure of this article. However, I want to thank authorities of C.V. Raman Global University (CGU), Bhubaneswar, Odisha, for their encouragement towards this specific aspect and supported during the write-up. The complete manuscript's conceptualization, collected data from the field, GIS work, methodology, investigation, validation, , formal analysis, writing-original draft preparation, review, and editing, is being achieved solely by the corresponding author. Rani Prativa Das of Ruchika Square, Bhubaneswar reviewed this article and gave their valuable corrections and suggestions.

Data Availability Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declaration of Competing Interest

The author declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

5. References

1. Ahmed, R., Sreeram, V., Mishra, Y., Arif, M.D., 2020b. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. Renew. Sust. Energ. Rev. 124, 109792.
2. Rao, A.P., Sarman, K.G., Kumar, G.V.P., Yerra, S.D., 2022. Water quality monitoring using remote control boat. In: International Conference on Cognitive Computing and Cyber Physical Systems. Springer Nature Switzerland, Cham, pp. 201–212.
3. Martinell, D.P., Cashion, T., Parker, R., Sumaila, U.R., 2020. Closing the high seas to fisheries: possible impacts on aquaculture. Mar. Policy 115, 103854.
4. Naylor, R., Fang, S., Fanzo, J., 2023. A global view of aquaculture policy. Food Policy 116, 102422.
5. Zhou, X., Wang, J., Huang, L., Li, D., Duan, Q., 2022. Modelling and controlling dissolved oxygen in recirculating aquaculture systems based on mechanism analysis and an adaptive PID controller. Comput. Electron. Agric. 192, 106583.
6. Zhou, C., Zhang, J., 2023. Simultaneous measurement of chemical oxygen demand and turbidity in water based on broad optical spectra using backpropagation neural network. Chemom. Intell. Lab. Syst. 237, 104830.
7. Chai, W.S., Tan, W.G., Munawaroh, H.S.H., Gupta, V.K., Ho, S.H., Show, P.L., 2021. Multifaceted roles of microalgae in the application of wastewater biotreatment: a review. Environ. Pollut. 269, 116236
8. Yang, X., Xu, M., Zou, R., Angelidaki, I., Zhang, Y., 2021. Microbial protein production from CO2, H2, and recycled nitrogen: focusing on ammonia toxicity and nitrogen sources. J. Clean. Prod. 291, 125921.
9. Moloantoa, K.M., Khetsha, Z.P., Van Heerden, E., Castillo, J.C., Cason, E.D., 2022. Nitrate water contamination from industrial activities and complete denitrification as a remediation option. Water 14 (5), 799.
10. Nabi, M., Liang, H., Cheng, L., Yang, W., Gao, D., 2022. A comprehensive review on the use of conductive materials to improve anaerobic digestion: focusing on landfill leachate treatment. J. Environ. Manag. 309, 114540.
11. Iber, B.T., Kasan, N.A., 2021. Recent advances in shrimp aquaculture wastewater management. Heliyon 7 (11), e08283.
12. Liu, Y.Z., Chen, Z., 2023. Prediction of biochemical oxygen demand with genetic algorithm-based support vector regression. Water Qual. Res. J. 58 (2), 87–98
13. Cao, X., Liu, Y., Wang, J., Liu, C., Duan, Q., 2020. Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network. Aquac. Eng. 91, 102122.
14. Owolabi, T.O., Saleh, T.A., Olusayo, O., Souiyah, M., Oyeneyin, O.E., 2021. Modeling the specific surface area of doped spinel ferrite nanomaterials using hybrid intelligent computational method. J. Nanomater. 2021, 1–13.

15. Melesse, A.M., Khosravi, K., Tiefenbacher, J.P., Heddam, S., Kim, S., Mosavi, A., Pham, B. T., 2020. River water salinity prediction using hybrid machine learning models. Water 12 (10), 2951.
16. Nafsin, N., Li, J., 2022. Prediction of 5-day biochemical oxygen demand in the Buriganga River of Bangladesh using novel hybrid machine learning algorithms. Water Environ. Res. 94 (5), e10718.
17. Jiange, J., Liqin, Z., Senjun, H., Qianqian, M., 2023. Water quality prediction based on IGRA-ISSA-LSTM model. Water Air Soil Pollut. 234 (3), 172.
18. Nagaraju, T.V., Mantena, S., Azab, M., Alisha, S.S., El Hachem, C., Adamu, M., Murthy, P. S.R., 2023b. Prediction of high strength ternary blended concrete containing different silica proportions using machine learning approaches. Results Eng. 17, 100973.
19. Kolli, M.K., Opp, C., Groll, M., 2021. Estimation of soil erosion and sediment yield concentration across the Kolleru Lake catchment using GIS. Environ. Earth Sci. 80 (4),
20. Jagaba, A.H., Kutty, S.R.M., Hayder, G., Baloo, L., Abubakar, S., Ghaleb, A.A.S., Almahbashi, N.M.Y., 2020. Water quality hazard assessment for hand dug wells in Rafin Zurfi, Bauchi State, Nigeria. Ain Shams Eng. J. 11 (4), 983–999
21. Qambar, A.S., Al Khalidy, M.M., 2022. Prediction of municipal wastewater biochemical oxygen demand using machine learning techniques: a sustainable approach. Process. Saf. Environ. Prot. 168, 833–845.
22. Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model. 81, 401–418.
23. Alade, I.O., Rahman, M.A.A., Hassan, A., Saleh, T.A., 2020. Modeling the viscosity of nanofluids using artificial neural network and Bayesian support vector regression. J. Appl. Phys. 128 (8).
24. Wahla, S.S., Kazmi, J.H., Tariq, A., 2023. Mapping and monitoring of spatio-temporal land use and land cover changes and relationship with normalized satellite indices and driving factors. Geol. Ecol. Landscapes 00, 1–17. https://doi.org/10.1080/ 24749508.2023.2187567.
25. Gunturi, S.K., Sarkar, D., 2021. Ensemble machine learning models for the detection of energy theft. Electr. Power Syst. Res. 192, 106904.
26. Shahraki, A., Abbasi, M., Haugen, Ø., 2020. Boosting algorithms for network intrusion detection: a comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost. Eng. Appl. Artif. Intell. 94, 103770
27. Devi, V.A., Naved, M., 2021. Dive in deep learning: computer vision, natural language processing, and signal processing. In: Machine Learning in Signal Processing. Chapman and Hall/CRC, pp. 97–126.
28. Dong, J., Chen, Y., Yao, B., Zhang, X., Zeng, N., 2022. A neural network boosting regression model based on XGBoost. Appl. Soft Comput. 125, 109067.
29. Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 1–74.
30. Feng, D.C., Wang, W.J., Mangalathu, S., Hu, G., Wu, T., 2021. Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. Eng. Struct. 235, 111979.
31. Behera, J., Pasayat, A.K., Behera, H., Kumar, P., 2023. Prediction based mean-value-at- risk portfolio optimization using machine learning regression algorithms for multi- national stock markets. Eng. Appl. Artif. Intell. 120, 105843.
32. Nyirandayisabye, R., Li, H., Dong, Q., Hakuzweyezu, T., Nkinahamira, F., 2022. Automatic pavement damage predictions using various machine learning algorithms: evaluation and comparison. Results Eng. 16, 100657.
33. Massaoudi, M., Refaat, S.S., Chihi, I., Trabelsi, M., Oueslati, F.S., Abu-Rub, H., 2021. A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting. Energy 214, 118874.
34. Alabdullah, A.A., Iqbal, M., Zahid, M., Khan, K., Amin, M.N., Jalal, F.E., 2022. Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. Constr. Build. Mater. 345, 128296.
35. Nasir, N.M., Jusoh, A., Manan, H., Kasan, N.A., Kamaruzzan, A.S., Ghani, W.A.W.A.K., Lananan, F., 2023. Utilization of microalgae, Chlorella sp. UMT LF2 for bioremediation of Litopenaeus vannamei culture system and harvesting using bio- flocculant, Aspergillus niger. Biocatal. Agric. Biotechnol. 47, 102596
36. Tan, L.V., Tran, T., Loc, H.H., 2020. Soil and water quality indicators of diversified farming systems in a saline region of the Mekong Delta, Vietnam. Agriculture 10 (2), 38.

37. Ji, S., Wang, X., Lyu, T., Liu, X., Wang, Y., Heinen, E., Sun, Z., 2022. Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: a non-linear and interaction effect analysis. J. Transp. Geogr. 103, 103414.
38. Khanjani, M.H., Alizadeh, M., Sharifinia, M., 2020. Rearing of the Pacific white shrimp, Litopenaeus vannamei in a biofloc system: the effects of different food sources and salinity levels. Aquac. Nutr. 26 (2), 328–337.
39. Li, T., Lu, J., Wu, J., Zhang, Z., Chen, L., 2022. Predicting aquaculture water quality using machine learning approaches. Water 14 (18), 2836.
40. He, M.-Y., Dong, J.-B., Jin, Z., Liu, C.-Y., Xiao, J., Zhang, F., Sun, H., Zhao, Z.-Q., Gou, L.- F., Liu, W.-G., Luo, C.-G., Song, Y.-G., Ma, L., Deng, L., 2021. Pedogenic processes in loess-paleosol sediments: Clues from Li isotopes of leachate in Luochuan loess. Geochim. Cosmochim. Acta 299, 151–162. https://doi.org/10.1016/j.gca.2021.02.0 21.
41. Singh, R. B., & Patra, K. C. (2024). Integrated PCA–RNN approach for surface water quality assessment in the Mahanadi river system. International Journal of Environmental Science and Technology, 1-16.
42. Rajput, P., Sinha, M. K., & Nigam, G. K. (2022). Strategic Evaluation of Water Quality Monitoring Network Using GIS-AHP Model in a Large River System. In Water Resources Management and Sustainability (pp. 281-303). Singapore: Springer Nature Singapore.
43. Pati, S. S., Nayak, S., Mishra, S., Panda, B. S., Mahala, S. S., Mohanty, S. K., ... & Murugesan, K. (2023). A comprehensive study of the estuary sea environment in the Bay of Bengal, near the Mahanadi River confluence. Discover Water, 3(1), 20.
44. Pattanayak, A. A., Swain, S., Behera, R. R., Sharma, S. D., Panda, C. R., & Mohanty, P. K. (2024). Variability in water quality of two meso-tidal estuaries of Odisha, East Coast of India. Journal of Marine Systems, 241, 103919.
45. Hussain, S.A., Han, F.Q., Ma, Z., Hussain, A., Mughal, M.S., Han, J., Alhassan, A., Widory, D., 2021. Origin and evolution of Eocene rock salts in Pakistan and implications for paleoclimate studies: insights from chemistry and Cl stable isotopes. Front. Earth Sci. 9, 1–13. https://doi.org/10.3389/feart.2021.644485
46. WHO, 2011. Guidelines for drinking-water quality. WHO Chron. 38 (4), 104–108 from. https://www.epa.gov/sites/default/files/2014-03/documents/guidelines_for_drinki ng_water_quality_3v.pdf. (Accessed 1 June 2023).
47. Ali, S., Khorrami, B., Jehanzaib, M., Tariq, A., Ajmal, M., Arshad, A., Shafeeque, M., Dilawar, A., Basit, I., Zhang, L., Sadri, S., Niaz, M.A., Jamil, A., Khan, S.N., 2023. Spatial downscaling of GRACE data based on XGBoost model for improved understanding of hydrological droughts in the indus basin irrigation system (IBIS). Rem. Sens. 15, 873. https://doi.org/10.3390/rs15040873.
48. Yin, L., Wang, L., Li, J., Lu, S., Tian, J., Yin, Z., Liu, S., Zheng, W., 2023. YOLOV4_CSPBi: enhanced land target detection model. Land 12, 1813. https://doi.org/10.3390/ land12091813.
49. Omonona, O. V., & Okogbue, C. O. (2024). Cumulative ecological risk of groups of trace metals in soils impacted by agricultural activity. International Journal of Environmental Science and Technology, 21(1), 687-702.
50. Khaled, E. B., Obeidat, M., Al-Ajlouni, A., Awawdeh, M., & Dalo, M. A. (2024). Demarcation of Groundwater Quality Using Drinking Water Quality Index (DWQI), Nitrate Pollution Index (NPI), and Irrigation Indices: A Case Study from Jerash Region.
51. Esmaeili, M., Abbasi-Moghadam, D., Sharifi, A., Tariq, A., Li, Q., 2024. ResMorCNN model: hyperspectral images classification using residual-injection morphological features and 3DCNN layers. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 17, 219–243. https://doi.org/10.1109/JSTARS.2023.3328389
52. Lu, S., Zhu, G., Meng, G., Lin, X., Liu, Y., Qiu, D., Xu, Y., Wang, Q., Chen, L., Li, R., Jiao, Y., 2024. Influence of atmospheric circulation on the stable isotope of precipitation in the monsoon margin region. Atmos. Res. 298, 107131. https://doi. org/10.1016/j.atmosres.2023.107131.
53. Monira, U., Sattar, G. S., & Mostafa, M. G. (2024). Assessment of surface water quality using the Water Quality Index (WQI) and multivariate statistical analysis (MSA), around tannery industry effluent discharge areas. H2Open Journal, 7(2), 130-148.
54. Ali, S., Verma, S., Agarwal, M. B., Islam, R., Mehrotra, M., Deolia, R. K., ... & Fattahi, M. (2024). Groundwater quality assessment using water quality index and principal component analysis in the Achnera block, Agra district, Uttar Pradesh, Northern India. Scientific Reports, 14(1), 5381.

55. Bokhari, R., Shu, H., Tariq, A., Al-Ansari, N., Guluzade, R., Chen, T., Jamil, A., Aslam, M., 2023. Land subsidence analysis using synthetic aperture radar data. Heliyon 9, e14690. https://doi.org/10.1016/j.heliyon.2023.e14690
56. Cruz, F., & Silva, T. F. D. G. (2024). Water quality emergency monitoring networks: a method for identifying non-critical variables based on Shannon's entropy. Journal of Hydroinformatics, jh2024256.
57. Felegari, S., Sharifi, A., Khosravi, M., Sabanov, S., Tariq, A., Karuppannan, S., 2023. Using Sentinel-2 data to estimate the concentration of heavy metals caused by industrial activities in Ust-Kamenogorsk, Northeastern Kazakhstan. Heliyon 9, e21908. https://doi.org/10.1016/j.heliyon.2023.e21908.
58. Cheng, Y., Lan, S., Fan, X., Tjahjadi, T., Jin, S., Cao, L., 2023. A dual-branch weakly supervised learning-based network for accurate mapping of woody vegetation from remote sensing images. Int. J. Appl. Earth Obs. Geoinf. 124, 103499.
59. Zhao, Z., Islam, F., Waseem, L.A., Tariq, A., Nawaz, M., Islam, I.U., Bibi, T., Rehman, N. U., Ahmad, W., Aslam, R.W., Raza, D., Hatamleh, W.A., 2024. Comparison of three machine learning algorithms using google earth engine for land use land cover classification. Rangel. Ecol. Manag. 92, 129–137. https://doi.org/10.1016/j. rama.2023.10.007.
60. Han, X., Tang, F., & Liu, A. L. (2024). Drinking water quality evaluation in supply systems in Wuhan, China: application of entropy weight water quality index and multivariate statistical analysis. Environmental Science and Pollution Research, 31(1), 280-292.
61. Zheng, X., Sarwar, A., Islam, F., Majid, A., Tariq, A., Ali, M., Gulzar, S., Khan, M.I., Sardar Ali, M.A., Israr, M., Jamil, A., Aslam, M., Soufan, W., 2023. Rainwater harvesting for agriculture development using multi-influence factor and fuzzy overlay techniques. Environ. Res. 238, 117189 https://doi.org/10.1016/j.envres.2023.117189.
62. Gautam, V. K., Kothari, M., Al-Ramadan, B., Singh, P. K., Upadhyay, H., Pande, C. B., ... & Yaseen, Z. M. (2024). Groundwater quality characterization using an integrated water quality index and multivariate statistical techniques. Plos one, 19(2), e0294533.
63. Shah, S.H.I.A., Yan, J., Ullah, I., Aslam, B., Tariq, A., Zhang, L., Mumtaz, F., 2021. Classification of aquifer vulnerability by using the DRASTIC index and geo-electrical techniques. Water 13, 2144. https://doi.org/10.3390/w13162144
64. Jayaraman, P., Nagarajan, K. K., Partheeban, P., & Krishnamurthy, V. (2024). Critical review on water quality analysis using IoT and machine learning models. International Journal of Information Management Data Insights, 4(1), 100210.
65. Wu, X., Guo, S., Qian, S., Wang, Z., Lai, C., Li, J., Liu, P., 2022. Long-range precipitation forecast based on multipole and preceding fluctuations of sea surface temperature. Int. J. Climatol. 42, 8024–8039. https://doi.org/10.1002/joc.7690.
66. Islam, F., Tariq, A., Guluzade, R., Zhao, N., Shah, S.U., Ullah, M., Hussain, M.L., Ahmad, M.N., Alasmari, A., Alzuaibr, F.M., Askary, A.El, Aslam, M., 2023. Comparative analysis of GIS and RS based models for delineation of groundwater potential zone mapping. Geomatics, Nat. Hazards Risk 14, 27. https://doi.org/ 10.1080/19475705.2023.2216852.
67. Prajapat, K., Sanwal, S. K., & Sharma, P. C. (2024). Screening of quinoa (Chenopodium quinoa Willd.) germplasms under high-SAR saline water on the basis of growth, yield, and multivariate analysis. Journal of Biosciences, 49(1), 23.
68. Khan, D., Raziq, A., Young, H.W.V., Sardar, T., Liou, Y.A., 2022. Identifying potential sites for rainwater harvesting structures in Ghazi Tehsil, Khyber Pakhtunkhwa, Pakistan, using geospatial approach. Remote Sens. 14, 5008. https://doi.org /10.3390/rs14195008.
69. Tariq, A., Qin, S., 2023. Spatio-temporal variation in surface water in Punjab, Pakistan from 1985 to 2020 using machine-learning methods with time-series remote sensing data and driving factors. Agric. Water Manag. 280, 108228 https://doi.org/ 10.1016/j.agwat.2023.108228
70. Xie, Y., Nie, X., Wang, C., Xu, X., & Zhang, F. (2024). A metabolomics approach for monitoring thermal processing temperature of bovine milk using ultra-performance liquid chromatography tandem Q-Exactive mass spectrometry and multivariate data analysis. Journal of Future Foods, 4(1), 83-90.
71. Aziz, S.F., Abdulrahman, K.Z., Ali, S.S., Karakouzian, M., 2023. Water harvesting in the Garmian Region (Kurdistan, Iraq) using GIS and remote sensing. Water (Switzerland) 15. https://doi.org/10.3390/w15030507.
72. Laghrib, F., Bahaj, T., El Kasmi, S., Hilali, M., Kacimi, I., Nouayti, N., ... & Hammani, O. (2024). Hydrogeochemical study of groundwater in arid and semi-arid regions of the Infracenomanian aquifers

(Cretaceous Errachidia basin, Southeastern Morocco). Using hydrochemical modeling and multivariate statistical analysis. Journal of African Earth Sciences, 209, 105132.

73. Das, A. (2022). Multivariate statistical approach for the assessment of water quality of Mahanadi basin, Odisha. Materials Today: Proceedings, 65, A1-A11.
74. Das, A. (2024). Assessment of Surface Water Quality Using Entropy-WQI, Fuzzy-TOPSIS Analysis, Irrigation Indices and Spatial Interpolation Approaches in Mahanadi River Basin, Odisha, India. Journal of Environment Nanotechnology., Volume 13, No (1), pp. 182-212.
75. APHA (2005) Standard methods for the examination of water and wastewater, 21st edn. New York
76. Das, A. (2024). Water Quality Assessment Using Water Quality Index (WQI) Under GIS Framework in Brahmani Basin, Odisha. In International Conference Innovation in Smart and Sustainable Infrastructure (pp. 131-150). Springer, Singapore.
77. Das, A. (2023). Characterization of Surface Water Quality Using Water Evaluation Indices, EDAS and Geo-Statistics in Brahmani River Basin (BRB), Odisha, India. In Hydraulic and Civil Engineering Technology VIII (pp. 826-834). IOS Press.
78. Das, A. (2023). Anthropogenic Effects on Surface Water Quality Assessment in Baitarani River Basin, Odisha Using GIS and MCDM Techniques. Engineering Research Transcripts, 5, 37-64.