Original Research Article

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

**A study of wells in Kerala**

**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:** Various wells located across districts in Kerala, India. Data were sourced from the Central Pollution Control Board (CPCB) under the National Water Quality Monitoring Programme (NWMP), 2022.**Methodology:** Water quality data from wells were collected and standardized for analysis. Sixteen parameters were included such as temperature, pH, conductivity, total/fecal coliforms, nitrogen, biochemical oxygen demand (BOD), total dissolved solids (TDS), and fluoride. PCA was applied using R software to reduce dimensionality and identify principal components (PCs). The first four principal components (PC1–PC4) were analyzed in depth, explaining 68.4% of total variance. Each well was scored on the PCs to evaluate dominant trends and regional variations in water quality.**Results:** PC1 (33.33%) revealed a general pattern of water quality deterioration, with higher conductivity and total coliform levels. PC2 (21.06%) highlighted variation due to pH and nitrogen, where wells like Punalur and Malappuram showed higher nutrient levels. PC3 (14%) focused on pH variability, and PC4 (11.06%) reflected localized anomalies. For example, the well at Vytilla (Ernakulam) had poor water quality (PC1 score: –9.33), while Kannur Municipality well showed high contamination (PC1 score: +2.26). These patterns suggest influences from sewage, agriculture, and industrial runoff.**Conclusion:** Principal Component Analysis effectively highlighted the key factors and spatial patterns influencing groundwater quality in Kerala. The study underscores the need for targeted water quality management through improved sanitation, pollution control, and localized monitoring strategies to ensure sustainable and safe groundwater resources. |

*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. 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. 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.

Shrestha and Kazama studied the Fuji River Basin in Japan and analysed that seasonal changes impacted pollution, mostly due to agricultural runoff and urban waste Shrestha & Kazama 2007. Panda *et al*., 2018 examined the groundwater in Odisha, India, and identified that fluoride, manganese, and iron are the major pollutants caused by both natural processes and human activities. In the Indus River Basin, Kazi *et al.*, 2009 used PCA to demonstrate that industrial discharge and agricultural practices extensively affected the quality of water. Vijith et al. 2014 studied groundwater in Palakkad, Kerala, and observed that salinity and nutrient levels were influenced by land use and natural factors. A similar study was reported by Dhanachandra et al. 2021 who studied the groundwater in Tamil Nadu and observed heavy metal pollution associated with the industrial activities. In order to examine surface water, Sharma *et al.,* 2022 employed PCA in the Ganga Basin and revealed that agricultural land usage improved the levels of nitrate and phosphate.

The shallow wells in Kerala were assessed by Kumar *et al.,* 2021 who observed that microbial contamination was primarily caused by poor sanitation and waste management. Sujitha *et al*., 2019 studied about the coastal aquifers in Kerala in which he stated that elevated levels of lead and cadmium are due to the industrial activities and fishing harbors. Further studies in India also demonstrate how PCA is helpful in water quality evaluation. Kumar et al. 2020 studied groundwater in Uttar Pradesh and found that nitrate levels were mainly due to the use of fertilizers (Kumar *et al*., 2020). Tiwari *et al*., 2017 looked at the Kosi River Basin in Bihar and according to him the most important reason for water pollution is domestic sewage and agricultural runoff. Nag & Ghosh 2013 utilized PCA in West Bengal to analyze arsenic contamination in groundwater, which was caused by natural geological processes. The groundwater in Karnataka was studied by (Mallikarjuna *et al*., 2019), who observed that fluoride levels were influenced by natural geochemistry and irrigation. Das & Mandal. 2020 studied the groundwater in Assam and stated that iron pollution is caused by both natural factors and human activities. Srinivas *et al*., 2023 used PCA in Telangana to assess groundwater for irrigation and observed that salinity and sodium adsorption ratio (SAR) were the most significant factors. sodium adsorption ratio (SAR) were the most significant factors. Outside India, Shrestha *et al*., 2010 used PCA to learn about the rivers in Nepal and discovered that domestic sewage and industrial waste are the main sources of pollution. Singh et al. 2004 studied surface water in the Gomti River Basin and revealed that high BOD and coliform levels were the main contributors to water quality issues. Yidana *et al*., 2008 used PCA in Ghana and revealed that groundwater salinity and nitrate levels were caused by urban and agriculture activities. Further studies from around the world also emphasize PCA’s effectiveness. Phung *et al*., 2015 used PCA in Vietnam to study groundwater, and he showed that industrial pollution and pesticide are the main problems. Ouyang *et al*. 2006 examined the surface water in China and identified that seasonal changes in pollution were caused by agriculture and industry. Mohamed *et al*., 2018 studied groundwater in Egypt’s Nile Delta and discovered that irrigation and wastewater practices were affecting water quality. Zhu *et al*., 2011 studied the rivers in Malaysia and he found that mining and urban runoff were the main sources of pollution. In Turkey, Kannel *et al*., 2007 utilized the PCA to examine the Melen River Basin, highlighting agricultural runoff and domestic sewage as the major problems. Ahmed *et al*., 2020 applied PCA in Bangladesh to observe the groundwater and found that arsenic contamination was caused mostly by natural geological factors. These reports illustrate how PCA simplifies complex data, making it easier to recognize the major causes of water pollution. It is an efficient tool for monitoring the water quality, particularly in regions like Kerala, where both natural and human activities affect groundwater. 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 anoxygenic 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, eigenvalues and eigenvectors are calculated in which eigenvalues 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 analysed 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, PCAtools, 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. 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 |
| Wellat Eloor, 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 |
| Wellat Chungapally, 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 |
| Wellat Ollur (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 |
| Wellat Laloor (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 scree 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.

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. It may be possible to see such an approach as a hierarchy that encompasses all possible approaches to water resource development and contaminant control.

**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.

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. The importances of pH-related loadings imply that this component is important for understanding variations in the acidity or alkalinity of the water.

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. 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.

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.

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. 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.

**Consent (where ever applicable)**

Not Applicable

**Ethical approval (where ever applicable)**

Not Applicable

**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. *Egypt. J. Pet.*, 25, 107–123.

Ahmed, M. J., Islam, M. N., & Rahman, M. M. (2020). Arsenic contamination in groundwater in Bangladesh: A case study using principal component analysis. *Environ. Sci. Pollut. Res.*, 27(3), 3426–3438.

Das, S., & Mandal, M. (2020). Assessment of iron contamination in groundwater in Assam, India. *Environ. Earth Sci.*, 79(6), 151.

Dhanachandra, L. K., Rajmohan, N., & Venkatesan, G. (2021). Groundwater quality and health risk assessment in Tamil Nadu, India: An integrated approach. *Environ. Sci. Pollut. Res.*, 28(15), 19020–19037.

Kannel, P. R., Lee, S., Lee, Y. S., Kannel, S. R., & Khan, S. P. (2007). Application of water quality indices and multivariate analysis for river water quality assessment: A case study of the Melen River, Turkey. *Water Air Soil Pollut.*, 189(1–4), 91–102.

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. *Ecotoxicol. Environ. Saf.*, 72(2), 301–309.

Kumar, N., Rajeev, R., & Suresh, G. (2021). Microbial contamination in shallow wells: A study from Kerala, India. *Water Sci. Technol.*, 84(12), 3482–3492.

Mallikarjuna, A. S., Kumar, B., & Reddy, M. V. (2019). Influence of natural geochemistry and irrigation on fluoride levels in groundwater of Karnataka, India. *Environ. Earth Sci.*, 78(1), 22.

Mohamed, H. A., & El-Zeiny, A. (2018). Impact of irrigation and wastewater practices on groundwater quality in the Nile Delta. *Water Sci. Technol.*, 78(10), 2034–2046.

Nag, S. K., & Ghosh, P. (2013). Arsenic contamination in groundwater in West Bengal: A statistical analysis using principal component analysis. *J. Water Resour. Prot.*, 5(8), 789–799.

Ouyang, Y., Nkedi-Kizza, P., Wu, Q. T., Shinde, D., & Huang, C. H. (2006). Assessment of seasonal variations in surface water quality. *Water Res.*, 40(20), 3800–3810.

Panda, S. K., Mishra, P. K., & Bastia, A. K. (2018). Groundwater quality assessment using water quality index and GIS techniques: A case study from Baripada block, Odisha, India. *Groundw. Sustain. Dev.*, 7, 170–180.

Phung, D., Huang, Y., & Yang, Z. (2015). Groundwater quality assessment in Vietnam using multivariate statistical techniques. *Environ. Monit. Assess.*, 187(10), 654.

Sharma, R., Bhardwaj, R., & Sharma, A. (2022). Multivariate statistical techniques for evaluating water quality of Ganga River Basin. *Environ. Monit. Assess.*, 194, 679.

Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan. *Environ. Model. Softw.*, 22(4), 464–475.

Singh, K. P., Malik, A., & Sinha, S. (2004). Water quality assessment and apportionment of pollution sources of Gomti River using multivariate statistical techniques: A case study. *Anal. Chim. Acta*, 538(1–2), 355–374.

Srinivas, Y., Rao, G., & Babu, D. (2023). Evaluation of groundwater quality for irrigation in Telangana, India. *Agric. Water Manag.*, 275, 107963.

Sujitha, S., Kumar, M., & Anilkumar, N. (2019). Assessment of heavy metal contamination in coastal aquifers of Kerala, India. *Environ. Monit. Assess.*, 191(4), 231.

Vijith, H., Sreedevi, P. D., & Das, S. (2014). Influence of anthropogenic activities on groundwater quality in Palakkad District, Kerala. *Environ. Earth Sci.*, 72(7), 2547–2562.

Zhu, X., Chen, Q., & Zheng, W. (2011). Mining and urban runoff as major sources of pollution in Malaysian rivers. *Environ. Earth Sci.*, 63(6), 1367–1377.