Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 11, No. 11, November, 2025

A Comparative Study of PCA and KPCA for Groundwater Quality Index Estimation

Shokry Abdelaziz [1, 2*], Marwan Kheimi [1], Mohamed A. H. Eizeldin [2, 3], Hassan Safi Ahmed [4]

[1] *Department of Civil and Environmental Engineering, Faculty of Engineering—Rabigh Branch, King Abdulaziz University, Jeddah 21589, Saudi Arabia.*

[2] *Department of Civil Engineering, Faculty of Engineering—Materia, Helwan University, Cairo, Egypt.*

[3] *Department of Civil Engineering, Faculty of Engineering, The British University in Egypt, Cairo Governorate 11837, Egypt.*

[4] *Department of Civil Engineering, Faculty of Engineering, South Valley University, Qena, Egypt.*

Received 28 September 2024; Revised 21 October 2025; Accepted 25 October 2025; Published 01 November 2025

Abstract

Groundwater quality assessment is crucial for ensuring human welfare and promoting sustainable economic development. This study evaluates the effectiveness of linear Principal Component Analysis (PCA) and nonlinear Kernel PCA (KPCA) in developing a reliable Groundwater Quality Index (GWQI) for Qena, Egypt. Using ten hydrochemical parameters from seventy-three groundwater samples, we compare the performance of four kernel functions within the KPCA framework. The PCA-based GWQI classified 71.0% of samples as suitable for irrigation, closely aligning with the Wilcox Diagram classification (76.7%). In contrast, KPCA with linear, polynomial, sigmoid, and radial basis function kernels yielded suitability rates of 58.9%, 52.1%, 63.0%, and 58.9%, respectively. These values are consistent with USSL (53.4%) and Na% (53.4%) classifications. Notably, the sigmoid kernel in KPCA demonstrated stronger correlations with Key hydrochemical parameters, effectively capturing nonlinear data structures. These findings underscore the importance of accounting for nonlinearity in groundwater quality assessment and demonstrate the potential of KPCA to improve GWQI accuracy. This comparative analysis highlights KPCA's superiority over PCA for nonlinear datasets, providing enhanced tools for groundwater management and more reliable quality evaluations.

Keywords: Groundwater Quality Index (GWQI); Principal Component Analysis (PCA); Kernel PCA (KPCA); Water Quality Assessment.

1. Introduction

Groundwater constitutes a vital resource, supplying nearly half of the world's domestic water consumption [1]. Consequently, protecting groundwater quality is essential to ensure human well-being and foster sustainable economic development. To facilitate informed decision-making, the Groundwater Quality Index (GWQI) is widely used as a composite metric that condenses complex hydrochemical data into a single, interpretable score. Since Horton [2] introduced the concept of water quality index, numerous models have been developed, including Canadian Water Quality Index (CWQI) [3] and the National Sanitation Foundation (NSFWQI) [4]. However, in the absence of a universally accepted standard, GWQI formulations vary widely in terms of parameter selection, weighting schemes, and aggregation methods. [5, 6], leading to inconsistent assessments across studies and regions. Traditionally, PCA has been extensively used in GWQI development due to its ability to reduce dimensionality and reveal dominant patterns in

* Corresponding author: sabdualazez@kau.edu.sa

http://dx.doi.org/10.28991/CEJ-2025-011-11-02

hydrochemical datasets. PCA has enabled researchers to classify groundwater samples and assign relative weights to chemical constituents [7, 8].

Ali et al. [9] analyzed 50 groundwater samples from the Achnera block in Agra district, Northern India, using the GWQI and PCA to identify the key geochemical factors influencing groundwater contamination. The results indicated that the groundwater was generally unsuitable for drinking. PCA revealed strong factor loadings for pH, Na+, Ca^{2+}, HCO_3^-, and fluoride, suggesting a geogenic origin for fluoride. The authors concluded that groundwater from contaminated areas requires treatment before it can be considered safe for human consumption. Nevertheless, one of PCA's main limitations is its assumption of linearity. Linear models can oversimplify the complex interdependencies present in real-world groundwater systems, which are often shaped by diverse geological formations, anthropogenic activity, and dynamic hydrochemical conditions. Moreover, PCA-based WQI models have been shown to produce inconsistent results with minor changes in input data [7], raising concerns about their reliability in dynamic or heterogeneous environments.

To overcome these limitations, the scientific community has increasingly adopted nonlinear dimensionality reduction techniques. Methods such as ISOMAP [10], t-distributed Stochastic Neighbor Embedding (t-SNE) [11], and Kernel Principal Component Analysis (KPCA) [12-15] have demonstrated improved performance in capturing hidden nonlinear structures within environmental datasets. Among them, KPCA is particularly promising, as it projects input data into a higher-dimensional feature space using kernel functions—such as radial basis, polynomial, and sigmoid kernels—thus allowing for the effective modeling of complex, nonlinear relationships. Several studies have demonstrated the superiority of KPCA over PCA for dimensionality reduction in nonlinear contexts [15-17]. For instance, Zhang et al. [17] successfully applied KPCA to achieve three objectives simultaneously: reducing data dimensionality, suppressing noise, and extracting meaningful features from water quality sensor data. This preprocessing step significantly improved the performance of a subsequent Recurrent Neural Network (RNN) model in estimating dissolved oxygen concentrations.

In the DWT-KPCA-GWO-XGBoost hybrid model for river dissolved oxygen prediction, KPCA is employed to reduce the dimensionality of meteorological input variables. Specifically, eight meteorological parameters are transformed into four principal components that collectively retain over 90% of the original data's variance. These components are then integrated as input features into the Grey Wolf optimizer (GWO)-tuned XGBoost model, enhancing its predictive accuracy by capturing the nonlinear interdependencies among climatic factors.

Jibrin et al. [18] utilized KPCA to extract nonlinear features from groundwater quality data, which were subsequently classified using a density-based spatial clustering algorithm (DBSCAN) to identify distinct water quality zones. Despite these advantages demonstrated, KPCA has been rarely applied to the direct computation of GWQI. A review of the existing literature reveals a notable gap: there is a lack of studies that employed KPCA-derived component loadings or weights in the formulation of GWQI. This presents a promising avenue for future research to improve the accuracy and robustness of groundwater quality assessments.

Given the inherent limitations of the linear assumption in PCA and the demonstrated potential of KPCA to capture complex, nonlinear relationships in environmental data, this study aims to evaluate and compare the effectiveness of PCA and nonlinear KPCA in developing a GWQI. The research is specifically applied to 73 groundwater samples collected from Qena Governorate, Egypt, based on ten hydrochemical parameters. The objectives are threefold:

- To develop and compare GWQI models using PCA and KPCA with four distinct kernel functions (linear, polynomial, radial basis function (RBF), and sigmoid) to assess their respective abilities in weighting and aggregating water quality parameters.
- To evaluate the performance of each method in terms of (a) dimensionality reduction efficiency, (b) strength of correlation between derived components and original hydrochemical variables, and (c) accuracy in spatial classification of groundwater suitability for irrigation.
- To determine whether the incorporation of nonlinear feature extraction via KPCA leads to a more robust, accurate, and physically interpretable GWQI compared to the traditional linear PCA approach.

This comparative analysis seeks to address a critical gap in current groundwater assessment methodologies by rigorously testing the hypothesis that nonlinear dimensionality reduction techniques offer superior performance in capturing the complex geochemical interactions prevalent in real-world aquifer systems.

2. Study Area

The study area covers approximately 3,415.36 km² and is situated in the narrow Nile valley of Upper Egypt, within Qena Governorate, about 600 km south of Cairo (Figure 1). It lies within UTM zone 36N, with coordinates ranging from 393,760 N to 503,099 N and from 2,841,700 E to 2,915,210 E. The eastern part of the region is characterized by significantly higher topography, with elevations reaching up to 530 m above mean sea level (a.m.s.l.), while the Nile River floodplain remains relatively low-lying at approximately 50 m a.m.s.l. [19].

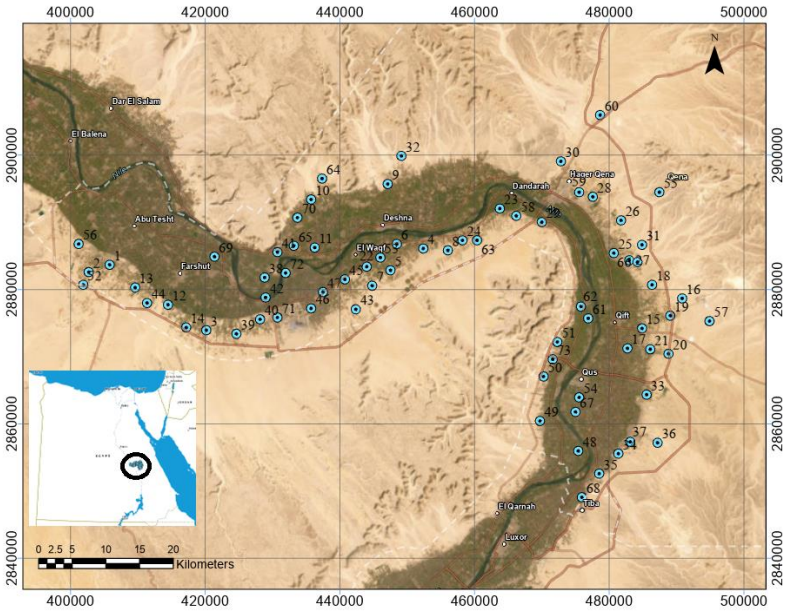


Figure 1. Map of the study area highlighting the locations of groundwater samples

Numerous researchers have previously investigated the geology of the Qena area, including [20-22]. As illustrated in Figure 2 [19], the geological succession in the study area from top to bottom is summarized as follows:

- Holocene Sediments: The uppermost layer consists primarily of silty clay derived from Wadi alluvium and the Nile floodplain [19].
- Late Pleistocene Sediments: This unit forms the primary aquifer in the region. The Quaternary aquifer is composed of sand and gravel with intercalated clay layers [19].
- Plio-Pleistocene Sediments: These deposits, associated with proto- and pre-Nile phases, consist of a heterogeneous mixture of clay, sand, and gravel [19].
- Pliocene deposits: Predominantly composed of clay with interbedded sand layers, these deposits serve as the basal confining layer for the overlying Quaternary aquifer [19].
- Eocene Limestone Rocks: This formation includes chalky and dolomitic limestone, marl, and localized concentrations of halite and gypsum, particularly in the western part of the Nile Valley [19].
- Paleocene-Late Cretaceous Shale: This thinly bedded unit contains alternating layers of chalk and phosphate [19].
- Late Cretaceous (Paleozoic) Sediment: Composed of sandstone with intercalated shale layers [19].

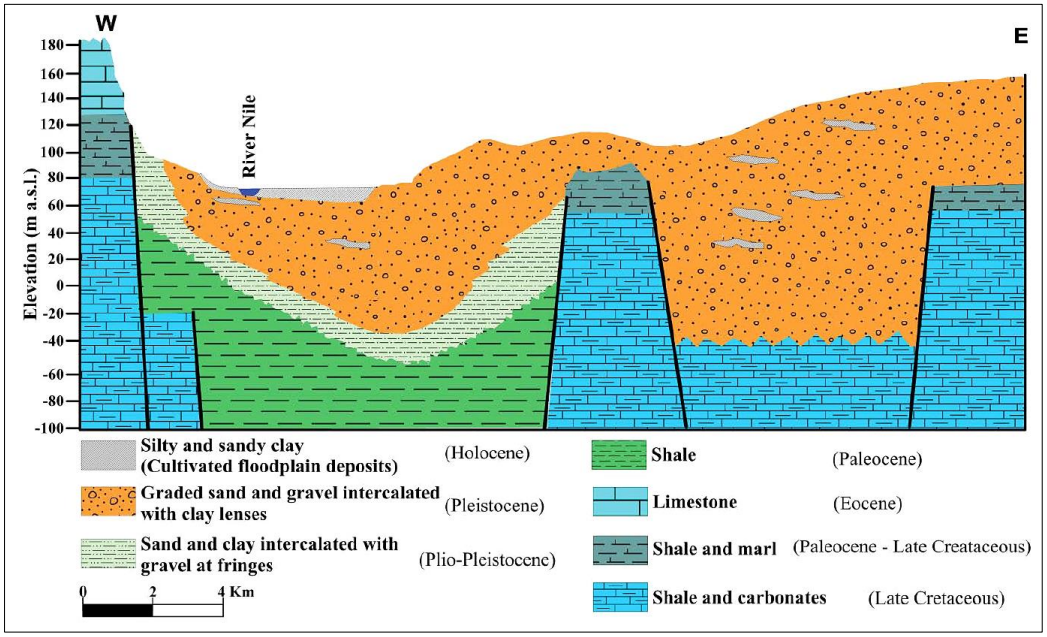


Figure 2. Geologic section of the region under investigation [19]

The Quaternary aquifer, consisting of fluvial sands and gravels with occasional thin clay interlayers, serves as the primary source of groundwater in the study area. It exhibits a transitional hydraulic behavior, functioning as a semi-confined aquifer beneath the Nile floodplain where the underlying Pliocene clay layer provides partial confinement and as an unconfined aquifer in the desert fringes where this confining layer is absent or discontinuous. The main source of recharge to the aquifer is Irrigation return flow [19].

Agricultural activities in the study area are primarily concentrated within the Nile Valley, with some expansion into adjacent desert regions where soil quality, land access, and water resources are allowed. Irrigation is supported by surface water drawn from the Nile River and its extensive canal system, as well as by groundwater extracted through a network of dispersed pumping wells. Farming practices in the region are marked by intensive fertilizer application, including magnesium phosphate, and the common use of lime during land preparation prior to cultivation [19].

3. Research Methodology

This study employs a groundwater hydrochemical dataset originally compiled by El-Rawy et al. [19]. The dataset is based on 73 groundwater samples collected from various locations across Qena Governorate, Upper Egypt, between July 2013 and January 2014. Samples were collected, filtered, preserved, and analyzed in the laboratory following standard hydrochemical protocols as shown in Figure 1. Ten key hydrochemical parameters were selected for analysis, categorized as follows:

- Physical parameters: Total Dissolved Solids (TDS), Electrical Conductivity (EC), and pH;
- Major cations: Calcium (Ca^{2+}), Magnesium (Mg^{2+}), Sodium (Na^{+}), and Potassium (K^{+});
- Major anions: Bicarbonate (HCO_3^{-}), Chloride (Cl^{-}), and Sulfate (SO_4^{2-}).

These parameters were chosen due to their significant influence on groundwater chemistry, their established role in Groundwater Quality Index (GWQI) development, their relevance to drinking water and irrigation suitability, and their alignment with international water quality guidelines, particularly those established by the Food and Agriculture Organization (FAO) [23]. Descriptive statistics for those parameters are presented in Table 1, along with the FAO maximum permissible limits [23]. and the percentage of samples conforming to these standards. The statistical analysis reveals that approximately 30% of the samples exceed the FAO threshold for Mg^{2+}, Na^{+}, and Cl^{-}.

Table 1. The statistical summary of groundwater's hydrochemical characteristics (after El-Rawy et al. [19])

Parameter	Units	Min	Max	Mean	Standard dev.	FAO Standard	% Valid samples
pH	µS/cm	6	9	7.81	0.66	6.5-8.5	86.3
EC	mg/L	8	6045	1506.8	961.6	3000	93.2
TDS	mg/L	224	4050	1034.9	640.5	2000	93.2
Ca^{2+}	mg/L	1	1240	80.0	160.7	400	97.3
Mg^{2+}	mg/L	8	191	54.3	45.3	60	**67.1**
Na^{+}	mg/L	24	865	175.5	129.7	207	**72.6**
K^{+}	mg/L	0.1	290	5.23	33.8	100	98.6
HCO_3^{-}	mg/L	0.8	392	196.3	87.7	500	100
Cl^{-}	mg/L	21	2130	344.9	331.2	350	**67.1**
SO_4^{2-}	mg/L	20	640	145.1	118.3	300	89

Elevated concentrations of Na^{+} and Cl^{-} are primarily attributed to evaporation during intensive irrigation and the subsequent return flow of irrigation water, which transports dissolved solutes into the aquifer. Increased Mg^{2+} levels may result from the application of nitrogen-based Fertilizers and the natural dissolution of dolomitic rock formations. High sodium content in irrigation water can disperse soil clay particles, degrading soil structure and reducing infiltration capacity [24]. An elevated Mg^{2+}/Ca^{2+} ratio may adversely affect soil quality and lead to calcium deficiency in crops [25]. furthermore, excessive chloride levels can cause soil salinization and induce chloride toxicity [25].

3.1. Principal Component Analysis (PCA)

PCA is a widely used technique for dimensionality reduction, particularly effective in capturing the complexity of high-dimensional datasets by transforming them into a lower-dimensional linear subspace while preserving the maximum amount of variance [26]. In groundwater quality assessment, PCA plays a pivoted role by classifying water samples and identifying the dominant geochemical process influencing their chemical composition. This is achieved by transforming a set of correlated variables into a smaller number of uncorrelated principal components [26], which collectively retain most of the information contained in the original dataset. The PCA procedures consists of several key steps: data standardization, covariance matrix computation, Eigenvalues decomposition and principal component selection and feature vector construction.

Standardization is a essential to ensure that all variables contribute equally to the analysis, regardless of their original scale or units. Without rescaling, variables with larger magnitudes could disproportionately influence the results. The standardized value (Z) for a variable X is calculated as:

$$Z = \frac{X_i - \bar{X}}{\sigma} \quad (1)$$

where, Z is standardized value, X is original value of the variable, $\bar{X}$ is mean of the variable, σ is standard deviation.

Next, the covariance matrix is computed to quantify the linear relationships between pairs of variables. The covariance between two variables X and Y is given by:

$$cov(X,Y) = \frac{1}{n-1}\sum_{n}^{1}(X_i - \bar{X})(Y_i - \bar{Y}) \quad (2)$$

Eigenvalues decomposition of the covariance matrix follows, Where the eigenvalues λ are obtained by solving the characteristic Equation:

$$|C - \lambda \mathrm{I}| = 0 \quad (3)$$

where, C is the covariance matrix and I is the identity matrix of the same dimension. The corresponding eigenvectors V are then determined by solving:

$$(C - \lambda \mathrm{I})V = 0 \quad (4)$$

According to Kaiser [27], principal components with higher eigenvalues retain more of the original data's variance and thus representing the most significant patterns in the dataset. The proportion of total variance explained by each component is calculated as $\lambda_i/\sum\lambda_j$, enabling the selection of subset of components that capture a desired percentage of the total variability.

3.2. Kernel Principal Component Analysis (KPCA)

While PCA is highly effective for linear data structures, it is inherently limited in its ability to capture nonlinear relationships among variables. To address this limitation, Kernel principal component analysis (KPCA) extends PCA by applying a kernel function to implicitly map the original data into a higher-dimensional feature space, where nonlinear patterns in the input space become linearly separable [28]. This transformation is achieved using the kernel trick, which avoids explicitly computing coordinates in the high-dimensional space. Instead, inner products between data points are replaced with a kernel function $K(\mathbf{x}_i,\mathbf{x}_j)$, which measures similarity in the transformed space [29].

The KPCA algorithm consists of three main steps:

- Compute a kernel matrix: $Kij=K(\mathbf{x}i,\mathbf{x}j)$ represents the similarity between data points in the feature space.
- Perform eigenvalue decomposition: Extract principal components by solving the eigenvalue problem on the kernel matrix, identifying directions of maximum variance.
- Project data: Transform the original dataset onto the most significant principal components to achieve nonlinear dimensionality reduction while preserving key structural patterns [30].

The choice of kernel function in KPCA significantly influences its performance especially in environmental applications such as groundwater quality analysis, where variable interactions are often complex and nonlinear [12]. In this study, four common kernel functions are evaluated:

- The linear kernel, which is equivalent to conventional PCA, captures only linear relationships in the data [31].

$$K(X_i, X_j) = X_i{}^T X_j \quad (5)$$

- For more complex interactions, the polynomial kernel can model higher-order dependencies among variables, making it suitable for capturing multiplicative effects [32].

$$K(X_i, X_j) = (aX_i{}^T X_j + b)^d \quad (6)$$

where a is scaling factor (controls the influence of the dot product), b is offset (or bias) term, d is degree of the polynomial (determines the order of interactions).

- The radial basis function (RBF) kernel, due to its ability to map data into an infinite-dimensional feature space, excels at identifying local nonlinear patterns, which is particularly useful for detecting subtle variations in water quality parameters [33].

$$K(X_i, X_j) = e^{\left(-\frac{\|X_i - X_j\|^2}{\sigma^2}\right)} \quad (7)$$

where, $\|X_i, -X_j\|$ is Euclidean distance between two data points, σ is kernel width (controls the radius of influence of each point).

- Additionally, the sigmoid kernel, inspired by neural network activation functions, offers further flexibility in modeling specific nonlinear relationships [34].

$$\text{Sigmoid: } K(X_i, X_j) = \tanh(\sigma X_i{}^T X_j + r) \quad (8)$$

where σ is scale parameter, r is shift (or bias) parameter.

By enabling the detection of complex, nonlinear relationships, KPCA is especially valuable for analyzing environmental datasets such as groundwater quality data, where variables interact through intricate, nonlinear processes. Its ability to uncover hidden patterns makes it a powerful tool for dimensionality reduction and feature extraction in real-world, nonlinear systems.

3.3. Groundwater Quality Index (GWQI) Estimation

The Groundwater Quality Index (GWQI) is computed through a multi-step process (illustrated in Figure 3) to integrate multiple hydrochemical parameters into a single, interpretable score that reflects overall groundwater suitability for irrigation. The procedure is outlined below::

- Scaling Hydrochemical Parameters: The collected hydrochemical parameters of groundwater (X_i) are scaled from 0 to 100. A score of 0 indicates that the measured concentration of the parameter (C_i) exceeds the maximum permissible limit (C_{is}) for irrigation water set by FAO [23], while a score of 100 represents the most desirable concentration (C_{id}).

$$X_i = 100 - \frac{100(C_i - C_{id})}{(C_{is} - C_{id})} \quad (9)$$

Note: If $C_i > C_{is}$, the score becomes less than 0 and is typically set to 0. Similarly, if $C_i \leq C_{id}$, the score is capped at 100.

- PCA or KPCA: The scaled hydrochemical data (X_i) values are used as input for dimensionality reduction via either PCA or KPCA:
  - In PCA, linear combinations of the scaled variables are used to compute eigenvalues, eigenvectors, and loadings.
  - In KPCA, a kernel function is first applied to the scaled data to capture nonlinear relationships, followed by eigen-decomposition in the transformed space to extract principal components and their loadings.
- Selection of Principal Components: based on the Kaiser criteria, Principal components with eigenvalues greater than one are retained for further analysis. For each selected component (i), a weight (Wi) is estimated based on the ratio of its eigenvalue (λ_i) to the sum of eigenvalues of all selected components.

$$\text{Wi} = \frac{\lambda_i}{\sum_n^1 \lambda_j} \quad (10)$$

The weights *Wi* represent the relative importance of each component in the final GWQI.

- Sorting of Principal Components: The selected principal components are then sorted based on the sign and magnitude of their loadings for key hydrochemical variables:
  - If higher values of the variables indicate poorer water quality, the component is sorted in descending order (i.e., higher scores reflect worse quality).
  - If higher values indicate better quality, sorting is done in ascending order.
- Rescaling Principal Component Scores (S_{ri}): The raw component scores (S_i) are rescaled to a 0–100 range to standardize their contribution to the index. The transformation depends on the sorting direction:
  - For ascending order (higher score = better quality):

$$S_{ri} = \frac{100(S_i - S_{min})}{(S_{max} - S_{min})} \text{ for ascending PCs} \quad (11)$$

  - For descending order (higher score = worse quality):

$$S_{ri} = 100 - \frac{100(S_i - S_{min})}{(S_{max} - S_{min})} \text{ for descending PCs} \quad (12)$$

where, S_{min}, S_{max} are Minimum and maximum scores across all samples for that component.

This rescaling ensures all components contribute consistently to the final index.

- GWQI Calculation: The final GWQI for each sample is computed as a weighted sum of the rescaled component scores:

$$GWQI = \sum_{i=1}^{n}(W_i \times S_{ri}) \quad (13)$$

The resulting GWQI ranges from 0 to 100 where: Very Poor: 0–25; Poor: 25–50; Fair: 50–65; Good: 65–80; Excellent: 80–100.

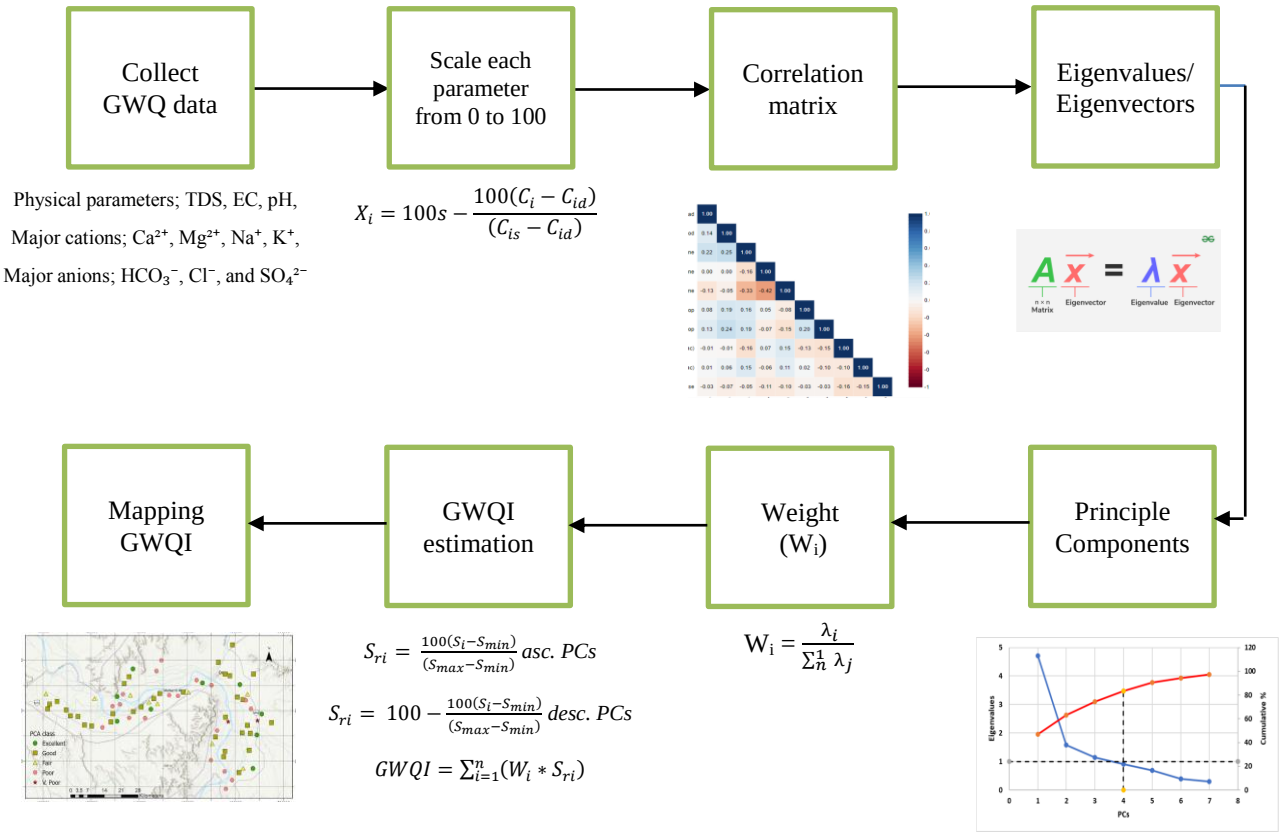


Figure 3. Graphical presentation of PCA-GWQI method

4. Results and Discussion

This section evaluates groundwater quality using a combination of conventional assessment techniques and advanced methodologies. The conventional methods include established salinity indices relevant for irrigation purposes, such as the Sodium Absorption Ratio (SAR), Soluble Sodium Percentage (SSP), Sodium Percentage (%Na), Residual Sodium Carbonate (RSC), Magnesium Hazard (MH), Permeability Index (PI), and Kelly Ratio (KR). These indices consider various hydrochemical parameters, such as dissolved salts, sodium content, magnesium concentration, and carbonate balance, to assess water toxicity and suitability for irrigation.

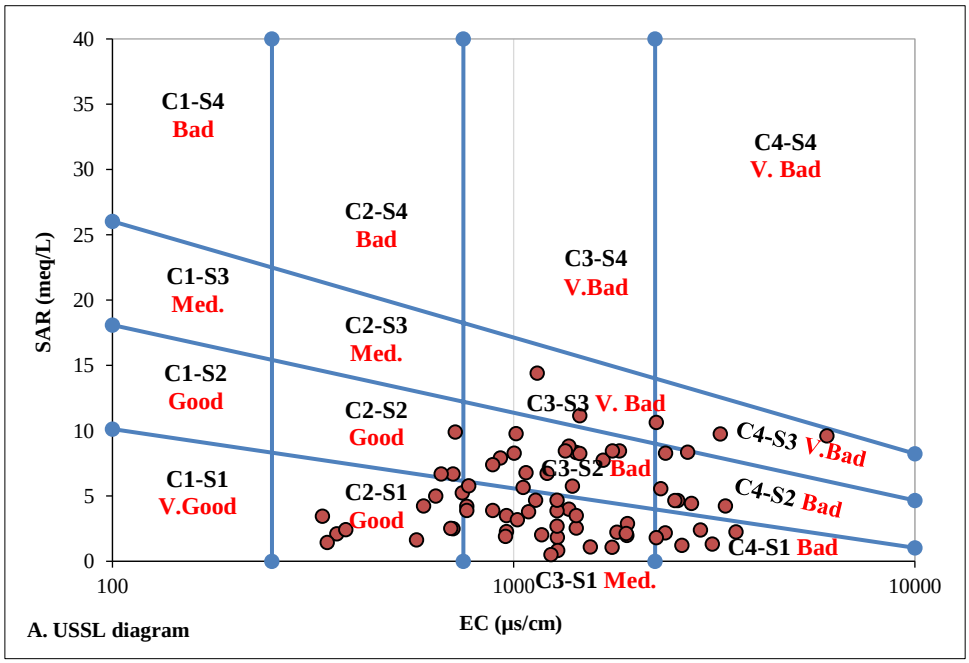
In addition to these conventional indices, newly developed approaches PCA and KPCA were employed. KPCA utilized four kernel functions: Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. Geographic Information Systems (GIS) were used to create spatial distribution maps of GWQI values derived from both PCA and KPCA methodologies, providing insights into the spatial patterns of groundwater quality.

Table 2 summarizes the results of the irrigation suitability assessments using these indices. The findings reveal significant variations in the percentage of samples classified as suitable for irrigation. For example: the Sodium Absorption Ratio (SAR) indicates that all samples are classified as suitable for irrigation, while Magnesium Hazard (MH) shows that only 12.3% are appropriate. This wide disparity in classification percentages among the different indices underscores the limitations of relying solely on conventional methods for groundwater quality assessment.

Table 2. Groundwater indices Based on hydrochemical parameters

Indices	Equation	Water Classification	No. well	% of Valid Samples
SAR: Sodium adsorption ratio [35]	$SAR = \frac{Na^+}{\sqrt{\frac{ca^{2+} + Mg^{2+}}{2}}}$	<10 Excellent 10-18 Good 18-26 Doubtful >26 Unsuitable	70 3 - -	100%
SSP: Soluble sodium percentage [36]	$SSP = \frac{Na^+}{ca^{2+} + Mg^{2+} + K^+} * 100$	<20 Excellent 20-40 Good 40-60 Permissible 60-80 Doubtful >80 Unsuitable	6 6 7 6 48	26%
Na%: Sodium percentage [37]	$\%Na = \frac{Na^+ + K^+}{ca^{2+} + Mg^{2+} + Na^+ + K^+}$	<20 Excellent 20-40 Good 40-60 Permissible 60-80 Doubtful >80 Unsuitable	7 14 18 25 9	53.4%
RSC: Residual Sodium Carbonate [35]	$RSC = HCO_3^- - (ca^{2+} + Mg^{2+})$	<1.25 safe 1.25-2.5 Marginal suitable >2.5 Unsuitable	62 6 5	93.2%
MH: Magnesium Hazard [38]	$MH = \frac{Mg^{2+}}{Mg^{2+} + ca^{2+}} * 100$	<50 % Suitable >50% unsuitable	9 64	12.3%
PI: Permeability index (PI) [39]	$PI = \frac{Na^+ + K^+ + \sqrt{HCO_3^-}}{(ca^{2+} + Mg^{2+} + Na^+ + K^+) \times 100}$	<60 Suitable 60-80 Permissible >80 Unsuitable	25 17 31	57.6 %
KR: Kelly Ratio [40]	$KR = \frac{Na^+}{Mg^{2+} + ca^{2+}}$	<1 Suitable >1 Suitable	29 44	39.7 %

Traditional graphical techniques such as the Piper diagram [41] and the Gibbs diagram [42] provide straightforward and rapid visualizations of groundwater chemistry. Figure 4 illustrates the results obtained from two such methods: the US Salinity Laboratory Staff (USSL) diagram [35] which assesses the combined risks of salinity and sodium hazard and the Wilcox diagram [43] which categorizes water based on EC and percent sodium. Application of these classification schemes revealed that 76.7% of the samples were classified as suitable for irrigation according to the Wilcox diagram, whereas the USSL classification indicated suitability for only 53.4%.



A. USSL diagram

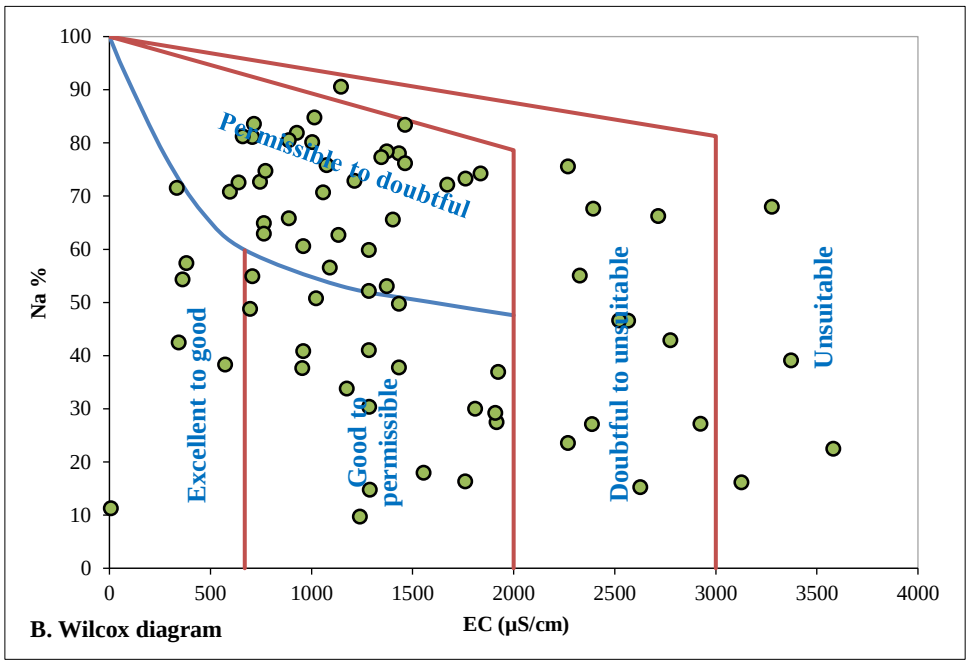


Figure 4. Classification of irrigation waters quality based on a. USSL diagram and b. Wilcox Diagram

An alternative approach to water quality assessment involves computing the GWQI. This study employs two techniques: PCA and KPCA. Both algorithms were implemented in Python using the scikit-learn library [44]. Figure 5 shows the relationship between eigenvalues and principal components (PCs). Based on this plot, the first four PCs were selected, as they collectively account for 83% of the total variance.

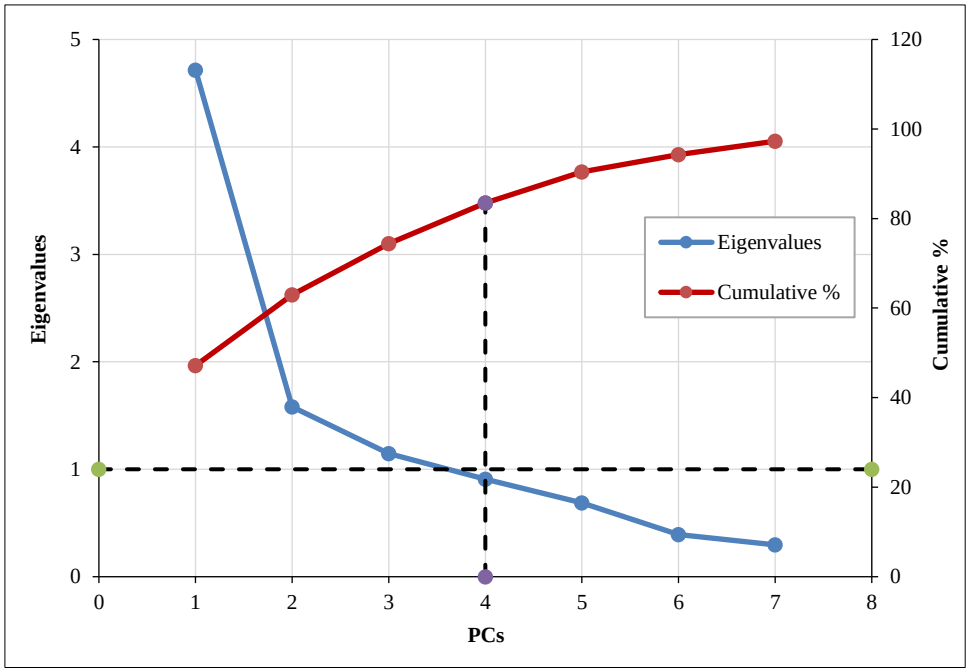


Figure 5. The scree plot and cumulative eigenvalue (%) of the water quality parameters

Groundwater flows from higher elevations into the Nile, with hydraulic heads between 36.34 and 504.34 m above sea level, supporting previous findings that the Nile gains water at Qena, while groundwater depth increases from 3 to 40 m toward the eastern and western desert edges, raising irrigation extraction costs [19, 45].

Table 3 presents the loadings of hydrochemical parameters on each principal component (PC). PC1, which explains 47.2% of the total variance, is influenced by six parameters: EC, TDS, Ca^{2+}, Mg^{2+}, Cl^{-}, and SO_4^{2-}. These loadings are negative, indicating that higher values of PC1 correspond to lower concentrations of these parameters. PC2 accounts for 15.8% of the variance and is primarily associated with K^{+}, while PC3 (11.5% of the variance) is dominated by pH. Finally, PC4 explains 9.1% of the variance and is linked to HCO_3^{-}. Na^{+} exhibits moderate loadings across all four PCs.

Table 3. Principal component loadings and adjusted parameter weight

Variable	PC1	PC2	PC3	PC4	Weight	Rel. Weight
	Desc.	Asc.	Asc.	Asc.		
Eigen V.	**4.716**	**1.581**	**1.147**	**0.907**		
pH	0.162	0.252	**0.809**	0.334	0.286	0.061
EC	**-0.941**	-0.247	0.032	-0.086	0.592	0.126
TDS	**-0.975**	0.025	0.037	0.035	0.564	0.120
Ca^{2+}	**-0.784**	0.404	-0.269	0.137	0.571	0.121
Mg^{2+}	**-0.784**	0.128	0.171	-0.252	0.518	0.110
Na^{+}	-0.426	-0.567	0.398	0.295	0.435	0.092
K^{+}	-0.170	**0.817**	0.020	0.359	0.292	0.062
HCO_3^{-}	-0.275	-0.445	-0.392	**0.663**	0.365	0.078
Cl^{-}	**-0.848**	-0.165	0.230	-0.222	0.566	0.120
SO_4^{2-}	**-0.787**	0.250	-0.150	0.014	0.514	0.109

The relative weight of each parameter can be calculated using these PC loadings and their corresponding eigenvalues. As expected, pH, K^+, and HCO_3^- have minimal weights due to their limited contribution to PC1, the dominant component. In contrast, EC, TDS, Ca^{2+}, and Cl^- exhibit the highest weights, reflecting their strong influence on PC1 and their substantial contribution to the explained variance.

Table 4 summarizes the results of groundwater quality assessment using the PCA-based method. For each sample, it presents the principal component (PC) scores, rescaled PC scores (computed using Equations 11 and 12, where applicable), and the final GWQI value along with its corresponding water quality classification.

Table 4. Irrigation groundwater quality index and classification based on PCA method

	PCs Score				PCs rescaled score				GWQI	Class
PCS	**PC1**	**PC2**	**PC3**	**PC4**	**PC1**	**PC2**	**PC3**	**PC4**		
Eigen V.	4.716	1.581	1.147	0.907	4.716	1.581	1.147	0.907		
Order	Desc	Asc	Asc	Asc	Desc	Asc	Asc	Asc		
Wt.	0.565	0.189	0.137	0.109	0.565	0.189	0.137	0.109		
1	-1.724	0.855	-1.259	-1.374	0.816	0.892	0.178	0.341	69.113	Good
2	0.069	-0.905	-0.381	1.266	0.579	0.729	0.333	0.938	61.257	Good
3	-0.546	1.049	-0.964	-1.861	0.660	0.910	0.230	0.231	60.167	Good
4	1.972	0.315	-1.573	0.801	0.327	0.842	0.122	0.833	45.142	Poor
5	-0.051	0.606	-1.093	-0.077	0.595	0.869	0.207	0.634	59.768	Fair
6	-2.081	-0.398	0.127	-0.146	0.863	0.776	0.422	0.619	75.960	Good
7	-2.797	-0.246	1.603	1.260	0.958	0.790	0.683	0.936	88.609	Excellent
8	3.135	0.013	0.225	-0.586	0.173	0.814	0.440	0.519	36.876	Poor
9	1.420	0.904	-1.564	0.604	0.400	0.896	0.124	0.788	49.834	Poor
10	0.060	-0.947	-0.365	1.541	0.580	0.725	0.335	1.000	61.961	Good
11	0.833	-0.339	0.747	-0.330	0.478	0.782	0.532	0.577	55.349	Fair
12	0.181	1.416	0.931	0.081	0.564	0.944	0.564	0.670	64.748	Good
13	-2.217	-0.888	-1.011	0.461	0.881	0.731	0.221	0.756	74.852	Good
14	-1.468	1.084	-0.032	-1.535	0.782	0.913	0.394	0.304	70.180	Good
15	-2.013	0.122	-1.019	-0.947	0.854	0.824	0.220	0.437	71.615	Good
16	-2.710	-0.485	1.634	1.492	0.947	0.768	0.688	0.989	88.187	Excellent
17	-1.440	0.805	0.220	-1.908	0.778	0.887	0.439	0.220	69.178	Good
18	4.446	0.267	1.293	0.375	0.000	0.838	0.628	0.736	32.483	Poor
19	4.248	-0.823	-1.414	0.167	0.026	0.737	0.150	0.689	24.981	V. Poor
20	-2.739	0.275	0.435	-0.941	0.950	0.838	0.477	0.439	80.848	Excellent
21	0.277	0.488	1.213	0.730	0.551	0.858	0.614	0.817	64.682	Good
22	0.409	0.762	-1.241	0.109	0.534	0.883	0.181	0.676	56.704	Fair
23	4.353	-0.302	-0.485	0.136	0.012	0.785	0.314	0.682	27.282	Poor
24	-1.004	0.182	-0.757	0.460	0.721	0.830	0.266	0.756	68.279	Good

	PCs Score				PCs rescaled score				GWQI	Class
PCS	**PC1**	**PC2**	**PC3**	**PC4**	**PC1**	**PC2**	**PC3**	**PC4**		
Eigen V.	4.716	1.581	1.147	0.907	4.716	1.581	1.147	0.907		
Order	Desc	Asc	Asc	Asc	Desc	Asc	Asc	Asc		
Wt.	0.565	0.189	0.137	0.109	0.565	0.189	0.137	0.109		
25	-2.339	-0.933	0.158	0.593	0.897	0.727	0.428	0.786	78.845	Good
26	-2.277	0.132	1.863	0.477	0.889	0.825	0.729	0.760	84.090	Excellent
27	-0.143	0.542	1.652	0.161	0.607	0.863	0.691	0.688	67.586	Good
28	-1.584	0.392	0.775	-1.351	0.797	0.849	0.537	0.346	72.241	Good
29	2.791	0.710	2.142	-0.149	0.219	0.878	0.778	0.618	46.390	Poor
30	-0.037	-1.034	-0.233	1.418	0.593	0.717	0.359	0.972	62.548	Good
31	1.150	-0.664	-0.016	0.946	0.436	0.752	0.397	0.866	53.699	Fair
32	1.862	-0.639	-0.054	0.463	0.342	0.754	0.390	0.756	47.143	Poor
33	-2.501	-0.277	-0.077	0.105	0.919	0.787	0.386	0.675	79.435	Good
34	-0.962	1.367	0.275	-2.081	0.715	0.939	0.448	0.181	66.296	Good
35	4.125	0.450	0.266	-1.147	0.042	0.854	0.447	0.392	28.972	Poor
36	-3.047	-0.892	0.093	0.696	0.991	0.730	0.416	0.809	84.303	Excellent
37	1.032	2.027	1.466	-0.956	0.452	1.000	0.659	0.435	58.208	Fair
38	3.666	0.540	0.378	1.446	0.103	0.863	0.467	0.978	39.195	Poor
39	2.324	1.322	0.440	0.850	0.281	0.935	0.477	0.844	49.274	Poor
40	-1.030	-0.371	-0.690	0.916	0.724	0.779	0.278	0.859	68.792	Good
41	-3.115	-0.675	0.252	0.398	1.000	0.751	0.444	0.742	84.841	Excellent
42	-1.380	0.195	-0.251	-0.383	0.771	0.831	0.355	0.565	70.265	Good
43	1.900	1.189	-0.915	-0.404	0.337	0.923	0.238	0.560	45.844	Poor
44	-0.829	-0.565	0.577	-0.217	0.698	0.761	0.502	0.603	67.234	Good
45	-1.084	0.794	-0.810	-0.476	0.731	0.886	0.257	0.544	67.518	Good
46	4.358	-0.086	3.400	-1.012	0.012	0.805	1.000	0.423	34.222	Poor
47	-1.391	0.153	-0.107	-0.435	0.772	0.827	0.381	0.553	70.495	Good
48	1.195	-0.670	-1.779	1.146	0.430	0.751	0.086	0.911	49.571	Poor
49	1.832	-0.156	-0.843	0.818	0.346	0.798	0.251	0.837	47.176	Poor
50	0.970	1.188	-0.416	-1.043	0.460	0.923	0.326	0.416	52.428	Fair
51	3.755	-0.043	2.525	0.172	0.091	0.809	0.846	0.690	39.587	Poor
52	-2.461	-0.063	0.456	-0.445	0.914	0.807	0.480	0.551	79.455	Good
53	-1.350	-0.515	0.152	0.607	0.767	0.765	0.427	0.789	72.208	Good
54	-2.283	-0.005	-0.051	-0.207	0.890	0.812	0.391	0.605	77.578	Good
55	-2.171	0.901	-0.197	-2.353	0.875	0.896	0.365	0.119	72.695	Good
56	0.442	-0.471	0.046	0.056	0.530	0.769	0.408	0.664	57.291	Fair
57	-1.282	-0.472	-0.268	0.816	0.758	0.769	0.352	0.836	71.268	Good
58	-2.513	-0.920	-0.115	1.011	0.920	0.728	0.380	0.880	80.533	Excellent
59	-2.335	-0.326	0.090	0.061	0.897	0.783	0.416	0.665	78.407	Good
60	-0.045	-0.330	-0.402	0.356	0.594	0.782	0.329	0.732	60.826	Good
61	2.547	-8.804	0.013	-2.882	0.251	0.000	0.402	0.000	19.700	V. Poor
62	-0.394	0.669	1.274	0.460	0.640	0.875	0.625	0.756	69.493	Good
63	-0.243	0.302	-2.265	0.250	0.620	0.841	0.000	0.708	58.634	Fair
64	-2.733	-0.354	-0.314	0.323	0.949	0.780	0.344	0.725	80.993	Excellent
65	2.019	0.578	-1.415	0.548	0.321	0.866	0.150	0.776	45.015	Poor
66	1.720	-0.112	-0.298	0.460	0.361	0.803	0.347	0.756	48.531	Poor
67	0.019	-0.463	-0.251	0.584	0.586	0.770	0.356	0.784	61.044	Good
68	2.686	0.918	-1.070	0.083	0.233	0.898	0.211	0.670	40.315	Poor
69	0.148	0.407	-1.747	-0.376	0.568	0.850	0.091	0.567	55.612	Fair
70	-1.728	-0.032	0.784	-0.723	0.817	0.810	0.538	0.488	74.139	Good
71	-3.004	-0.425	1.755	1.353	0.985	0.774	0.710	0.958	90.436	Excellent
72	3.405	0.451	-0.695	-0.078	0.138	0.854	0.277	0.634	34.651	Poor
73	-0.300	1.257	-0.821	-0.630	0.628	0.929	0.255	0.509	62.065	Good

Figure 6 illustrates the spatial distribution of GWQI values estimated using PCA method. The analysis reveals that 9 out of 73 wells (12.3%) have excellent water quality for irrigation, 34 (46.6%) are classified as good, 9 (12.3%) as fair, 19 (26%) as poor, and 2 (2.7%) as very poor. Overall, 52 wells (71%) fall within the "suitable" category (excellent to good), indicating that a majority of the groundwater sources assessed are appropriate for irrigation purposes.

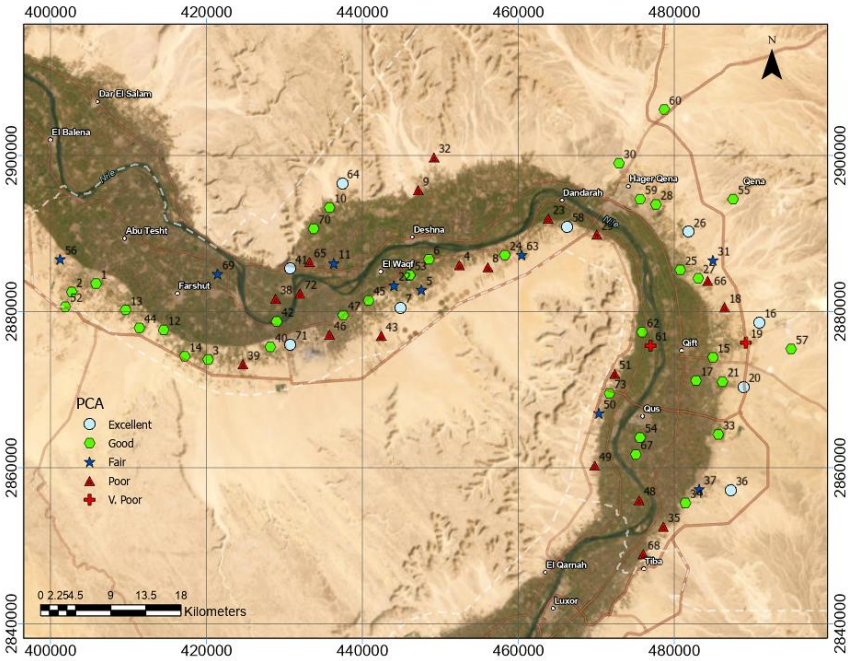


Figure 6. Mapping of GWQI based on PCA

The KPCA analysis employed four kernel functions (linear, polynomial, sigmoid, and radial basis function (RBF) and considered four principal components (PCs). The cumulative percentage of variance explained by the PCs for each kernel function is as follows: linear (87.7%), polynomial (87.8%), sigmoid (87.75%), and RBF (83.8%).

The groundwater quality distribution, as determined by Kernel Principal Component Analysis (KPCA) with a linear kernel (Figure 7), shows a moderate variation across the study area. Of the 73 wells sampled, 43 wells, representing 58.9%, were deemed suitable for irrigation. The suitability categories were further broken down: 6 wells (8.2%) were classified as excellent, 22 (30.1%) as good, and 15 (20.5%) as fair. Conversely, 23 wells (31.5%) were categorized as poor, and 7 (9.6%) were found to be very poor. When compared to the PCA model, the linear KPCA method yielded a more conservative assessment, with a lower percentage of suitable samples (58.9% versus 71.0%). Notably, several wells that were classified as "Good" under PCA are downgraded to "Fair" or "Poor" under this model.

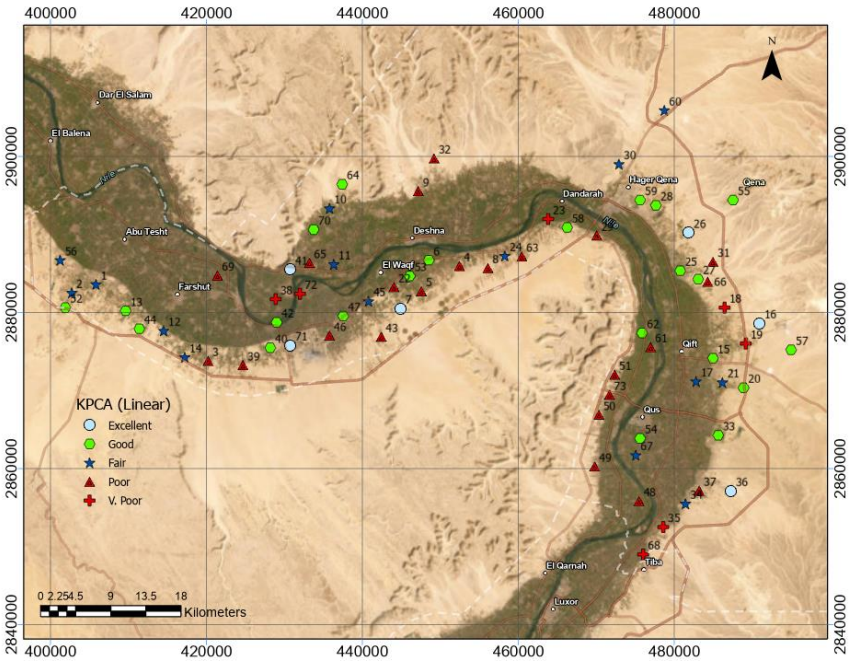


Figure 7. Mapping of GWQI based on KPCA with linear kernel

The polynomial kernel (Figure 8) produced the most restrictive classification among the KPCA models, with only 38 wells (52.1%) classified as suitable. This includes 6 wells (8.2%) classified as excellent, 14 (19.2%) as good, 18 (24.7%) as fair, 29 (39.7%) as poor, and 6 (8.2%) as very poor. A notable number of wells downgraded from good in the PCA and linear results to fair or poor in this kernel.

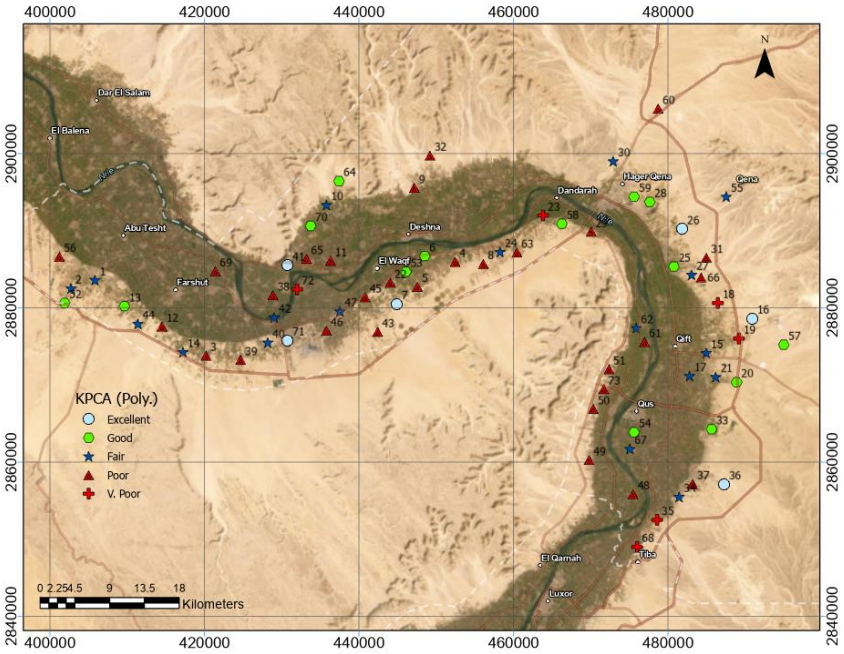


Figure 8. Mapping of GWQI based on KPCA with Polynomial kernel

The KPCA model with a sigmoid kernel produces the highest proportion of suitable wells among all KPCA variants, with 46 samples (63.0%) classified as suitable for irrigation (Figure 9). This includes 29 wells (39.7%) in the "Good" category, 17 (23.3%) as "Fair", 21 (28.8.1%) as "Poor", and 6 (8.2%) as "Very Poor". No wells were found to be excellent. Several wells that were downgraded in the polynomial and linear models are restored to "Good" status under the sigmoid kernel, suggesting a more balanced weighting of parameters.

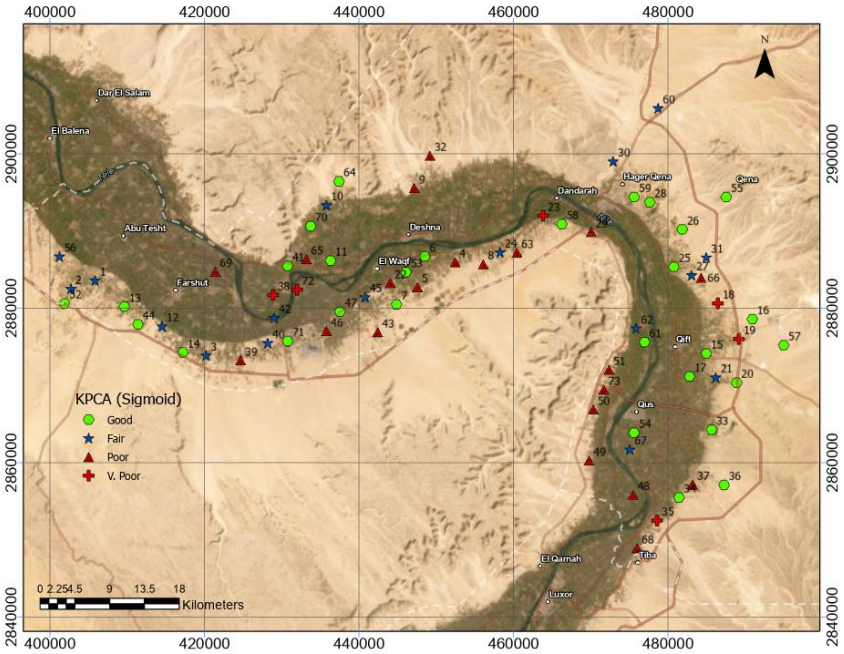


Figure 9. Mapping of GWQI based on KPCA with Sigmoid kernel

The radial basis function (RBF) kernel-based KPCA model (Figure 10) classifies 43 wells (58.9%) as suitable for irrigation, matching the performance of the linear kernel. The distribution includes 6 wells (8.2%) as “Excellent”, 22 (30.1%) as “Good”, 15 (20.5%) as “Fair, 25 (34.2%) as “Poor”, and 5 (6.8%) as “Very Poor”. The RBF kernel, known for its ability to detect local nonlinear patterns, tends to emphasize spatial heterogeneity in water quality, particularly in transitional zones between the floodplain and desert margins.

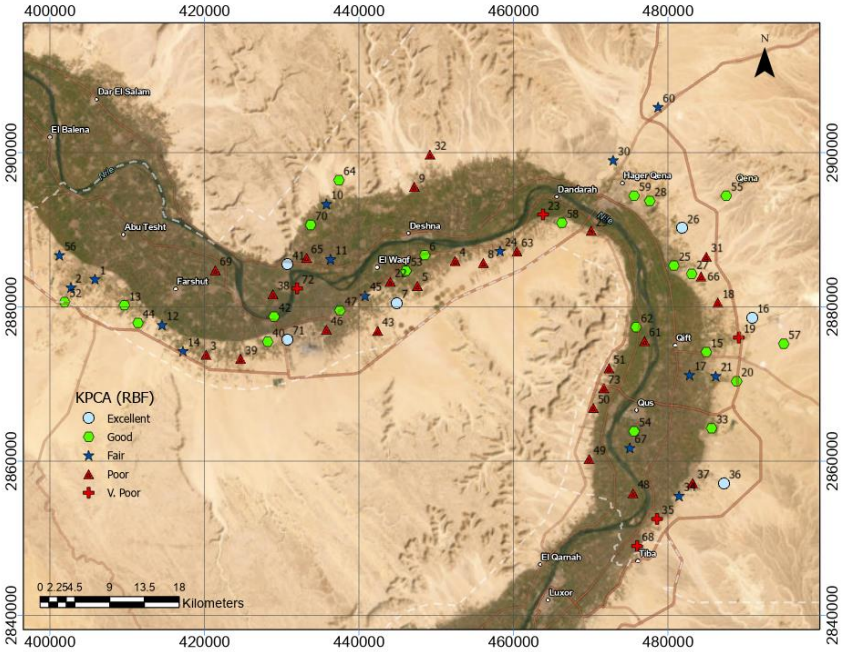


Figure 10. Mapping of GWQI based on KPCA with RBF kernel

Figure 11 presents a comparative bar chart summarizing the number of groundwater samples classified into each GWQI category (excellent, good, fair, poor, and very poor) using PCA and the four KPCA kernels. The results clearly highlight the differences between the linear PCA approach and the nonlinear KPCA methods. PCA classified the largest proportion of wells as suitable for irrigation, with a total of 43 samples (71%) falling within the excellent and good categories. In contrast, the KPCA approaches yielded more conservative classifications.

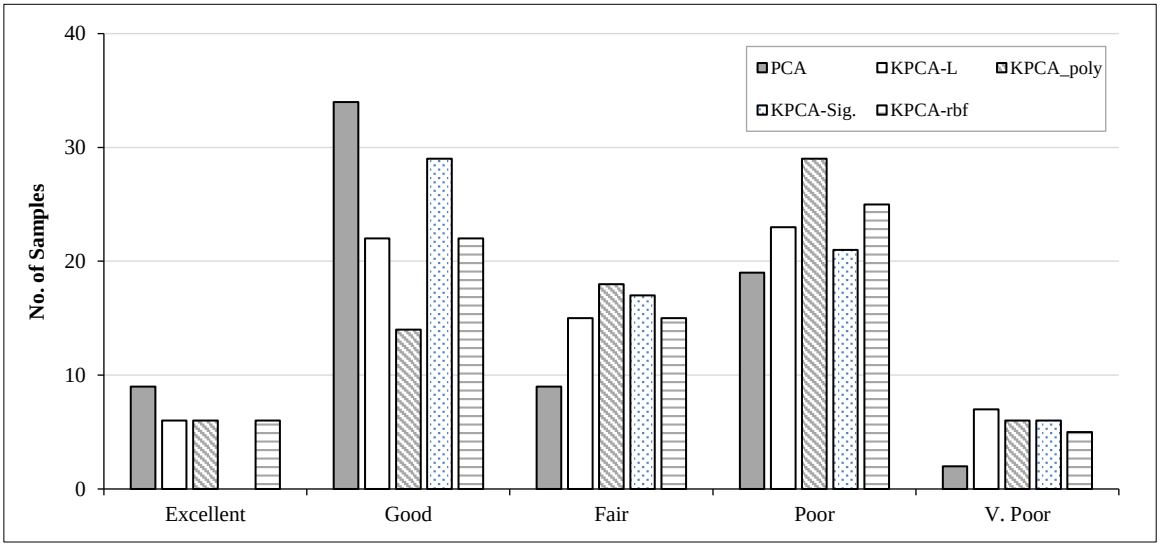


Figure 11. Distribution of GWQI classes using PCA and KPCA methods

The percentage of samples classified as suitable for irrigation using the four KPCA kernel methods (ranging from 52.1% to 63.0%) shows some agreement with the results obtained from the USSL method (53.4%), Na% method (53.4%), and PI method (57.6%). However, PCA yielded a higher percentage of suitable samples (71.0%), which aligns more closely with the Wilcox diagram classification (76.7%).

Figure 12 illustrates the correlation between hydrochemical parameters and GWQI for both PCA and the four KPCA methods. The figure demonstrates a negative correlation between most parameters and GWQI, apart from pH which exhibits a positive correlation. This indicates that higher GWQI values generally correspond to lower concentrations of these parameters. Notably, pH, K^+, and HCO_3^-, which had minimal weights in PCA, also displayed weak correlations with GWQI across all methods. Additionally, PCA showed a weak correlation with Na^+ while the KPCA approaches exhibited a weak correlation with Ca^{2+}. Interestingly, the Sigmoid kernel function within KPCA yielded stronger correlations with all hydrochemical parameters compared to the other KPCA methods.

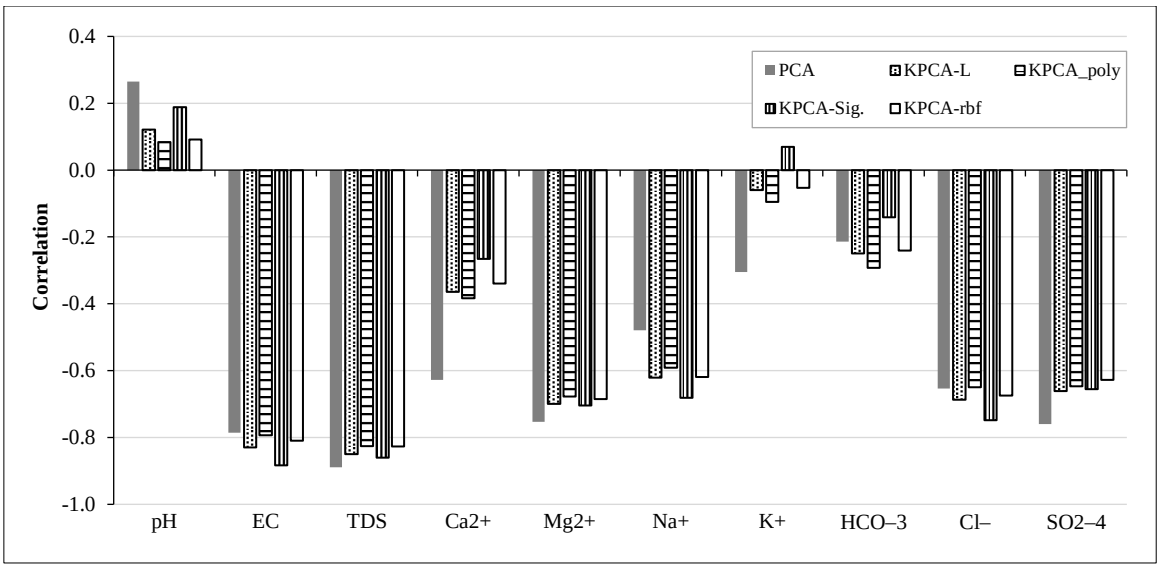


Figure 12. Correlation between the studied methods and the hydrochemical parameters

This study emphasizes the advantages of nonlinear dimensionality reduction techniques, particularly KPCA, for developing WQI. Our findings demonstrate that applying KPCA with a sigmoid kernel achieved a slightly stronger correlation with the final GWQI compared to other methods. This aligns with the growing body of research successfully employing KPCA for water quality assessment, as exemplified by Zhang et al. [17], Yu et al. [15], and Ni et al. [46].

It's worth noting that some studies utilize a single, easily measurable parameter like BOD or COD for WQI estimation, followed by nonlinear dimensionality reduction for further analysis. While this approach offers simplicity, our study takes a more comprehensive approach. We directly apply nonlinear dimensionality reduction to multiple hydrochemical parameters, potentially capturing a richer picture of groundwater quality within the GWQI framework.

5. Conclusion

This study undertook a comparative assessment of Principal Component Analysis (PCA) and Kernel Principal Component Analysis (KPCA) for developing a Groundwater Quality Index (GWQI) within the Qena Governorate of Egypt. Utilizing a dataset of 73 groundwater samples and ten hydrochemical parameters, the research demonstrated that both PCA and KPCA are effective for dimensionality reduction, successfully simplifying complex hydrochemical data while preserving critical information.

The findings indicate that while PCA produced the highest proportion of wells suitable for irrigation (71.0%), closely matching conventional classification methods such as the Wilcox Diagram (76.7%), the KPCA models provided a more conservative and potentially more realistic assessment. The suitability rates for KPCA varied depending on the kernel function: 58.9% for linear, 52.1% for polynomial, 63.0% for sigmoid, and 58.9% for radial basis function (RBF). These values show closer alignment with established salinity hazard indices such as USSL (53.4%), Na% (53.4%), and PI (57.6%), suggesting that KPCA may better reflect the constraints imposed by ionic interactions and salinity-related risks in irrigation contexts. The application of nonlinear kernels in KPCA, particularly the sigmoid kernel, proved more adept at capturing complex, nonlinear interactions among hydrochemical variables—such as ion exchange, saturation effects, and overconcentration—that conventional linear PCA may overlook. Notably, Figure 12 reveals that the sigmoid kernel achieved stronger correlations between the derived principal components and original hydrochemical parameters compared to other KPCA variants and PCA, indicating superior feature representation and interpretability. This enhanced performance suggests that the sigmoid kernel effectively models bounded, saturation-like geochemical behaviors, making it particularly suitable for groundwater quality assessment in arid environments.

Despite these contributions, the study acknowledges several limitations. The reliability of both PCA and KPCA is contingent upon data preprocessing methods, such as scaling and normalization, which can directly influence component loadings and final index values. For KPCA, the selection of the kernel type and its corresponding parameters (e.g., degree and coefficient for the polynomial kernel, γ and σ for RBF, or α and c for sigmoid) significantly impacts the results; however, this study did not perform systematic hyperparameter optimization or cross-validation, which could affect the generalizability of the models. Additionally, the dataset represents a single temporal snapshot from 2013–2014, thereby precluding an analysis of seasonal fluctuations, long-term trends, or the impact of climate variability and anthropogenic activities over time. Furthermore, the GWQI computation relies on eigenvalue-based weighting, which assumes that higher-variance components are more significant—a premise that may not always align with hydrochemical importance (e.g., trace contaminants with low variance but high toxicity).

Future research should focus on expanding the dataset to include multi-seasonal and multi-year sampling to facilitate the development of spatiotemporal GWQI models that more accurately reflect groundwater dynamics under changing environmental and human pressures. Methodological improvements should include systematic hyperparameter tuning for KPCA using techniques such as grid search or Bayesian optimization, as well as validation through bootstrapping or k-fold cross-validation to enhance model robustness. Furthermore, integrating KPCA with advanced machine learning models such as KPCA-LSSVM for classification or KPCA-RNN for time-series prediction could significantly improve predictive capabilities for early warning systems in water quality monitoring. Finally, comparative studies with other nonlinear dimensionality reduction techniques such as t-SNE or ISOMAP could provide further insights into optimal methods for hydrochemical data analysis.

In conclusion, this study highlights the advantages of nonlinear approaches like KPCA, especially with the sigmoid kernel, in improving the accuracy and physical interpretability of groundwater quality assessments. By capturing hidden nonlinear patterns in hydrochemical data, KPCA offers a promising advancement over traditional PCA-based indices, supporting more informed and sustainable groundwater management decisions in arid and semi-arid regions like Upper Egypt.

6. Declarations

6.1. Author Contributions

Conceptualization, S.A.; methodology, M.K. and S.A.; software, S.A. and H.A.; validation, M.K., H.A., and M.E.; formal analysis, M.E.; investigation, M.K.; resources, H.A.; data curation, S.A. and M.E.; writing—original draft preparation, M.K.; writing—review and editing, S.A.; visualization, H.A.; supervision, S.A.; project administration, M.K.; funding acquisition, S.A. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding and Acknowledgements

This Project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under Grant No. (IFPIP: 1289-829-1443). The authors also acknowledge, with thanks, DSR, for technical and financial support.

6.4. Conflicts of Interest

The authors declare no conflict of interest.

7. References

[1] UNESCO. (2022). Groundwater, Making the invisible visible. The United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, France.

[2] R. K. Horton. (1965). An Index Number System for Rating Water Quality. Journal of the Water Pollution Control Federation, 37(3), 300–306.

[3] Lumb, A., Sharma, T. C., Bibeault, J.-F., & Klawunn, P. (2011). A Comparative Study of USA and Canadian Water Quality Index Models. Water Quality, Exposure and Health, 3(3–4), 203–216. doi:10.1007/s12403-011-0056-5.

[4] Hoseinzadeh, E., Khorsandi, H., Wei, C., & Alipour, M. (2015). Evaluation of Aydughmush River water quality using the National Sanitation Foundation Water Quality Index (NSFWQI), River Pollution Index (RPI), and Forestry Water Quality Index (FWQI). Desalination and Water Treatment, 54(11), 2994–3002. doi:10.1080/19443994.2014.913206.

[5] Chidiac, S., El Najjar, P., Ouaini, N., El Rayess, Y., & El Azzi, D. (2023). A comprehensive review of water quality indices (WQIs): history, models, attempts and perspectives. Reviews in Environmental Science and Biotechnology, 22(2), 349–395. doi:10.1007/s11157-023-09650-7.

[6] Patel, P. S., Pandya, D. M., & Shah, M. (2023). A systematic and comparative study of Water Quality Index (WQI) for groundwater quality analysis and assessment. Environmental Science and Pollution Research, 30(19), 54303–54323. doi:10.1007/s11356-023-25936-3.

[7] Mahanty, B., Lhamo, P., & Sahoo, N. K. (2023). Inconsistency of PCA-based water quality index – Does it reflect the quality? Science of the Total Environment, 866, 161353. doi:10.1016/j.scitotenv.2022.161353.

[8] Subba Rao, N., Das, R., Sahoo, H. K., & Gugulothu, S. (2024). Hydrochemical characterization and water quality perspectives for groundwater management for urban development. Groundwater for Sustainable Development, 24, 101071. doi:10.1016/j.gsd.2023.101071.

[9] Ali, S., Verma, S., Agarwal, M. B., Islam, R., Mehrotra, M., Deolia, R. K., Kumar, J., Singh, S., Mohammadi, A. A., Raj, D., Gupta, M. K., Dang, P., & 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. doi:10.1038/s41598-024-56056-8.

[10] Lehr, C., Dannowski, R., Kalettka, T., Merz, C., Schröder, B., Steidl, J., & Lischeid, G. (2018). Detecting dominant changes in irregularly sampled multivariate water quality data sets. Hydrology and Earth System Sciences, 22(8), 4401–4424. doi:10.5194/hess-22-4401-2018.

[11] Liu, H., Yang, J., Ye, M., James, S. C., Tang, Z., Dong, J., & Xing, T. (2021). Using t-distributed Stochastic Neighbor Embedding (t-SNE) for cluster analysis and spatial zone delineation of groundwater geochemistry data. Journal of Hydrology, 597, 126146. doi:10.1016/j.jhydrol.2021.126146.

[12] Liu, Y., Zheng, N., Yang, S., Liu, F., Liu, M., & Chen, Y. (2025). Identification of water pollution sources in the Daluxi River using kernel principal component analysis and gradient boosting decision tree. Environmental Earth Sciences, 84(9), 236. doi:10.1007/s12665-025-12241-0.

[13] Zhao, Y., & Chen, M. (2025). Prediction of river dissolved oxygen (DO) based on multi-source data and various machine learning coupling models. PLOS ONE, 20(3), e0319256. doi:10.1371/journal.pone.0319256.

[14] Liu, W., Wang, J., Li, Z., & Lu, Q. (2024). ISSA optimized spatiotemporal prediction model of dissolved oxygen for marine ranching integrating DAM and Bi-GRU. Frontiers in Marine Science, 11, 1473551. doi:10.3389/fmars.2024.1473551.

[15] Yu, T., Yang, S., Bai, Y., Gao, X., & Li, C. (2018). Inlet water quality forecasting of wastewater treatment based on kernel principal component analysis and an extreme learning machine. Water (Switzerland), 10(7), 873. doi:10.3390/w10070873.

[16] Abba, S. I., Pham, Q. B., Usman, A. G., Linh, N. T. T., Aliyu, D. S., Nguyen, Q., & Bach, Q. V. (2020). Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. Journal of Water Process Engineering, 33, 101081. doi:10.1016/j.jwpe.2019.101081.

[17] Zhang, Y. F., Fitch, P., & Thorburn, P. J. (2020). Predicting the trend of dissolved oxygen based on the kPCA-RNN model. Water (Switzerland), 12(2), 585. doi:10.3390/w12020585.

[18] Jibrin, A. M., Al-Suwaiyan, M., Yaseen, Z. M., & Abba, S. I. (2025). New perspective on density-based spatial clustering of applications with noise for groundwater assessment. Journal of Hydrology, 661, 133566. doi:10.1016/j.jhydrol.2025.133566.

[19] El-Rawy, M., Ismail, E., & Abdalla, O. (2019). Assessment of groundwater quality using GIS, hydrogeochemistry, and factor statistical analysis in Qena governorate, Egypt. Desalination and Water Treatment, 162, 14–29. doi:10.5004/dwt.2019.24423.

[20] Said, R. (2017). The Geology of Egypt. Routledge, Milton Park, United Kingdom. doi:10.1201/9780203736678.

[21] El-Belsasy, M. I. (1994). Quaternary geology of some selected drainage basins in upper Egypt (Qena-Idfu area). Ph.D. Thesis, Cairo University, Faculty of Science, Giza, Egypt.

[22] Beshr, A. M., Kamel Mohamed, A., ElGalladi, A., Gaber, A., & El-Baz, F. (2021). Structural characteristics of the Qena Bend of the Egyptian Nile River, using remote-sensing and geophysics. The Egyptian Journal of Remote Sensing and Space Science, 24(3), 999–1011. doi:10.1016/j.ejrs.2021.11.005.

[23] Ayers, R. S., & Westcot, D. W. (1985). Water quality for agriculture. Food and agriculture organization of the United Nations, Rome, Italy.

[24] Yan, S., Zhang, T., Zhang, B., Zhang, T., Cheng, Y., Wang, C., Luo, M., Feng, H., & Siddique, K. H. M. (2023). The higher relative concentration of K+ to Na+ in saline water improves soil hydraulic conductivity, salt-leaching efficiency and structural stability. Soil, 9(1), 339–349. doi:10.5194/soil-9-339-2023.

[25] Yu, J., Shi, J. G., Ma, X., Dang, P. F., Yan, Y. L., Mamedov, A. I., Shainberg, I., & Levy, G. J. (2017). Superabsorbent Polymer Properties and Concentration Effects on Water Retention under Drying Conditions. Soil Science Society of America Journal, 81(4), 889–901. doi:10.2136/sssaj2016.07.0231.

[26] Missaoui, R., Ncibi, K., Abdelkarim, B., Bouajila, A., Choura, A., Hamdi, M., & Hamed, Y. (2023). Assessment of hydrogeochemical characteristics of groundwater: link of AHP and PCA methods using a GIS approach in a semi-arid region, Central Tunisia. Euro-Mediterranean Journal for Environmental Integration, 8(1), 99–114. doi:10.1007/s41207-023-00345-7.

[27] Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. Educational and Psychological Measurement, 20(1), 141–151. doi:10.1177/001316446002000116.

[28] Satour, N., Benyacoub, B., El Mahrad, B., & Kacimi, I. (2021). KPCA over PCA to assess urban resilience to floods. E3S Web of Conferences, 314, 03005. doi:10.1051/e3sconf/202131403005.

[29] Wang, Z., van der Laan, T., & Usman, M. (2025). Self- Adaptive Quantum Kernel Principal Component Analysis for Compact Readout of Chemiresistive Sensor Arrays. Advanced Science, 12(15), 202411573. doi:10.1002/advs.202411573.

[30] Lajmi, F., Mhamdi, L., Abdelbaki, W., Dhouibi, H., & Younes, K. (2023). Investigating Machine Learning and Control Theory Approaches for Process Fault Detection: A Comparative Study of KPCA and the Observer-Based Method. Sensors, 23(15), 6899. doi:10.3390/s23156899.

[31] Jalili, A., Saleki, Z., Luo, Y. A., Pan, F., Chen, A. X., & Draayer, J. P. (2025). Performance of various kernel functions for mass prediction with support vector machine. European Physical Journal A, 61(6), 143. doi:10.1140/epja/s10050-025-01610-9.

[32] Saha, S., Das, M., Mondal, B. S., Sarkar, S., & Maiti, J. (2021). DiPSVM: A Polynomial Kernel-free Support Vector Machine. 2021 International Conference on Data Analytics for Business and Industry (ICDABI), 448–452. doi:10.1109/icdabi53623.2021.9655976.

[33] Mustafa, B. (2024). New developments and applications of radial basis functions in interpolation, approximation and data science. Ph.D. Thesis, University of Granada Faculty of Science, Granada, Spain.

[34] 34-Jiang, H. J., You, Z. H., & Huang, Y. A. (2019). Predicting drug-disease associations via sigmoid kernel-based convolutional neural networks. Journal of Translational Medicine, 17(1), 382. doi:10.1186/s12967-019-2127-5.

[35] Richards, L. A. (1954). Diagnosis and Improvement of Saline and Alkali Soils. Soil Science, 78(2), 154. doi:10.1097/00010694-195408000-00012.

[36] Todd, D. and Mays, L. (2005) Groundwater Hydrology (3rd Ed.). John Wiley and Sons, Hoboken, New Jersey.

[37] Wilcox, L.V. (1955) Classification and Use of Irrigation Water. US Department of Agriculture, Circular 969, Washington, United States.

[38] Szabolcs, I. (1964). The influence of irrigation water of high sodium carbonate content on soils. Agrokémia és talajtan, 13(SUP), 237-246.

[39] Doneen, L. D. (1962). The influence of crop and soil on percolating water. Proc. 1961 Biennial conference on Groundwater recharge.

[40] Dregne, H. E. (1952). Alkali Soils, their Formation, Properties, and Reclamation. Agronomy Journal, 44(6), 339–339. doi:10.2134/agronj1952.00021962004400060019x.

[41] Teng, W. C., Fong, K. L., Shenkar, D., Wilson, J. A., & Foo, D. C. Y. (2016). Piper diagram – A novel visualisation tool for process design. Chemical Engineering Research and Design, 112, 132–145. doi:10.1016/j.cherd.2016.06.002.

[42] Gibbs, R. J. (1970). Mechanisms controlling world water chemistry. Science, 170(3962), 1088–1090. doi:10.1126/science.170.3962.1088.

[43] Wilcox, L.V. (1948) The Quality of Water for Agricultural Use. US Department of Agriculture, Technical Bulletin, Washington, United States.

[44] Bisong, E. (2019). Introduction to Scikit-learn. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA, United States. doi:10.1007/978-1-4842-4470-8_18.

[45] Dawoud, M. A., & Ismail, S. S. (2013). Saturated and unsaturated River Nile/groundwater aquifer interaction systems in the Nile Valley, Egypt. Arabian Journal of Geosciences, 6(6), 2119–2130. doi:10.1007/s12517-011-0483-4.

[46] Ni, J., Ma, H., & Ren, L. (2012). A time-series forecasting approach based on KPCA-LSSVM for lake water pollution. 9th International Conference on Fuzzy Systems and Knowledge Discovery, 1044–1048. doi:10.1109/fskd.2012.6234207.