**Assessing water quality variability using principal component analysis:**

**A study of wells in Kerala, India**

**ABSTRACT**

| **Aims:** To identify the major factors influencing groundwater quality across different wells in Kerala using Principal Component Analysis (PCA), and to evaluate spatial and seasonal variations in water quality parameters including conductivity, nitrogen, pH, total coliforms, and total dissolved solids. The study aims to provide actionable insights for improved water resource management.  **Study design:** Cross-sectional, observational, analytical study based on statistical analysis using PCA.  **Place and Duration of Study:** Groundwater samples were collected from multiple wells across districts in Kerala, India. Data were obtained from the Central Pollution Control Board (CPCB) under the National Water Quality Monitoring Programme (NWMP), 2022.  **Methodology:** Sixteen parameters, including temperature, pH, conductivity, total coliforms, nitrogen, biochemical oxygen demand (BOD), total dissolved solids (TDS), and fluoride, were used to obtain and standardize data on groundwater quality. In order to reduce dimensionality and identify important principle components (PCs), PCA was carried out using R software. A detailed analysis of the first four components (PC1–PC4), which accounted for 68.4% of the total variance, was conducted. To evaluate regional variations and dominating trends in water quality, each well was given a score based on the PCs.  **Results:**PC1 (33.33%) showed increased conductivity and total coliform levels along with an overall pattern of declining water quality. PC2 (21.06%) emphasized pH and nitrogen-related variance, with Punalur and Malappuram wells exhibiting higher nutrient levels. PC4 (11.06%) represented localized anomalies, while PC3 (14%) concentrated on pH fluctuation. For instance, the Kannur Municipality well revealed high levels of contamination, whereas the Vytilla (Ernakulam) well had low water quality. These trends indicate the effects of industrial runoff, sewage, and agriculture.  **Conclusion:** Principal Component Analysis successfully identified the main variables and geographical trends affecting Kerala's groundwater quality. In order to ensure sustainable and secure groundwater resources, the study emphasizes the necessity of focused water quality management through enhanced sanitation, pollution prevention, and localized monitoring techniques. |
| --- |

*Keywords: Principal Component Analysis (PCA); water quality; groundwater; contamination; nutrient levels; pH variations; coliforms; total dissolved solids; wastewater management; water monitoring*

**1. INTRODUCTION**

Water is a precious resource which makes up about two-thirds of the Earth's surface. Even though only a minute percentage of it is appropriate for human use. The world's drinking water requirements are met by approximately one-third of surface water sources, but the activity of humans is posing an increasing threat to this restricted water supply. The most critical environmental problems of our time is water pollution, which can be caused by a combination of sources such as trash from homes and construction sites, industrial waste, sewage leaks, and agricultural runoff. Water quality is significantly impacted by pollutants, which commonly break down into contaminants including carbon dioxide, methane, and organic compounds. Some heavily populated areas of India are affected by Pollution. Water sources like wells are contaminated by Seepage from nearby pits, septic tanks, and latrines. Kerala is still facing issues with water quality predominantly throughout the rainy season when too much water dilutes some pollutants despite having a lot of rainfall each year and having an abundance of water resources. Mishra. 2010 suggested that Principal Component Analysis (PCA) is a useful statistical approach for determining major elements influencing water quality in the Ganges. Nutrient load, sewage and fecal contamination, physicochemical variability, and wastewater pollution account for more than 90% of the reported variance in water quality.

This exhibits how crucial it is to examine water quality since it is helpful in detecting and treating water pollution, guaranteeing safe drinking water and protecting public health. Hence, this study intends to analyze the spatial and seasonal variability in water quality of wells in Kerala using PCA.

Principal Component Analysis (PCA) is a statistical method extensively used in water quality evaluation to handle complex datasets with multiple interconnected parameters. It transforms the original variables, such as pH, dissolved oxygen, and heavy metal concentrations, into a smaller set of uncorrelated principal components, each one capturing a vital portion of the data's variance. This dimensionality decline highlights key patterns and trends, making it easier to recognize critical water quality parameters and sources of pollution, such as agricultural runoff, industrial effluents, or natural processes. PCA is particularly valuable for identifying temporal and spatial trends, grouping similar water bodies, and reducing the need for monitoring all parameters. Regardless of its merits, PCA has several demerits, including its assumption of linear relationships and sensitivity to outliers, which can distort results. Tripathy & Singal. 2019 used Principal Component Analysis to select and weigh critical criteria for creating a Water Quality Index (WQI) for the Ganga River, reducing 28 elements to 9. It promotes global water security goals and offers a framework for developing region-specific WQIs based on historical data. The interpretation of chief components can also be subjective and need expertise. However, when used cautiously, PCA offers valuable insights into the underlying factors affecting water quality, supports efficient resource management, and complements other statistical tools for environmental monitoring. Evaporation and water-rock interactions affect the main water chemistry, which calls for immediate treatment for long-term use. CaHCO₃ is the predominant water type, and silicate weathering and carbonate dissolution are the key factors affecting the water chemistry. Rock-water interaction and silicate weathering are the two main hydrochemical processes, highlighting the necessity of continuous quality monitoring. Dera Bassi's groundwater is largely adequate for drinking and irrigation, while some places exceed the alkalinity and magnesium limitations. According to hydrochemical study, rock weathering and ion exchange processes are major factors influencing water quality (Sharma et al., 2021,Jain et al.,2024, Ndah Anyang et al.,2021). The present study intends to identify and analyze the main factors influencing the quality of groundwater in wells across Kerala with the use of Principal Component Analysis (PCA). It will be used to evaluate the spatial variations in groundwater quality across different areas of the state, considering the seasonal variations. Moreover, the study seeks to offer actionable recommendations for enhancing groundwater quality monitoring and management, based on the analysis of these factors and variations.

**2. methodology**

Water quality is influenced by a variety of factors, one of which is temperature. Temperature plays an important role in determining oxygen levels, the rates of chemical reactions and biological activities of organisms. Warmer water expands the capacity of depleting oxygen and generating harmful algal blooms, while its cooler counterpart sustains ecosystem stability but slows down the turnover of pollutants and nutrients. Formation of pH measures the degree of acidity or alkalinity in water and acts as a deciding factor for the health of aquatic organisms, interference with chemical processes, and solubility of toxic substances reaching extreme levels. Conductivity, which is a direct measurement of dissolved salts and minerals present in the water, would generally be on the high side in the presence of pollutants or nutrients and may therefore threaten aquatic ecosystems. Pollution indicators such as Biochemical Oxygen Demand (BOD), which denotes oxygen absorbed by microorganisms in the degradation of organic matter, are the trademark of a pollution incident proportionally represented by decreasing levels of dissolved oxygen. Nitrate-N and Nitrite-N, substances often identified in agricultural runoff or wastewater, can lead to the rapid exhaustion of available oxygen and eutrophication in water bodies and act as a health hazard, mostly with respect to drinking water. The occurrence of fecal coliform and total coliform bacteria, indicators of fecal contamination, indicates the possible occurrence of harmful pathogens.

High concentrations of total dissolved solids (TDS), which quantify dissolved load, influence the taste of water, create an oxygenic conditions, and kill aquatic life; indicate contamination by high or low levels of flavor up or down, respectively. Lastly, fluorine is important for tooth health at very low levels; at high levels, however, it is a precursor of dental and skeletal fluorosis, is not good for the human body, and pollutes water.

The quality of water data used for this study are from the Water Quality Assessment by the Central Pollution Control Board (CPCB) under the National Water Quality monitoring Programme (NWMP), 2022. To prevent the problem of equitability in associated contribution by various parameters, therefore applications of standardization methods are always adopted as variables may otherwise assume the role of contribution in their own respective units or scales. Hence forth, PCA begins with either the covariance matrix or correlation matrix being computed to recognize the relationships and variances with respect to the variables by pointing those that may be under some influences from common underlying factors. From this matrix, eigen values and eigenvectors are calculated in which eigen values represent the variance explained by the principal component, while eigenvectors characterize the direction of these components in data space. Principal components will then be ranked based on the amount of variance they explain, with mostly the first few principal components being the only ones to assist interpretation. Finally, the components are analyzed to observe any trends and correlations, and the different principal components could represent different aspects of water quality from the splitting due to specific parameters into pollution or possibly seasonal effects.

Statistical methods and software used are significant to the analysis and interpretation of applicable water quality data. In order to assess the relation between water quality parameters and multicollinearity, it is necessary to perform Principal Component Analysis (PCA) in advance. PCA is a widely accepted statistical method that cuts dimensionality while keeping variance in the dataset as high as possible; the extraction of eigenvalues and eigenvectors for the computation of principal components suggests the procedure's proof of identity of dominant aspects that control water quality. Exploratory Data Analysis (EDA) methods have been used, examples being scatter plots, histograms, and box plots, which are also visualized for trends, outliers, and cleaning of the dataset. This study employs multiple software applications in carrying out these methods. Favorite R with its packages like stats, FactoMineR, Factoextra, PCA tools, ade4, cluster, vegan, ggfortify, bigstatsr, tensorflow, RSpectra, glmnet, caret; etc. is for PCA, data visualization, and correlation analysis. These statistical methods and software thereby make an all round analysis of water quality variability.

**3. results and discussion**

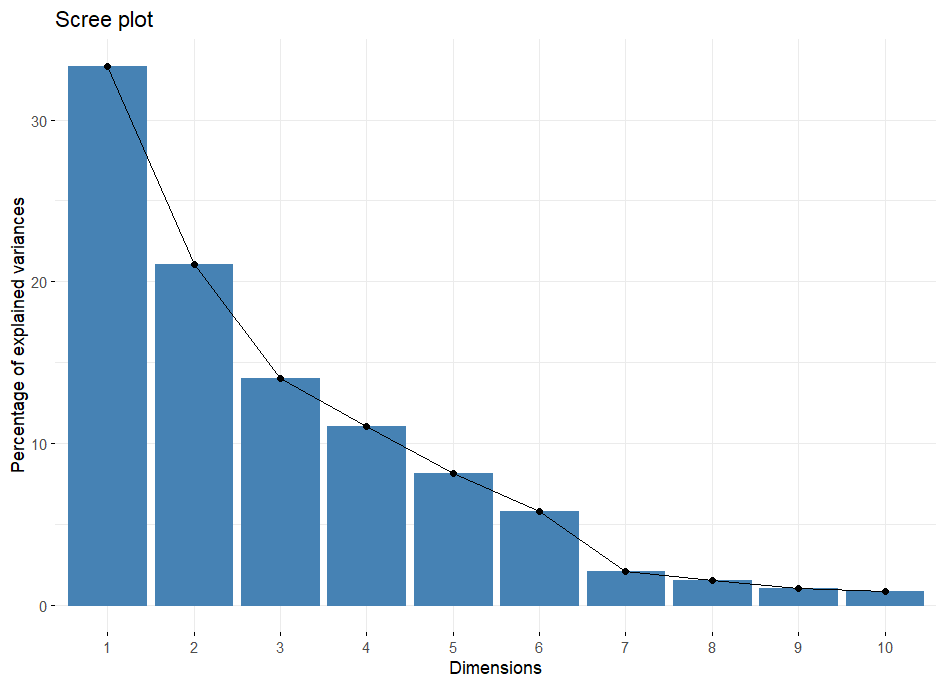
PCA study was performed on various wells in Kerala, India, for the understanding of key influencing factors on water quality or related variables. The data for these wells comprise measurements along sixteen principal components which were calculated as a part of the PCA. The respective PCA component scores of each well on the sixteen principal components are given in Table 1.

PCA results depicted in figure 1 present the ability of each component to explain variance in the dataset. Principal Component 1 (PC1) is the most significant due to the highest standard deviation of 2.3094, responsible for 33.33% of the total variance. This means that PC1 is important to explain the most substantial variation in the water quality parameters. This is consistent with research by Kumar et al. (2021), who found that inadequate waste management and poor sanitation were the primary causes of microbial contamination in shallow wells throughout Kerala. In the same manner, Principal Component 2 (PC2) also represents a significant water quality pattern, explaining 21.06% of variance and having 1.8358 of standard deviation, which reflects its ability to produce other significant patterns. Based on the cumulative variance

**Table 1. PCA Components (1-4) Scores for Wells in Kerala.**

| **Well Name** | **PC1** | **PC2** | **PC3** | **PC4** |
| --- | --- | --- | --- | --- |
| Well At Pappanamkode,  Thiruvananthapuram, Kerala | 1.455172545 | -0.730320385 | 0.87295803 | -0.09371065 |
| Well At Kannur Municipality Kannur | 2.266491017 | -1.481964203 | 0.76226511 | 0.05390029 |
| Well At Hazardous Waste Dump  (Ambalamugal Borewell) | 1.207383781 | -1.977568937 | 0.0783878 | -0.81257667 |
| Well At Vaikom | 0.883792492 | 0.326442813 | 0.0269742 | -1.6668404 |
| Well At Karoor (Pala) | 1.639991933 | -0.186409984 | -0.0253023 | -2.79987344 |
| WellatEloor, Ernakulam | 1.482661378 | -0.861161181 | 0.71655556 | -0.84163492 |
| Well At Kalamassery Ernakulam  Dist. ,Kerala | 0.536322836 | 0.955611112 | 0.26702163 | -1.75008788 |
| Well At Karimbam, Kannur | 2.253447225 | -1.429380276 | 0.42083569 | 0.55161268 |
| Well At  Nedumangad, Thiruvananthapuram,  Kerala | -0.055897863 | -0.980735993 | 1.39053503 | 1.65469001 |
| Well At Chellora Trenching  Ground (Kannur) | 1.409539229 | -0.902567255 | 0.11653195 | -0.64788342 |
| Well At Vadavathoor (Kottayam) | 1.586665532 | -1.262080702 | 0.44715133 | -0.89137335 |
| Well At Edayar Ernakulam Distt.,  Kerala | -2.274907429 | -1.160427155 | -0.29577228 | -1.99228512 |
| Well At Brahmapuram M.S.W.  Dumpark | 0.395333012 | -0.773564392 | 0.35216865 | -0.84662829 |
| WellatChungapally, Kerala | 1.713134457 | -0.304750266 | 0.67411153 | 0.01779789 |
| Well At Payyannur, Kannur  Distt., Kerala | 0.295749955 | 0.004190987 | -0.39061758 | -0.26622599 |
| Well At  Fathimapuram (Changanassery) | 0.785351302 | -0.62890959 | 1.38751131 | 1.05132109 |
| Well At Punnalpettippalam | 0.215155505 | 0.933084735 | 0.25027048 | 0.31681872 |
| Well At Vellparamba, Kozhikode | 1.623001414 | -0.730201664 | -0.29823841 | 0.37572177 |
| Well At Mavoor, Kozhikode  Distt., Kerala | 1.436979778 | -0.815960726 | -0.50685181 | 0.55665016 |
| Well At Kundara, Kollam Distt.,  Kerala | 0.20908251 | 2.231226912 | -2.60246877 | 0.2287938 |
| Well At Kureepuzha (Kolam) | 0.497629435 | 3.695804884 | -5.87905186 | 1.28215455 |
| Well At Manjeri | 0.345514378 | 2.445195191 | 0.92303015 | -0.879542 |
| Well Of Temple  Parassinikadavu, Kannur | -0.970001073 | 0.301994604 | 0.59840458 | 0.33227646 |
| WellatOllur (Thrissur) | 0.475099837 | -0.546828963 | -1.05421154 | -0.21146535 |
| Well At Cherthala, Alleppey, Kerala | -0.654403308 | -0.084233021 | 0.22329217 | 1.27793985 |
| Well At Punkunnam Thrissur  Distt, Kerala | -0.154753439 | 0.698139991 | -1.5279183 | 1.53263199 |
| Well At Sarvoday Puram, Alappuzha | -1.714186305 | -2.433537322 | 0.05185071 | 2.17828847 |
| Well At Vytila, Ernakulam Distt.  Kerala | -9.333946036 | -4.447979443 | -2.4545945 | -1.27765865 |
| Well At Malappuram, Kerala | -0.607572231 | 3.173856952 | 1.12294512 | -0.69653943 |
| Well At Payyanur, Kannur | 0.902728647 | -1.230623797 | 0.84217712 | 3.74225056 |
| Well At Sri Kirathamoorthy Siva  Temple, Kanjikkode, Palakkad | -0.379736738 | -1.374082498 | -0.09981865 | -0.13409536 |
| WellatLaloor (Thrissur) | -0.000365178 | 0.149869437 | -0.69033441 | 1.52582499 |
| Well At Punalur, Kerala | 0.423380754 | 2.198143049 | -0.12030476 | -1.96698237 |
| Wellat K.M.M.L. (Kollam) | -0.897095664 | 3.083312776 | 0.71743266 | 0.25378797 |
| Well At Karukamani | -6.996743685 | 4.146414307 | 3.70307433 | 0.84294206 |

proportion, it can be said that PC1 and PC2 together account for 54.40% of the total variance, thereby stressing the combined design usefulness on the dataset. These extra factors, such as PC3 and PC4, explain relatively little of the total variance but tell meaningful spread to the total variance. For instance, PC3 has a standard deviation of 1.497 and accounts for 14.00% of variance and PC4 contribute a similar 11.06% at a standard deviation of 1.3306, adding up to 68.40%. Thus, it indicates that the first four components cover around 68.40% of the total variability of the dataset, stressing their common importance.



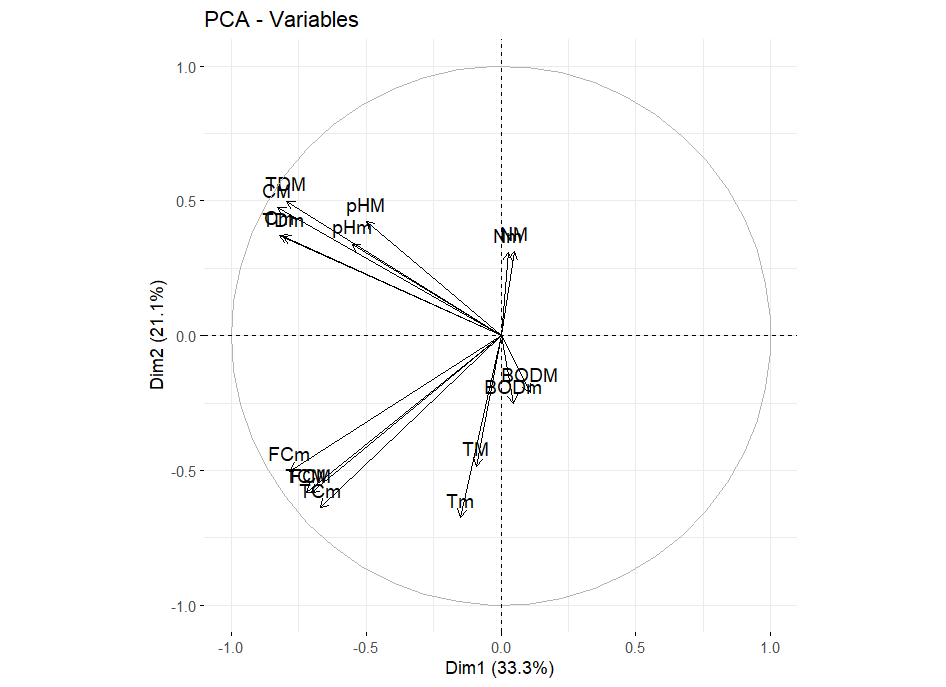
**Fig. 1. PCA Component Selection (Scree Plot)**

After PC4, the remaining components see their contribution to the variance continuously decreasing. From PC5 to PC12, it sounds progressive, with PC5 capturing a portion of variance accounting for 8.15%, PC6 5.78%, and all others, less than 2%. The cumulative ratio increases to 93.39% by PC6, and virtually all the variance can be explained by PC16 at a final cumulative ratio of 100%. This trend indicates that while the first few of such components are important in understanding the main components of variation in the data, the rest account for smaller contributions toward variance that are often not so easily interpreted.

Also, the same inference can be drawn from the screen plot presented in figure 2. So, the first few principal components are the ones where focusing on them gives much more meaningful insight into all the practical business analysis and decision-making.

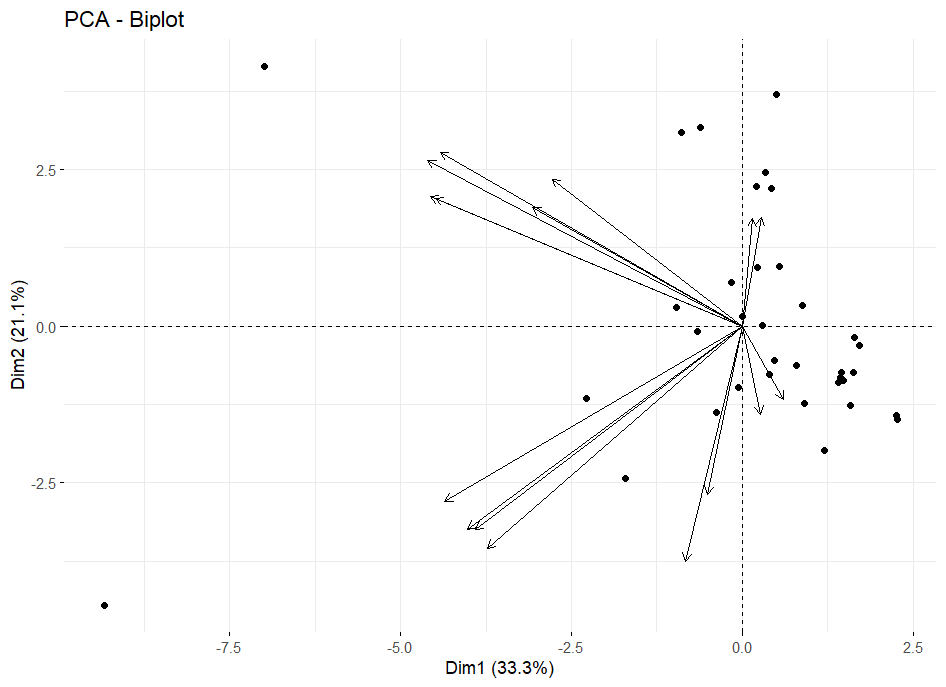
**3.1 Variability In Water Quality Across Different Wells In Kerala**

The PCA results for water quality parameters across different wells in Kerala have given varied information about the differences in the water quality at different locations. Each principal component, from PC1 to PC4, gives a different aspect of water quality, and the scores for each well provide insight into how those aspects differ regionally. Figure 3 shows the PCA Variable Plot, where all loadings of each variable are projected into the first two principal components. It expresses the importance of original variables in reduced dimensional space.



**Fig. 2. PCA Variable Loadings Plot**

Figure 3 presents the PCA Biplot that brings together both the observations and variable loadings into the first two principal components. It further allows for visualizing the relationship between the observations regarding each variable in the context of PCA.



**Fig. 3. Principal Component Biplot**

The PCA result has an understanding of the differences in water quality at different locations in Kerala in terms of water quality parameters across wells. Each principal component represents various aspects of water quality, while the score at each well indicates the regional variance of those aspects. Kazi *et al.*, 2009 used PCA to demonstrate that industrial discharge and agricultural practices extensively affected the quality of water.

PC1 or Principal Component 1 outlines a general pattern of degrading water quality. The positive locations of `WELL AT CHUNGAPALLY, KERALA` and `WELL AT KANNUR MUNICIPALITY KANNUR' indicate that these wells exhibit a higher number of parameters related to poor water quality such as increased conductivity, coliforms, etc. On the other hand, wells like `WELL AT VYTILA, ERNAKULAM DISTT. KERALA' and `WELL AT EDAYAR ERNAKULAM DISTT., KERALA' have negative PC1 scores and signify low to none of such parameters, implying comparatively good water quality. Thus, it's possible to categorize the wells on the basis of the extent of contamination/pollution with this component.

Principal Component 2 or PC2 refers to the diversities of nutrients and pollutants. Here, a location such as `WELL AT PUNALUR, KERALA' along with `WELL AT MALAPPURAM, KERALA' scores very positively on the metric, suggesting high pH, nitrogen, and total dissolved solids-thus perhaps implying better or more balanced water quality conditions and much higher nutrient values in these wells. Conversely, there are negative scores given by wells such as `WELL AT SARVODAY PURAM, ALAPPUZHA' and `WELL AT VADAVATHOOR (KOTTAYAM)', indicating lower pH and higher contamination-obviously suggesting poorer water quality and/or different water chemistry conditions.

Principal Component 3 (PC3) corresponds to the variations in pH levels. Wells such as `WELL AT NEDUMANGAD, THIRUVANANTHAPURAM, KERALA` and `WELL AT KUREEPUZHA (KOLAM)` with high positive scores of PC3 are corroborated with significant variations in high and low pH. This indicates extreme pH levels in these areas in comparison to other wells. Negatively scored areas such as `WELL AT VYTHILA, ERNAKULAM DIST. KERALA` can be considered to have a rather neutral pH. This component broadens understanding about variations in terms of pH, which could influence the chemical reactivity and biological processes occurring in the water.

Principal Component 4 (PC4) captures other more intricate or specific variations of water quality parameters. Positive scores in wells such as `WELL AT PAYANOOR, KANNUR` and `WELL AT KARIMBAM, KANNUR` seem to suggest that these wells tend to have distinct characteristics in their water quality profiles that cannot easily be explained by the other principal components. Negative scores, such as those observed for `WELL AT KAROOR (PALA)` and `WELL AT SRI KIRATHAMOORTHY SIVA TEMPLE, KANJIKKODE, PALAKKAD`, may indicate features with different tendencies or arguably much less pronounced effects on water quality. This component tends to capture more subtle variations and interactions between the parameters.

In the same way, each principal component gives a different view of the variables affecting the groundwater quality. While PC5 to PC16 are for temporal, spatial, or minor variations, PC1 to PC4 are for primary trends in contamination of general quality, salinity, microbiological aspects, industrial and agricultural influences.Highly contaminated wells were frequently found in areas with significant population densities or industrial activity, such Kannur and Vytilla, according to the spatial analysis of PC scores. Dhanachandra *et al.* (2021) observed similar patterns in Tamil Nadu and linked industrial zones to heavy metal contamination. Further supporting the dual nature of Kerala's pollution drivers, Das and Mandal (2020) discovered that both natural and anthropogenic sources had an impact on the iron levels in Assam groundwater. It may be possible to see such an approach as a hierarchy that encompasses all possible approaches to water resource development and contaminant control. Aref & Roosta. 2016 examined that Farashband plain's groundwater is generally unfit for drinking and irrigation due to its high levels of TDS and TH, as well as its extreme hardness, minor alkalinity, and brackishness.

**3.2 Practical Implications**

The PCA results offer a comprehensive view of the variations in water quality across different wells. A general trend of water quality deterioration is captured by Principal Component 1 (PC1). It displays strong negative loadings for maximum and minimum conductivity (`CM`, `Cm`), and total coliforms (`TCM`, `TCm`), suggesting that higher levels of these parameters are allied with lower PC1 scores. This designates that PC1 reflects an overall increase in pollution, where eminent conductivity and coliform counts, along with high total dissolved solids, are indicative of inferior water quality. The dominance of negative loadings for these pollution indicators emphasizes PC1 as a measure of water quality degradation. Sheela *et al*., 2012 used PCA, HCA, and FA to examine seasonal water quality fluctuations in Akkulam-Veli Lake, identifying important pollution sources such as organic contamination and seawater intrusion. The bund between the lakes has a considerable impact on water flow and quality, with organic pollutants rising before the monsoon and diminishing throughout the wet season.

Principal Component 2 (PC2) exposes a contrasting trend related to nutrient levels and contamination. This component demonstrate positive loadings for pH (`pHm`, `pHM`), nitrogen (`Nm`, `NM`), and total dissolved solids (`TDm`, `TDM`), signifying that higher values of these parameters are associated with higher PC2 scores. On the other hand, negative loadings for temperature (`TM`, `Tm`) and fecal coliforms (`FCM`, `FCm`) show that increased values of these parameters are connected to lower PC2 scores. PC2 comes into view to stand for a dimension where elevated pH and nitrogen levels, along with total dissolved solids, are generally related with better water quality, while higher temperatures and fecal contamination are associated with lower water quality. Principal Component 3 (PC3) mainly reflects variations in pH conditions. This component explains positive loadings for maximum and minimum pH values (`pHM`, `pHm`), which propose that PC3 is sensitive to changes in pH levels. The varied or weak negative loadings for other parameters, including temperature and conductivity, entail that PC3 might represent specific conditions affecting pH rather than a broad water quality trend. This is consistent with the findings of Nag and Ghosh (2013) in West Bengal, who documented arsenic pollution as a result of geogenic processes that affect groundwater chemistry and pH. Likewise Ahmad & Khurshid. 2019 stated that Ghaziabad's groundwater is largely good for irrigation, its high nitrate and TDS levels make it somewhat unsuitable for human consumption. The main cause is industrial and agricultural pollution, which needs to be addressed right away.The importances of pH-related loadings imply that this component is important for understanding variations in the acidity or alkalinity of the water. Kouser *et al*.2022, claimed that the majority of samples in the Kathua region fall under acceptable quality levels, making the groundwater there generally appropriate for industrial, agricultural, and drinking uses. Rock-water interaction mechanisms, primarily of the bicarbonate type, dominate the water chemistry. According to Lal et al. (2023), Ambagarh chowki's groundwater is generally safe for agriculture and drinking. However, certain samples had significant levels of salinity, hardness, and fluoride.

Principal Component 4 (PC4) inserts another layer of complexity to the understanding of water quality. It illustrates a varied pattern with positive and negative loadings across different parameters. While specific loadings for PC4 are less well-defined, the components likely confine additional aspects of water quality that are not fully explained by the other three components. The mixed loadings in PC4 point out that it may reflect more nuanced or specific conditions impacting water quality, such as concentrate changes in various parameters or interactions between them. By construe the PCA results, water quality managers and policymakers can better understand regional variations and tailor their strategies to deal with the specific needs of different locations. Sudhakaran *et al.,*2020 concluded that the river water in the Netravati basin exhibits seasonal fluctuations in quality, with well water continuously being of great quality while certain locations close to the coast have low drinking quality as a result of human inputs. Although specific pollution treatment is required for improvement, both WQI and IWQI show that the water is typically fit for drinking and irrigation. Sukanya & Sabu. 2020 declared that Karamana River has decent water quality upstream, but downstream areas show excessive BOD, NO₃⁻, and eutrophication, as proven by Eutrophication Index (EI), Comprehensive Pollution Index (CPI), and Organic Pollution Index (OPI). PCA and Pearson correlation show contamination from natural weathering, marine intrusion, and sewage inflow, stressing the necessity for pollution control in the pollution effect zone (ZPI). Sreedhar & Nagaraju. 2017 found that Tummalapalle groundwater has low sodium hazard (SAR < 10), tolerable residual sodium carbonate (RSC < 1.25), and a favorable permeability index, making it appropriate for irrigation. However, only 12.5% of the samples are appropriate for drinking since the majority of samples have high total dissolved solids (TDS) and total hardness (TH) that exceed the recommended limits. Singh *et al*., 2015 concluded that the majority of groundwater in the Chandauli-Varanasi region is safe for drinking and irrigation, while some samples are contaminated by agricultural runoff and home effluents, particularly during the monsoon. Effective water management can be achieved through embattled interventions based on these insights, improving overall water quality and ensuring safe and sustainable water resources.

Water quality is influenced in great part through analysis by some elements like conductivity, total coliform and fecal coliform, pH, nitrogen, total dissolved solids, temperature, and biochemical oxygen demand. Conductivity along with TDS, show increased levels of pollution. On the other hand, high levels of total and fecal coliform indicate contamination through sewage and animal waste. Differences in pH indicate changes in the degree of acidity or alkalinity which do biological and chemical processes. Nitrogen adds to eutrophication and pollution, particularly from agricultural runoff. Temperature of water has a direct impact on microbial activity, and indirectly on BOD which is a measure of organic pollution.

This review does a good job at analyzing the water quality data, yet it contains some gaps and potential biases. The geographic boundaries that focus on certain areas and the limited access to comprehensive water quality data coupled with short data collection periods presents a challenge to the accuracy and adaptability of the findings. Techniques like PCA that do not account for non-linear relationships have methodological limitations, and representativeness and generalizability are compromised by inadequate sampling and lack of control groups. The scaling of distress further limits the ability to generalize to larger or smaller water bodies. Examples of potential errors include instrumental mistakes, human error from sampling or data entry, and environmental factors such as weather or time that manipulate the measurements. Basic modeling reasoning along with a lack of consideration for spatial and temporal context contributes to data uniformity and oversimplifies complex patterns which are rigid and restrictive.

**3.3 Recommendations for Water Quality Management and Monitoring**

In order to remove contaminated sources and maintain natural balance, we need systematic activities to effectively manage and monitor water quality. Since conductivity, Coliform levels, and total dissolved solids (TDS) indicate the general deterioration of the water quality, the PCA results indicate an important area of ​​intervention. To solve this problem, housing and companies need not only more stringent rules for wastewater discharge, but also the rules necessary to encourage environmentally friendly agricultural methods such as precision farming. The buffer zones and vegetation near the reservoir can help you lower the temperature and increase the resistance to heat stability, but if they improve the hygiene infrastructure and increase the hygiene awareness, fecal contamination can be further reduced. Chaudhary & sateesh kumar. 2018 the high levels of EC, TDS, TH, and fluoride in northwest Rajasthan make the groundwater mostly unfit for human consumption. The two main factors influencing groundwater chemistry are anthropogenic activities and evaporation-crystallization. Olusegun  & Folashade.2021 declared that the research area's groundwater falls within WHO guidelines and is fit for irrigation and consumption.

To maintain a balanced water quality, pH fluctuations and controlling fertilizers are required. To prevent acidification, the monitoring program must focus on nitrogen, pH and TD while ensuring the requirements of industrial emissions. PH and nutrient levels can be adjusted using water land and natural buffer. Pollution sources in the field are an example of localized problems that require a special monitoring system and preliminary participation of stakeholders. Quality management is possible in general monitoring, which helps to detect new problems, including multidimensional technologies such as PCA. Ali *et al*., 2024 employed GWQI, PCA, and Piper diagrams to examine groundwater in the Achhnera block of Agra, revealing elevated EC, TDS, TH, fluoride, and chloride levels due to rock-water interaction. The majority of the samples were unfit for drinking due to the presence of sodium bicarbonate and calcium chloride. Mustapha & Abdu. 2012 used PCA and multiple regression to evaluate Jakara River water quality, identifying five major pollution sources and significant indicators such as DO, BOD₅, SS, and salinity. The findings stress the importance of better trash management and human activity monitoring in order to protect the river. Seasonal fluctuations affect groundwater quality, with slightly greater contamination levels seen following rainfall. Sharma et al., 2017 has claimed that groundwater in Bathinda shows seasonal contamination with fluoride, sulfate, nitrate, and high hardness, making some samples unsuitable for drinking and irrigation. Geochemical processes such as mineral dissolution, ion exchange, and leaching from fertilizers and livestock waste influence water quality. PCA detected geogenic fluoride pollution. Ravikumar & Somashekar. 2017 classified the hydrochemistry of groundwater using Piper and Chadha plots, identifying alkaline earth metals and reverse ion exchange processes as the main sources of permanent hardness. Key water quality characteristics that explained the majority of the variance were revealed by PCA, suggesting that groundwater is generally appropriate for irrigation and drinking. Venkatesan *et al*., 2022 observed that while groundwater in sections of Salem is chemically appropriate for irrigation, its high TDS, Na, Cl, and fluoride concentrations primarily from industrial effluents make many samples unfit for drinking without treatment.

Finally, long term and stable strategies are important, including the implementation of climate water management and the restoration of damaged ecosystems. Continuous improvement of water quality will be guaranteed as a national education campaign, a political compliance improvement and a general automatic monitoring of important indicators. In addition to solving the current problems, these measures protect the water supply of future generations to ensure the wells of the ecosystem and the population.

**4. Conclusion**

The principal component analysis (PCA) performed according to the parameters of the water quality provided the penetration analysis of the main tendency of the data. International studies further support the robustness of PCA as an analytical tool PCA’s utility in distinguishing pollution origins, whether from wastewater, urban runoff, or mining. Analysis, including 16 principal components (from PC1 to PC16), showed the relationship between a variety of patterns and temperature, pH, conductivity, oxygen (BOD), nitrogen (BOD), nitrogen, stool cake, general column and general dissolution solids (TDS). Thus, the PCA results provide a multifaceted type of water quality. The PC1 distinguishes the tendency to increase high conductivity and contamination associated with coliforms. Assuming complex interactions between pH, nitrogen and temperature, PC2 contrasts with nutrients and pollution levels. The PC3 highlights the importance of pH change, while the PC4 confines more specific or subtle conditions of water quality. This thoughtful is important for effectively managing water quality and helps to control the main parameters for monitoring and intervention, to solve the pollution problem and make sure safe and healthy water conditions.

**Disclaimer (Artificial intelligence)**

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

**References**

Abdel-Shafy, H. I., & Mansour, M. S. M. (2016). A review on polycyclic aromatic hydrocarbons: Sources, environmental impact, effect on human health, and remediation. *Egyptian Journal of Petroleum*, 25, 107–123. <https://doi.org/10.1016/j.ejpe.2015.03.011>.

Ahmad, S., & Khurshid, S. (2019). Hydrogeochemical assessment of groundwater quality in parts of the Hindon River basin, Ghaziabad, India: implications for domestic and irrigation purposes. SN Applied Sciences, 1, 1-12. <https://doi.org/10.1007/s42452-019-0161-9>.

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. <https://doi.org/10.1038/s41598-024-56056-8>.

Aref, F., & Roosta, R. (2016). Assessment of groundwater quality and hydrochemical characteristics in Farashband plain, Iran. Arabian Journal of Geosciences, 9, 1-14. <https://doi.org/10.1007/s12517-016-2781-3>.

Chaudhary, V., & Satheeshkumar, S. (2018). Assessment of groundwater quality for drinking and irrigation purposes in arid areas of Rajasthan, India. Applied Water Science, 8, 1-17. <https://doi.org/10.1007/s13201-018-0865-9>.

Ige, O. O., Adewoye, F. O., & Obasaju, D. O. (2021). Hydrochemical evaluation of groundwater quality: a case study from parts of North-Central, Nigeria. Sustainable Water Resources Management, 7, 1-16. <https://doi.org/10.1007/s40899-021-00577-x>.

Islam, S. D. U., Majumder, R. K., Uddin, M. J., Khalil, M. I., & Ferdous Alam, M. (2017). Hydrochemical characteristics and quality assessment of groundwater in Patuakhali district, southern coastal region of Bangladesh. Exposure and health, 9, 43-60. <https://doi.org/10.1007/s12403-016-0221-y>.

Jain, S., Bargah, R. K., &amp; Vaishnav, M. (2024). Study of Seasonal Variability of Ground Water Quality in Sarguja District (C.G.), India. Asian Journal of Environment &amp; Ecology, 23(12), 120 133.https://doi.org/10.9734/ajee/2024/v23i12638.

Kazi, T. G., Arain, M. B., Jamali, M. K., Jalbani, N., Afridi, H. I., & Sarfraz, R. A. (2009). Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. *Ecotoxicology and Environmental Safety*, 72(2), 301–309.<https://doi.org/10.1016/j.ecoenv.2008.02.024.>

Lal, B., Sengar, S. S., Singh, R., Jhariya, M. K., & Raj, A. (2023). Hydrogeochemistry and groundwater quality assessment in Ambagarh Chowki, Chhattisgarh, India. *Environmental Monitoring and Assessment*, *195*(1), 43. <https://doi.org/10.1007/s10661-022-10650-3>.

Mishra, A. (2010). Assessment of water quality using principal component analysis: a case study of the river Ganges. *Journal of Water Chemistry and Technology*, *32*, 227-234. <https://doi.org/10.3103/S1063455X10040077>.

Mustapha, A., & Abdu, A. (2012). Application of principal component analysis & multiple regression models in surface water quality assessment. *Journal of environment and earth science*, *2*(2), 16-23.

Ndah Anyang, G. B., Ngwabie, N. M., &amp; Ayonghe, S. N. (2021). Evaluation of the Variability of Drinking Water Quality within Bamenda Metropolis, North West Region, Cameroon. Current Journal of Applied Science and Technology, 40(27),54–70. <https://doi.org/10.9734/cjast/2021/v40i2731527>.

Ravikumar, P., & Somashekar, R. K. (2017). Principal component analysis and hydrochemical facies characterization to evaluate groundwater quality in Varahi river basin, Karnataka state, India. *Applied Water Science*, *7*, 745-755. <https://doi.org/10.1007/s13201-015-0287-x>.

Sharma, D. A., Rishi, M. S., & Keesari, T. (2017). Evaluation of groundwater quality and suitability for irrigation and drinking purposes in southwest Punjab, India using hydrochemical approach. *Applied Water Science*, *7*, 3137-3150. <https://doi.org/10.1007/s13201-016-0456-6>.

Sharma, N., Vaid, U., & Sharma, S. K. (2021). Assessment of groundwater quality for drinking and irrigation purpose using hydrochemical studies in Dera Bassi town and its surrounding agricultural area of Dera Bassi Tehsil of Punjab, India. *SN Applied Sciences*, *3*, 1-13. <https://doi.org/10.1007/s42452-021-04199-y>.

Sheela, A. M., Letha, J., Joseph, S., Chacko, M., Sanal Kumar, S. P., & Thomas, J. (2012). Water quality assessment of a tropical coastal lake system using multivariate cluster, principal component and factor analysis. *Lakes & Reservoirs: Research & Management*, *17*(2), 143-159. <https://doi.org/10.1111/j.1440-1770.2012.00506.x>.

Singh, S., Raju, N. J., & Ramakrishna, C. (2015). Evaluation of groundwater quality and its suitability for domestic and irrigation use in parts of the Chandauli-Varanasi region, Uttar Pradesh, India. *Journal of Water Resource and Protection*, *7*(07), 572. [10.4236/jwarp.2015.77046](http://dx.doi.org/10.4236/jwarp.2015.77046).

Sudhakaran, S., Mahadevan, H., Arun, V., Krishnakumar, A. P., & Krishnan, K. A. (2020). A multivariate statistical approach in assessing the quality of potable and irrigation water environs of the Netravati River basin (India). *Groundwater for Sustainable Development*, *11*, 100462. <https://doi.org/10.1016/j.gsd.2020.100462>.

Sukanya, S., & Sabu, J. (2020). Water quality assessment using environmetrics and pollution indices in a tropical river, Kerala, SW Coast of India. *Current World Environment*, *15*(1), 11.

Tripathi, M. and Singal, S.K., 2019. Use of principal component analysis for parameter selection for development of a novel water quality index: a case study of river Ganga India. *Ecological indicators*, *96*, pp.430-436. <https://doi.org/10.1016/j.ecolind.2018.09.025>.

Venkatesan, D., Gandhi, M. S., Vamsi, K. S., Das, P., Reddy, N. G., & Gayathri, G. S. (2022, April). Hydrogeochemical Characterization, Groundwater Quality and Water Resource Management in Salem, Tamil Nadu, India. In *Recent Developments in Sustainable Infrastructure (ICRDSI-2020)—GEO-TRA-ENV-WRM: Conference Proceedings from ICRDSI-2020 Vol. 2* (pp. 47-60). Singapore: Springer Singapore. <https://doi.org/10.1007/978-981-16-7509-6_5>.