Assessment of Drinking Water Suitability Using Water Quality Index in Patna District, Bihar, India

[1]Mr.Pappu Kumar Yadav, [2]Dr.Pappu Kumar

[1] *Research Scholar, Department of Civil Engineering, Sandip University, Madhubani, Bihar, India*

[2] *Assistant Professor, Department of Civil Engineering, Sandip University, Madhubani, Bihar, India*

Abstract - Anthropogenic activities-related degradation of groundwater quality has emerged as a major concern. By creating a water quality index as a single value, several water quality variables are used to explain the overall state of the water quality. The following water quality factors make up the index: calcium, magnesium, chloride, sulphate, fluoride, nitrate, manganese, pH, EC, total alkalinity, and total hardness. Using a drinking water quality index approach, the current study attempts to evaluate the drinking water quality of the study area in and around urban Patna. Ten criteria related to water quality have been chosen for assessment. The quality of the water samples is assessed using an arithmetic and geometric index using a data set of twenty ground water samples that were taken from the study area in and around urban Patna, Patna. framework. The groundwater quality data collected from 67 sites spread throughout the city were subjected to cluster analysis (CA), principal component analysis (PCA), and discriminant analysis (DA). The data was measured on 10 different parameters. The 67 sampling stations were divided into two groups using hierarchical cluster analysis (CA), with cluster 1 exhibiting high pollution and cluster 2 exhibiting lower pollution. To identify the most significant characteristics that account for the temporal and spatial fluctuations in the groundwater quality of the research area, discriminant analysis (DA) was used. The most significant parameter found by Temporal DA is pH, which accounts for the seasonal assignment of cases and distinguishes between water quality during the pre- and post-monsoon seasons. Mg, Cl, and NO3 were found by spatial DA to be the three most significant factors that differentiated between two clusters and accounted for 89% of the differences.

Keywords: principal component analysis, cluster analysis (CA), water Quality Index, Groundwater

INTRODUCTION

Groundwater used for domestic, irrigation and industrial fulfilment of water in the world. Now a days ground water resources decreasing and water demand increasing due to use of different fields(Nampak et al. 2014). In the last few decades, there has been a tremendous increase in the demand for fresh water due to rapid growth of population and the accelerated pace of industrialization. Due to ground water is contaminated as such groundwater should be protected from contamination by analysing , planning and management Human health is threatened by most of the agricultural development activities particularly in relation to excessive application of fertilizers and unsanitary conditions(Ramakrishnaiah et al. 2009). In a drinking water quality assessment, the decision making based on water quality data is a crucial issue. Traditionally, water resource professional communicates their decision on drinking water quality status by comparing the individual parameters with guideline values. While this decision is too technical and detailed, without providing a whole picture of drinking water quality(Cude 2001). For determination of water status WQI is a mathematical tool that determine the value of sample water after determination of WQI we can say that the water is good or poor or excellent(Singh et al. 2008).for the water quality index deployment many parameter have been taken. These index are aim to overall condition of water in different environmental condition(Neshat et al. 2015). The main focus of this study is to analysis the water quality and map all WQI. The groundwater data consist of 9 variable and sample taken at 32 locations in Patna city. Therefore, applying advanced data analysis might help to understand quality of water and to improve produced water quality. In this paper we study the consequences of using a well-known method of cluster analysis to partition they sample treatment means in a balanced design and show how a corresponding likelihood ratio test gives a method of judging the significance of the differences among groups obtained.

STUDY AREA

The Patna district is located near the Ganga River's southernmost point. The city is 16 km wide and 35 km long. The Patna district is 51 metres above mean sea level and is located between 25º13′ and 25º45′ North latitude and 84º 43′ and 86º 44′ East longitude. It is the most populous city in eastern India and serves as the capital of Bihar. According to Singh and Singh (2017), Patna's climate is humid subtropical, with scorching summers from late March to early June, monsoon rains from late June to September, and a pleasant winter from November to February.

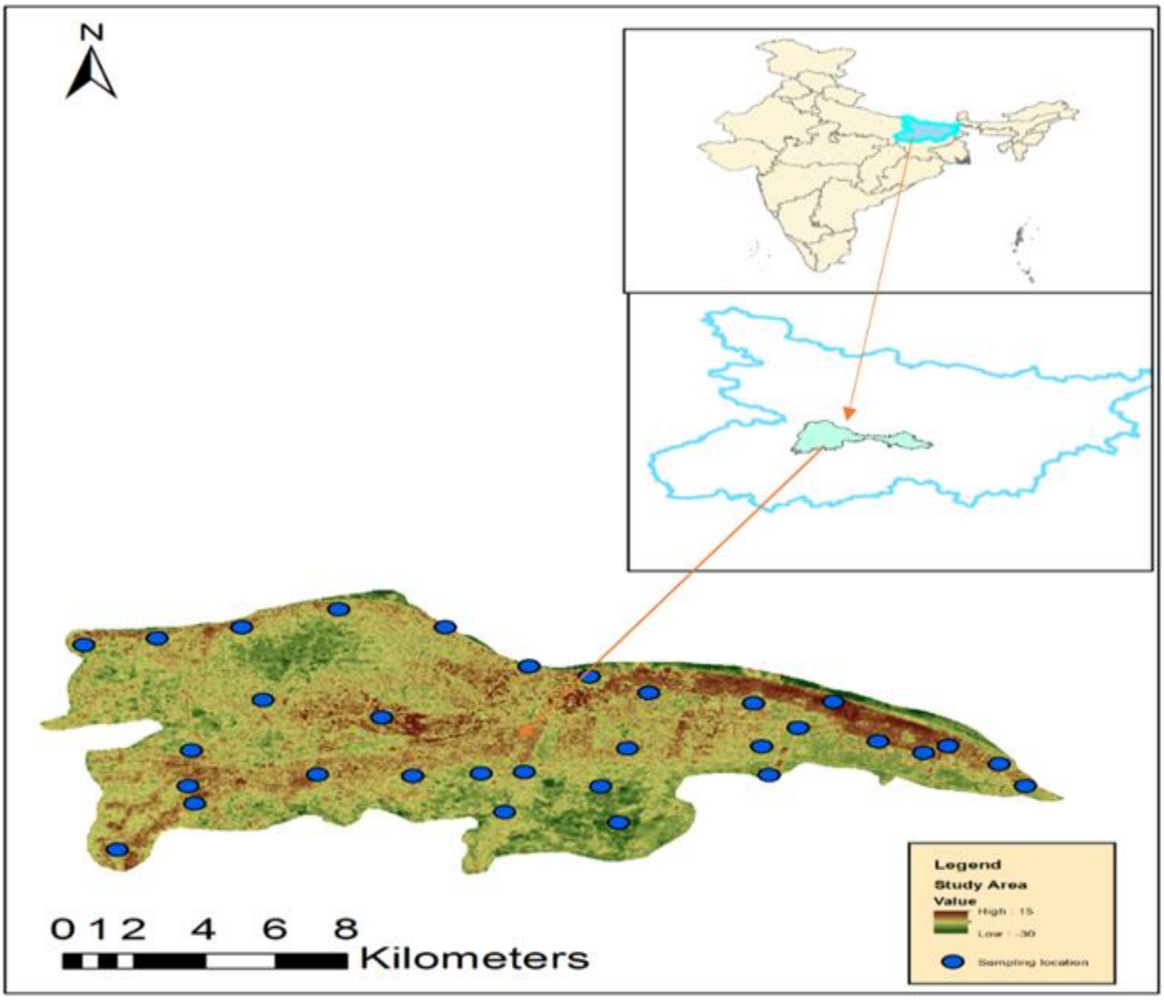


Fig.: Study Area

Regional hydrogeology

depicts the research area, which spans 87 km2 and is located between 25° 32' 8.12" and 25° 39' 18.14"N latitude and 85° 0' 16.83" to 85° 16' 2.37"E longitude. Three blocks in the urban Patna districts of "Patna Sadar," "Phulwarisarif," and "Danapur" make up the research area. The district's terrain is made up of plains with a small westward slope. The western portion of the district generally slopes directly north and northeast, with a land surface elevation that varies from 68 metres in the south to 48 metres in the north and from 67 metres in the west to 45 metres in the east. The district experiences 1076 mm of precipitation on average and 26°C on average annually.

Water Quality Index Calculation

The 17 variables (EC, TDS, turbidity, pH, total hardness, Ca, Mg, sulphate, phosphate, nitrate, nitrite, fluoride, chloride, Fe, Mn, Cu, and Cr(VI)) that make up the groundwater quality data used in this study were periodically sampled from 65 wells for eight years between 2006 and 2013 by the Andimeshk Health Network and the Ministry of Energy in Iran.

To calculate WQI, we used the annual mean of each parameter. Each parameter was then given a weight (wi) based on how important they were in terms of the overall quality of the water that could be consumed. The parameters that had the most negative impact on groundwater users' health were given the highest weights, while the parameters that had the least negative impact in this area were given the lowest weights. Step two:

$$W_i = w_i / \sum_{i=1}^{n} w_i$$

where Wi is the relative weight, wi is the weight of each parameter, and n is the number of parameters. In the third step, a quality rating scale (qi) for each parameter was assigned by dividing its concentration in each water sample by its respective standard and the result multiplied by 100:

$$q_i = \frac{C_i}{S_i} * 100$$

where Ci is the concentration of each chemical parameter in each water sample, and Si is the associated drinking water standard. The final WQI was determined by the product of Wi and qi as the following:

$$WQI = W_i * q_i$$

WQI AND STATUS

S.no.	WQI value	Water quality
1	<50	excellent
2	50-100	good water
3	100-200	poor water
4	200-300	very poor water

Water quality classification based on WQI value

Multivariate statistical methods

Sparse Principal Component Analysis

It is a classical method for the reduction of dimensionality of data in the form of n observations (or cases) of a vector with p variables. Existing data sets often have p comparable to, or even much larger than, n. Principal component analysis (PCA) is a statistical technique that is deeply used as pattern deciphering in data sets on their resemblance by reducing the dimensions and complexity in the input data matrix. In the form of n observations (or cases) of a vector with p variables, it is a traditional method for reducing the dimensionality of data. P is generally larger than n or at least comparable in existing data sets. It is essentially the extraction of the independent variables' eigenvalues and eigenvectors from the correlation matrix. Each original variable's importance to a certain component is represented by principal component (PC) loading, which is calculated by multiplying the eigenvectors by the square root of the eigenvalue. Accordingly, loading is frequently regarded as a crucial element for the indiscernibility link between PC and a particular variable (Olsen et al. 2012). The more dispersion in the data is described by the PC with eigenvalues of 1.0 and above, which are regarded as significant (Wang et al. 2013).

Exploratory Factor Analysis (EFA)

In order to simplify the connected patterns among the set of variables, the broad purpose factor analysis is used to find complicated patterns by examining the data set and testing predictions based on shared variance (Child 1990). One of the statistical analyses used in research the most frequently is exploratory factor analysis (EFA). Reducing dimensionality, or the idea that measurable variables can be reduced to fewer latent variables that are unobservable and have a common variance, is the basis for factor analysis (Bartholomew et al. 2011). It is used when a researcher has to identify the elements that are impacting the variables and choose which variables "go together" (Decoster 1998). If there are m common "latent" elements in the dataset that need to be found, the EFA hypothesis states that the objective is to is to determine the fewest common elements necessary to explain the relationships (McDonald 2014). This programme factor was created as an all-purpose, user-friendly EFA calculator.

Cluster Analysis (CA)

Hierarchical cluster analysis is an agglomerative technique based on Ward method. This ward method treats each inputs as cluster based on Euclidean distance of Ward criteria, i.e. to reduce the error sum of squares (ESS) as shown in the equation (1). Suppose N_k clusters were developed at each step of k for given cluster of Z_l with n_l observation $(l = 1, \ldots, N_k)$,where Z_l is sum of the squared Euclidean distance for each element in Z_l from mean value.

$$ESS_{Zl} = \sum_{i=1}^{nl} \left(O_i - \overline{O}\right)\left(O_i - \overline{O}\right)' \quad (1)$$

Where $O_i (i = 1, \ldots, n_l)$ is the i-observation of the cluster Z_l and O is the corresponding mean value of all input observations in Z_l and $\left(O_i - \overline{O}\right)'$ is the transposed vector. Hence the total ESS for over-all cluster for this step is calculated based on the equation (2).

$$ESS = \sum_{i=1}^{N_K} EES_{Z_L} \quad (2)$$

The method starts by treating each observation as a cluster. Each successive step, clusters are merged according to the Ward criterion, i.e., the minimization of the error sum of squares (ESS) defined bellow. Suppose that at step k, there are Nk clusters. Given cluster Al with nl observations (l = 1, ... ,Nk), the ESSl for Al is defined as the sum of the squared Euclidean distance of each element in Al from the mean value:

RESULTS AND DISCUSSION

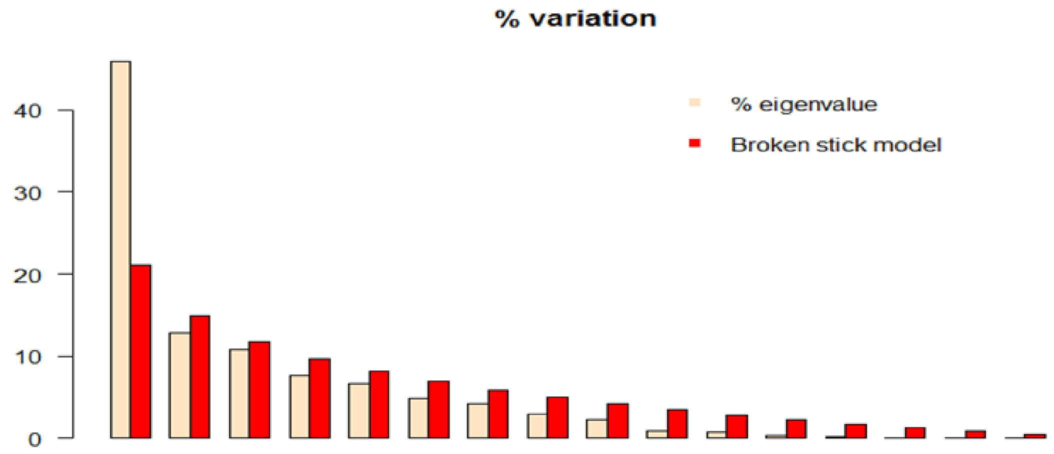
Evaluation of water quality

When statistical approaches are applied to evaluate the water quality in a given region using physiochemical data, the quality of the data, noise reduction techniques, and comprehension of the data before statistical methods are used all influence the outcome (Olsen et al. 2012). Numerous studies in the past failed to adequately address these problems when using multivariate statistical techniques to assess the quality of the water. (2012) Olsen et al. Thus, the authors agree with Olsen et al. (2012)'s findings and understand the data before applying any complex statistical procedures. Thus, we used our experience to examine the Patna groundwater quality data. Multivariate statistical analysis is now used to assess the data. The quantity of uranium and other chemical components in groundwater samples is summarised

Groundwater had an average aqueous uranium concentration of 2.3 ppb, ranging from 0.1 to 14.5 ppb. Comparatively speaking, the majority of the water samples appear to have uranium values that are lower than those advised by internationally recognised organisations including the WHO (2011), USEPA (2000b), and Canadian standards (2001). According to earlier research, there isn't expected to be any substantial harm at this level. The related physical parameters, on the other hand, fall into the following ranges: pH (6.7–7.8), EC (338–1844), DO (0.85–3.9 mg/l), Alkalinity (8.64–89.76 mg/l as $CaCO_3$), and temperature (29.7 to 34.5 °C). which show that, with the exception of a small amount of conductivity, chloride, and hardness, the most of the sample average concentration fall within the range of allowable levels.

Sparse Principal component analysis

The authors examine the use of correlation coefficients in principal component analysis in this section. These coefficients can be used to characterise the relationship between U and other factors related to the groundwater system in the research area. Therefore, to determine if PCA is required for the current variables or which variables should have been removed, the "naive" strategy of performing PCA first check has been utilised. Additionally, the chosen variable advances to the second step in order to determine PCs for the provided dataset. Where there is sufficient information available beforehand and where the goal is not scientific, this procedure could have been skipped. Numerous researchers have used a variety of techniques in the past to choose principal components (PCs). The "percentage explained criteria" is primarily utilised because it may account for a "large fraction," ideally greater than 60 to 70 percent, of the total variance. As was previously said, other.



An alternative strategy uses the "broken stick" paradigm. It is assumed that in the event that a dataset lacks correlation structure, the eigenvalues that result should be distributed in accordance with the p randomly broken pieces from a unit-length stick:

According to the above model, PCs with corresponding eigenvalues greater than are thought to be more accurate than statistically developed approaches or the thumb rule method Frontier (1976). Only one PC is nontrivial when comparing the eigenvalues for this water quality data set to the broken stick random model. The relationship between loading, the geochemical process, and the source of pollution is necessary for the interpretation of the PCs of the water quality.

REFERENCE

[1] Cude CG (2001): Oregon water quality index a tool for evaluating water quality management effectiveness. JAWRA Journal of the American Water Resources Association 37, 125-137

[2] Nampak H, Pradhan B, Manap MA (2014): Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. Journal of Hydrology 513, 283-300

[3] Neshat A, Pradhan B, Javadi S (2015): Risk assessment of groundwater pollution using Monte Carlo approach in an agricultural region: an example from Kerman Plain, Iran. Computers, Environment and Urban Systems 50, 66-73

[4] Ramakrishnaiah C, Sadashivaiah C, Ranganna G (2009): Assessment of water quality index for the groundwater in Tumkur Taluk, Karnataka State, India. Journal of Chemistry 6, 523-530

[5] Singh AK, Mondal G, Kumar S, Singh T, Tewary B, Sinha A (2008): Major ion chemistry, weathering processes and water quality assessment in upper catchment of Damodar River basin, India. Environmental geology 54, 745-758

[6] Singh V, Singh A (2017): Variation of temperature and rainfall at Patna. MAUSAM 68, 161-168

[7] Akhtar, Naseem, et al. "Modification of the water quality index (WQI) process for simple calculation using the multi-criteria decision-making (MCDM) method: a review." Water 13.7 (2021): 905.

[8] Ram, Arjun, et al. "Groundwater quality assessment using water quality index (WQI) under GIS framework." Applied Water Science 11 (2021): 1-20.

[9] Uddin, Md Galal, et al. "A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment." Water Research 219 (2022): 118532.

[10] Patel, Praharsh S., Dishant M. Pandya, and Manan Shah. "A systematic and comparative study of Water Quality Index (WQI) for groundwater quality analysis and assessment." Environmental Science and Pollution Research 30.19 (2023): 54303-54323.