PLANT SCIENCE TODAY
Vol 12(sp4): 1-11
https://doi.org/10.14719/pst.11964

eISSN 2348-1900

RESEARCH ARTICLE

Integrated assessment of soil quality using PCA-derived indices and spatial zonation in a micro-watershed

Aishwarya Golshetti[1*], Yogananda S B[1], Ananthakumar M A[2], M N Thimmegowda[3], P Thimmegowda[4], Praveen Puttaswamaiah[5] & Shivashankar K[6]

[1]Department of Agronomy, College of Agriculture, V C Farm, Mandya, University of Agricultural Science, Bangalore 571 405, Karnataka, India

[2]Department of Soil Science and Agricultural Chemistry, Water Technology Centre Zonal Agricultural Research Station V C Farm,Mandya, University of Agricultural Science, Bangalore 571 405, Karnataka, India

[3]Department of Agricultural Meteorology, GKVK, University of Agricultural Science, Bangalore 560 065, Karnataka, India

[4]Zonal Agricultural Research Station, V C Farm, Mandya, University of Agricultural Science, Bangalore 571 405, Karnataka, India

[5]Department of Agricultural Engineering, College of Agriculture, V C Farm, Mandya, University of Agricultural Science, Bangalore 571 405, Karnataka, India

[6]Carbon Department, Growindigo Pvt. Ltd., New Delhi 110 020 , India

*Correspondence email - bharatiashwa1976@gmail.com

Received: 24 September 2025; Accepted: 30 October 2025; Available online: Version 1.0: 11 November 2025

Cite this article: Aishwarya G, Yogananda SB, Ananthakumar MA, Thimmegowda MN, Thimmegowda P, Praveen P, Shivashankar K. Integrated assessment of soil quality using PCA-derived indices and spatial zonation in a micro-watershed. Plant Science Today. 2025; 12(sp4): 1-11. https://doi.org/10.14719/pst.11964

Abstract

Soil quality assessment is crucial for sustainable land management, integrating physical, chemical and biological soil properties into a single measurable index. This study evaluated soil quality in a micro-watershed using Principal Component Analysis (PCA) and cluster analysis tools to delineate management zones. Sixty three soil samples were collected horizon-wise and analyzed for texture, pH, electrical conductivity, organic carbon, exchangeable cations, Cation Exchange Capacity (CEC) and base saturation (BS %). PCA revealed four key components explaining 90.5 % of total data variability. A Minimum Data Set (MDS) was derived based on strong factor loadings, including sand, silt, clay, base saturation percentage and CEC. The Soil Quality Index (SQI) was computed using standardized z-scores, weighted by each component's variance contribution, resulting in a comprehensive SQI (CSQI) ranging from -0.48 to 1.18. Approximately 20 % of soils were classified as very good, 40 % as good and the remaining 40 % as fair to very poor. Cluster analysis (k-means and hierarchical) grouped the soils into three fertility zones, with fine-textured, nutrient-rich soils corresponding to the highest CSQI values. Spatial analysis and thematic mapping highlighted that soils with higher clay, CEC and BS % were concentrated in depositional zones and exhibited superior quality, whereas coarse-textured upland soils were nutrient-poor and vulnerable to degradation. These findings underscore the need for site-specific interventions in low-SQI areas, including organic matter addition, nutrient balancing, erosion control and adoption of conservation agriculture practices. In contrast, high-quality zones can maintain current management to sustain productivity. Overall, the integration of PCA-based SQI with GIS spatial modelling proved effective for identifying soil heterogeneity and prioritizing area-specific management strategies, ensuring improved soil resilience and sustainable productivity in the Bankanahalli micro -watershed, characterized predominantly by Alfisols.

Keywords: cluster analysis; comprehensive SQI; minimum data set; principal component analysis; soil quality index

Introduction

Healthy soil is the cornerstone of agricultural sustainability and food security, providing essential ecosystem services such as water regulation, nutrient cycling, plant growth support and maintenance of biological diversity (1). However, soils in many regions are experiencing increased stress because of intensive cultivation, cropping without rotation, nutrient mining and inappropriate management practices which together drive widespread soil degradation and reduce both current and future crop productivity (2, 3). This degradation poses significant threats to the resilience and long-term viability of farming systems, making a holistic understanding of soil health a prerequisite for effective restoration and management (1, 2).

Soil quality is best described as the capacity of soil to perform essential ecological functions such as sustaining productivity, regulating hydrological and biochemical cycles and maintaining environmental quality (1, 3). Because soils have multidimensional physical, chemical and biological attributes, no single property can represent overall soil quality. As such, research has shifted toward developing composite indices that integrate several indicators into a single numerical value (1, 2).

A widely used approach is the Soil Quality Index (SQI), which summarizes multiple soil indicators in a single score indicating functionality (3). The reliability of the SQI depends on precise selection and weighting of indicators, a process often achieved through Principal Component Analysis (PCA) a multivariate statistical tool that reduces data redundancy and identifies a MDS of the most informative soil properties (1, 2). Once the MDS is identified, indicator values are standardized and combined, with weights proportional to their contributions to total variability. Methodological reviews continue to identify conceptual and practical challenges, including indicator selection, data normalization and scale of assessment (4, 5). To further integrate multiple principal components, a Comprehensive Soil Quality Index (CSQI) may be computed by weighting individual SQI scores with the variance explained by each principal component, resulting in a robust measure of overall soil health (2).

Understanding spatial variability in soil quality is equally crucial for site-specific and precision management. Geostatistical tools and Geographic Information Systems (GIS) allow point data interpolation and the creation of thematic maps, displaying soil quality patterns for targeted interventions (1). Cluster analysis methods such as k-means and hierarchical clustering complement SQI by grouping soils with similar properties, enabling delineation of management zones sharing fertility constraints and production potential (2).

While individual studies often focus on constrained aspects of soil fertility or single-parameter limitations, relatively few have integrated PCA-derived SQI, CSQI, cluster analysis and GIS mapping in a unified framework at the micro watershed scale (2, 3). There remains a need for such holistic assessments that combine statistical rigor and spatial visualization, supporting actionable recommendations for farmers and policymakers.

Unlike many studies that focus on isolated soil parameters or limited geographic scales, this study uniquely integrates multivariate statistical techniques with spatial analysis to delineate precise soil fertility zones and management recommendations. Some of the recent studies also emphasized the role of slope, land use and management in driving SQI variation, with site-specific indicators improving predictive accuracy for both forestry and agricultural landscapes (6, 7).

The present study addresses this need by conducting a detailed multivariate and spatial analysis of the quality of soil in a representative micro watershed. The main objectives were: (i) to characterize the variability of key physico-chemical soil properties; (ii) to identify an MDS through PCA and compute SQI and CSQI; (iii) to classify soils into distinct fertility clusters using k-means and hierarchical clustering; and (iv) to produce a thematic map of soil quality to visualize spatial heterogeneity and support site-specific management. By integrating these approaches, the study delivers a robust framework for monitoring soil health and prioritizing interventions aimed at improving productivity and sustainability.

Materials and Methods

Study area and data collection

The study was carried out in a Bankanahalli micro watershed in Dudda sub watershed, where data were collected from 9 soil profiles representing various land uses and physiographic conditions, the dominant soil order was Alfisols with sandy clay loam located between north latitude 12°35‘ 6.863” to 12° 36' 56.68" N and east longitude 76° 45' 13.14" to 76° 46' 23.722" E covering an area of 489 ha (8, 9). A total of 63 soil samples were collected horizon-wise to capture vertical variability in physical and chemical properties. Profile data included texture fractions (sand, silt, clay) international pipette method, pH (potentiometric), electrical conductivity (conductometric) (EC), organic carbon (wet oxidation) (OC), exchangeable cations (Ca and Mg-Versanate titration, Na and K- flame photometry), exchange capacity of cations (CEC) by neutral normal ammonium acetate or sodium acetate method and base saturation (BS %).

Descriptive statistics and normality testing

Initial data exploration was performed using descriptive statistics (mean, standard deviation, minimum, maximum, skewness and kurtosis) to understand the variability of soil parameters. The Shapiro-Wilk test was employed to check normality, as it is sensitive to deviations in both skewness and kurtosis.

Multivariate analysis

To reduce dimensionality and identify key soil quality critical indicators, PCA was conducted (1, 10). The Kaiser-Meyer-Olkin (KMO) measure (0.75) and Bartlett's Test of Sphericity (χ^2 = 1641.2, $p < 0.001$) confirmed the adequacy of the data for PCA. Principal components with eigenvalues > 1 and contributing ≥ 5 % variance were retained, resulting in four major PCs explaining 90.49 % of total variance.

Soil Fertility Index (SFI) calculation

For each selected soil indicator, standardized z-scores were calculated (10) as:

$$Z = \frac{x - \bar{x}}{\sigma} \quad \text{(Eqn.1)}$$

Where

z, x, $\bar{x}$ and σ implies the standardized values, the value of the soil parameter, the average of the parameter and the standard deviation of the soil parameters, respectively

The Soil Quality Index (SQI) was then computed as:

$$SQI = \sum_{i=1}^{N} Wi * Si \quad \text{(Eqn.2)}$$

Where

Wi is the pc weight calculated form the communality, Si is the standardized values from eqn.1 (2, 10),

Finally, a Comprehensive Soil Quality Index (CSQI) was obtained by combining the SQI with the percentage variability of each PC:

$$CSQI = \sum_{i=1}^{N} \text{Variability of each PC} * \text{SQI PC} \quad \text{(Eqn.3)}$$

The soil quality is classified as very good (0.8-1), good (0.6-0.79), fair (0.35-0.39), bad (0.2-0.34) and very bad (<0-0.19) (3, 10).

Cluster analysis

To group soils with similar characteristics, both k-means clustering and hierarchical clustering were applied. The number of clusters was optimized using a scree plot and dendrogram analysis, leading to the identification of three distinct soil clusters based on silhouette method (2, 11).

Statistical analysis

Descriptive statistics were analyzed using the library “pastecs” in the R programming version 4.2.3 (12), indicating the descriptive statistics and normality distribution of the data of 13 parameters, which were taken from the existing Land resource inventory (LRI) data collected via profile sheets.

N is the number of samples subjected to analysis, the minimum and maximum value, mean and its standard deviation for a particular parameter based on the data, including 3 physical parameters and 10 chemical parameters.

Results and Discussion

Descriptive statistics and normality

A total of nine profiles were studies in the micro watershed depending on the physiographic features. The soil profile data exhibited considerable variability depth-wise across the micro watershed. Sand content ranged from 35.49 to 79.98 % (mean: 54.08 %), whereas silt ranged from 4.48 to 17.22 % (mean: 10.12 %) and clay from 14.45 to 47.83 % (mean: 35.11 %). Soils were generally neutral to moderately alkaline, with pH values varying between 5.41 and 10.43 (mean: 7.9). Electrical conductivity (EC) values indicated predominantly non-saline soils (mean: 0.28 dS/m), though some sites showed localized salinity buildup (EC up to 1.45 dS/m). Organic carbon levels were low to moderate (0.12-0.84 %, mean: 0.34 %), reflecting limited organic matter inputs. Exchangeable cations displayed wide ranges: calcium varied from 2.0-16.0 cmol(+)/kg, magnesium from 1.0-6.5 cmol(+)/kg and sodium from 0.07-3.38 cmol(+)/kg. Cation exchange capacity (CEC) ranged from 6.02-26.01 cmol(+)/kg (mean: 14.94 cmol(+)/kg) (Table 1). Table 2 shows the shape and symmetry of the data distribution for each soil parameter, which can help to understand the overall data structure. Table 3 summarizes the results of Shapiro-Wilk test.

Shapiro-Wilk Test Statistic (W Value)

The test statistics form the Shapiro-Wilk determination (table 3), for most of the soil parameters it was closer to one indicating the normal distribution in the dataset (10).

P-Value

After testing for normality, most parameters followed normal distribution ($p < 0.05$) except sand pH and base saturation (%). Then the data were subjected to Kaiser Meyer Olkin measures of sampling adequacy and Bartletts's sphericity test (10).

KMO value closer to 1 indicate (Table 4), the dataset is suitable for factor analysis, the determined value of KMO was 0.75, while values below 0.5 indicate that factor analysis may not be appropriate. The Chi-Squared test evaluates whether the distribution of sample categorical data matches an expected distribution. A p-value less than 0.05 typically indicates a statistically significant association (13, 14).

Table 1. Soil site characteristics of entire micro-watershed based on the soil profile study

Parameters	N	Minimum	Maximum	Mean	Std. Deviation	Kurtosis	Skewness	Shapiro wilk	
								W	P
Sand	63	35.49	79.98	54.08	10.35	-0.40	0.27	0.9731	0.1833
Silt	63	4.48	17.22	10.12	3.82	-1.18	0.34	0.9292	0.0014
Clay	63	14.45	47.83	35.11	9.09	-0.54	-0.68	0.9230	0.0007
pH	63	5.41	10.43	7.9	1.26	-0.68	0.37	0.9608	0.0426
EC (dS/m)	63	0.05	1.45	0.28	0.26	7.77	2.47	0.7470	0.0000
OC (%)	63	0.12	0.84	0.34	0.15	0.61	0.72	0.9434	0.0060
Ca	63	2	16	7.7	3.51	-0.11	0.77	0.9302	0.0015
Mg	63	1	6.5	3.52	1.41	-0.48	0.62	0.9205	0.0006
Na	63	0.07	3.38	0.9	0.91	0.47	1.25	0.7815	0.0000
K	63	0.03	1.2	0.19	0.22	12.02	3.34	0.5330	0.0000
Total Cations (Meq/100g soil)	63	3.32	23.67	12.46	5.35	-0.63	0.59	0.9273	0.0011
CEC	63	6.02	26.01	14.94	5.12	-0.38	0.66	0.9319	0.0018
BS (%)	63	55.18	97.27	80.9	9.72	-0.17	-0.67	0.9493	0.0114

Table 2. Description for kurtosis and skewness for normality distribution of the LRI data

Parameter	Kurtosis	Interpretation	Skewness	Interpretation
Sand	-0.40		0.27	Slightly right-skewed
Silt	-1.18	Light-tailed distribution, fewer extreme values	0.34	Right-skewed
Clay	-0.54		-0.68	Left-skewed
pH	-0.68		0.37	Slightly right-skewed
EC (dS/m)	7.77	Heavy-tailed distribution, more extreme values	2.47	Strongly right-skewed
OC (%)	0.61		0.72	
Ca	0.77	Slightly heavy-tailed distribution	0.93	Right-skewed
Mg	0.62		0.92	
Na	0.47		1.25	Strongly right-skewed
K	2.02	Heavy-tailed distribution, more extreme values	3.34	
Total Cations	-0.63		0.59	Right-skewed
CEC	-0.38	Light-tailed distribution, fewer extreme values	0.66	
BS (%)	-0.17		-0.67	Left-skewed

Table 3. Description for Shapiro-Wilk test for normality distribution of the LRI data

Parameter	Shapiro-Wilk Test	Interpretation
Sand	0.9731	Approximate normality with data close to normal.
Silt	0.9292	Some deviation from normality.
Clay	0.9230	Notable deviation from normality.
pH	0.9608	Data is reasonably close to normality.
EC (dS/m)	0.7470	Significant deviation from normality.
OC (%)	0.9434	Data reasonably close to normality.
Ca	0.9302	Some deviation from normality.
Mg	0.9205	Notable deviation from normality.
Na	0.7815	Significant deviation from normality.
K	0.5330	Substantial deviation from normality.
Total Cations	0.9273	Some deviation from normality.
CEC	0.9319	
BS (%)	0.9493	Data reasonably close to normality.

Table 4. Kaiser Meyer Olkin measures of sampling adequacy and Barletts's sphericity test

Kaiser meyer olkin measures of sampling	KMO	0.75
Barlett's sphericity test	Chi-square (Observed value)	1641.2
	Chi-square (Critical value)	3.074
	Degrees of freedom (df)	12
	p-value	$< 2.2e^{-16}$
	alpha	0.05

Principal Component Analysis (PCA)

The PCA extracted four principal components (PCs) with eigenvalues of 7.46, 2.24, 1.13 and 0.94, explaining 57.37 %, 17.26 %, 8.66 % and 7.20 % of the total variance, highlighting more the 5 % to the variation hence four PCs were retained, respectively (Table 5). Together, these components accounted for 90.49 % of the overall variability in soil properties, which exceeds the commonly accepted 80 % threshold recommended for environmental and soil quality studies (15, 16). Hence, all four PCs were retained for further interpretation and selection of the MDS, since each component contributed more than 5 % of the total variation and thus carried meaningful information about soil quality.

PC1 was the dominant component, explaining 57.37 % of the total variance and was characterized by high positive loadings for base saturation (BS %, 0.500) and clay (0.477), along with a strong negative loading for sand (-0.573). Base saturation and clay were retained in PC1 as they met the criterion of being within 90 % of the highest loading and exhibited strong correlations (absolute loading ≥ 0.45) (Fig. 1). This component represents the texture-fertility gradient, where higher PC1 scores indicate fine-textured soils with higher nutrient-holding capacity, while negative scores represent coarser soils with lower fertility potential (17).

PC2 contributed 17.26 % of the variance and was strongly associated with clay (0.597) and negatively with BS % (-0.626). This component mainly represents the nutrient availability and fine-particle interactions, making Base saturation and clay essential indicators for soil quality evaluation (15).

PC3 accounted for 8.66 % of the total variance and was dominated by cation exchange capacity (CEC, 0.513) and base saturation (-0.562). These variables capture the nutrient reservoir and cation exchange potential of soils, which are key chemical fertility attributes influencing crop productivity (16).

PC4, though explaining a lower proportion of variance (7.20 %), still contributed valuable information and was therefore retained. It was strongly associated with silt (0.657), which was selected as a representative indicator to capture the residual variability linked to coarse-textured soils and their influence on water movement, infiltration and nutrient leaching potential (17).

Based on the selection criterion of factor loadings ≥ 0.6 (or within 90 % of the highest loading per PC), the final MDS consisted

Table 5. Summarization of PCA

Parameters	PC1		PC2		PC3		PC4	
Eigenvalue	7.46		2.24		1.13		0.94	
Variability (%)	57.37		17.26		8.66		7.20	
Cumulative %	57.37		74.63		83.29		90.49	
Factor loadings					**Component score coefficient matrix**			
Sand	**-0.573**	-0.290	-0.096	0.506	0.109	0.042	-0.036	-0.030
Silt	0.111	-0.317	0.264	**-0.657**	0.109	0.042	-0.036	-0.030
Clay	0.477	**0.597**	-0.231	0.205	-0.246	-0.504	0.237	0.360
Ph	0.047	-0.095	-0.036	-0.152	-0.210	-0.059	0.065	0.104
EC (dS/m)	0.005	-0.020	-0.025	-0.028	-0.092	0.046	0.062	0.225
OC (%)	-0.002	0.004	0.017	0.001	0.143	1.460	-0.102	-0.529
Ca	0.176	-0.079	0.318	0.216	0.043	0.355	-0.244	-0.198
Mg	0.070	-0.039	0.125	0.061	0.406	0.117	-0.116	-0.179
Na	0.032	-0.048	-0.038	-0.044	0.335	0.099	-0.109	-0.172
K	0.003	-0.004	0.009	0.022	-0.058	-0.268	-0.200	-0.159
Total Cations (Meq/100g soil)	0.278	-0.182	0.409	0.281	-0.224	-0.145	1.233	0.320
CEC	0.256	-0.117	**0.513**	0.303	0.355	0.086	-0.104	-0.155
BS (%)	0.500	**-0.626**	**-0.562**	0.151	0.398	0.108	-0.114	-0.172

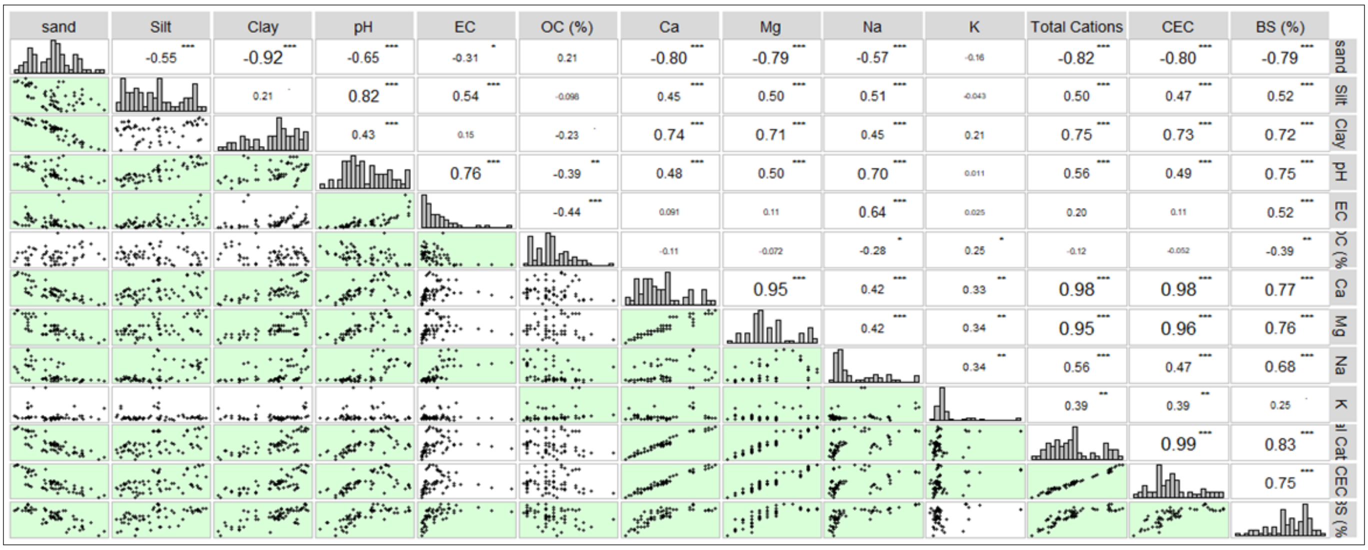

Fig. 1. Correlation of principal components with original variable.

of sand, silt, clay, base saturation (BS %) and cation exchange capacity (CEC). Collectively, these indicators represent physical (texture), chemical (CEC, cations, BS %) aspects of soil health and together explain over 90 % of the variability observed in the micro watershed. The use of these selected parameters ensures that the Soil Quality Index (SQI) derived from this MDS is comprehensive, scientifically robust and sensitive to changes in soil properties across different land uses (15-17).

Soil Quality Index (SQI) and Comprehensive SQI (CSQI)

The SQI was calculated for each sample using standardized z-scores and weighted contributions of the selected indicators from each PC (18, 19). The CSQI (Table 6), integrating variance-weighted SQIs from all PCs, ranged from -0.48 to 1.18, reflecting strong spatial heterogeneity in soil quality across the micro watershed and findings are in line with (20). Approximately 20 % of the samples were classified as very good (CSQI ≥ 0.80), 40 % as good

Table 6. Comprehensive soil quality index (CSQI) based on studied soil indicators using PCA

Samples	Sand	Silt	Clay	CEC	BS (%)	SQI_pc1	SQI_pc2	SQI_pc3	SQI_pc4	CSQI
1	2.11	0.97	-0.74	-1.18	-0.41	-0.312	-0.468	0.250	0.424	-0.106
2	-8.06	1.36	0.58	-0.43	0.39	-0.603	-1.118	0.688	0.852	-0.180
3	-1.22	1.51	0.66	-0.30	0.75	0.240	-0.909	0.459	0.675	0.464
4	-1.26	1.33	0.78	-0.18	0.76	0.264	-0.815	0.412	0.603	0.465
5	-1.39	1.28	0.95	-0.19	1.15	0.262	-0.824	0.409	0.598	0.445
6	0.44	1.57	-1.24	-0.51	0.80	0.476	-0.786	0.314	0.486	0.490
7	0.24	1.62	-1.03	-0.76	0.57	0.326	-0.851	0.382	0.583	0.440
8	-0.92	1.04	0.51	0.98	0.91	0.631	-0.519	0.177	0.256	0.546
9	-1.13	1.50	0.56	0.87	0.94	0.685	-0.777	0.310	0.452	0.671
10	-1.39	1.86	0.71	0.70	1.07	0.709	-1.000	0.429	0.629	0.767
11	-0.46	0.31	0.33	0.29	0.98	0.190	-0.201	0.054	0.079	0.122
12	-0.17	0.64	-0.15	-0.10	0.51	0.117	-0.353	0.148	0.223	0.135
13	1.31	0.81	-1.95	-0.54	-1.33	0.567	-0.244	0.111	0.182	0.617
14	0.15	-0.60	0.01	0.03	0.06	-0.119	0.308	-0.152	-0.226	-0.189
15	-0.03	-0.48	0.16	0.16	0.03	-0.026	0.248	-0.122	-0.184	-0.084
16	-0.17	-0.28	0.21	0.27	0.05	0.064	0.146	-0.078	-0.120	0.013
17	-0.01	0.09	-0.11	0.20	-0.07	0.126	-0.014	-0.007	-0.012	0.093
18	-1.22	1.14	0.80	1.72	0.72	0.981	-0.513	0.152	0.211	0.830
19	-1.34	1.14	0.95	1.82	0.75	1.039	-0.519	0.155	0.214	0.889
20	-1.53	1.43	1.04	1.98	0.75	1.174	-0.660	0.217	0.304	1.035
21	-1.72	1.73	1.11	2.16	1.04	1.308	-0.814	0.271	0.384	1.149
22	-1.81	1.54	1.30	2.13	0.85	1.284	-0.732	0.252	0.352	1.156
23	-1.89	1.60	1.40	2.01	0.74	1.262	-0.778	0.290	0.409	1.183
24	-0.04	0.39	-0.21	-0.66	-0.06	-0.129	-0.254	0.157	0.235	0.009
25	-0.18	0.05	0.09	-0.23	0.59	-0.093	-0.085	0.035	0.053	-0.090
26	-0.40	0.59	0.16	-0.05	0.80	0.095	-0.358	0.150	0.223	0.111
27	1.21	-0.22	-1.40	-1.27	-1.91	-0.090	0.178	0.006	0.018	0.112
28	0.06	-0.95	0.22	-0.71	-0.39	-0.452	0.404	-0.121	-0.185	-0.354
29	-0.13	-0.75	0.38	-0.56	-0.32	-0.333	0.302	-0.077	-0.120	-0.228
30	0.71	-0.75	-1.71	-1.14	-1.71	-0.161	0.450	-0.143	-0.220	-0.073
31	-0.35	-1.28	0.85	-0.79	-0.88	-0.469	0.529	-0.124	-0.199	-0.264
32	-0.43	-1.20	0.91	-0.65	-0.80	-0.389	0.491	-0.115	-0.186	-0.199
33	-0.52	-1.02	0.94	-0.62	-0.43	-0.348	0.386	-0.081	-0.134	-0.176
34	-0.57	-1.01	1.00	-0.61	-0.96	-0.320	0.399	-0.062	-0.109	-0.093
35	2.40	-1.33	-2.27	-1.45	-2.01	-0.118	0.822	-0.333	-0.473	-0.102
36	0.57	-1.45	-0.13	-0.66	-2.15	-0.481	0.768	-0.242	-0.368	-0.323
37	0.32	-1.48	0.18	-0.51	-0.43	-0.485	0.708	-0.281	-0.422	-0.480
38	0.94	-1.29	-0.63	-0.27	-0.41	-0.182	0.715	-0.340	-0.500	-0.308
39	0.32	-1.20	0.05	-0.19	-0.24	-0.319	0.603	-0.265	-0.396	-0.377
40	0.23	-1.17	0.15	-0.12	-0.05	-0.280	0.580	-0.261	-0.390	-0.351
41	0.18	-1.19	0.21	-0.03	0.12	-0.245	0.588	-0.275	-0.411	-0.344
42	0.11	-1.11	0.28	-0.01	0.46	-0.216	0.527	-0.258	-0.386	-0.333
43	0.05	-0.97	0.30	0.09	0.39	-0.143	0.468	-0.232	-0.348	-0.254
44	0.00	-1.05	0.38	0.17	0.52	-0.122	0.504	-0.256	-0.383	-0.257
45	2.12	-0.80	-2.16	-1.74	-2.65	-0.145	0.528	-0.141	-0.191	0.052
46	1.40	-0.69	-1.40	-1.44	-1.72	-0.258	0.395	-0.096	-0.131	-0.090
47	1.32	-0.61	-1.35	-1.26	-0.58	-0.212	0.326	-0.121	-0.163	-0.169
48	1.08	-0.50	-1.10	-1.18	-0.39	-0.236	0.251	-0.085	-0.112	-0.183
49	0.90	-0.88	-0.75	-1.15	0.06	-0.421	0.400	-0.161	-0.227	-0.408
50	1.33	0.18	-1.72	-0.93	-1.36	0.213	0.015	0.024	0.051	0.302
51	0.96	0.11	-1.25	-0.67	-1.35	0.159	0.036	0.021	0.040	0.255
52	0.94	0.48	-1.41	-0.75	-1.47	0.253	-0.148	0.112	0.176	0.392
53	-0.41	0.11	0.32	1.67	0.88	0.693	0.055	-0.150	-0.229	0.370
54	-1.01	-0.13	1.11	1.55	0.93	0.688	0.091	-0.123	-0.198	0.458
55	-1.13	-0.09	1.21	1.44	1.21	0.655	0.038	-0.097	-0.160	0.437
56	-1.24	0.09	1.28	1.35	1.68	0.652	-0.087	-0.048	-0.085	0.432
57	-1.25	0.05	1.32	1.32	1.57	0.643	-0.072	-0.047	-0.084	0.440
58	-1.25	0.10	1.30	1.26	1.51	0.627	-0.099	-0.028	-0.056	0.444
59	1.73	-0.25	-1.98	-0.68	-0.77	0.288	0.273	-0.154	-0.207	0.199
60	0.05	-0.45	0.01	-0.14	-0.37	-0.150	0.227	-0.083	-0.127	-0.132
61	-0.02	-0.81	0.28	0.05	-0.10	-0.120	0.400	-0.176	-0.267	-0.163
62	-0.21	-0.77	0.48	0.25	0.02	-0.012	0.375	-0.171	-0.262	-0.071
63	0.54	0.13	-0.75	0.22	0.76	0.317	-0.004	-0.081	-0.109	0.124

(0.60-0.79) and the remaining distributed across fair, poor and very poor categories following the (3, 10). Negative CSQI values reflect soil samples where the summed, weighted standardised soil quality indicators are below the overall mean, representing less favourable or degraded soil conditions relative to others in your data set. The classification based on the obtained negative value as, very poor- <0 (3).

Areas with higher clay, CEC and BS % consistently exhibited the higher CSQI values, indicating fertile, fine-textured soils with greater nutrient retention and the same was observed by 23. In contrast, coarse-textured soils with low OC and CEC scored poorly, suggesting a need for soil improvement interventions as recorded from (18, 21).

Spatial representation of CSQI values showed that higher soil quality zones were concentrated in areas with finer texture and higher cation exchange capacity, likely corresponding to lower landscape positions or alluvial deposits and a similar opinion was given by previous researchers (20). Conversely, upland or coarse-textured regions displayed lower CSQI values, indicating vulnerability to erosion, leaching and reduced crop productivity potential (22, 23).

Fig. 2 indicates that PCA 1 and PCA 2 contribute 57.4 % and 17.3 % to the variation. Points close together are similar in terms of the original variables (e.g., 42 and 62 in Fig. 2), while points far apart are different (for instance, 11 and 6). The potassium in Fig. 2 pointing towards right, highlighting that the soil samples have a higher K value (22, 24). K and Ca are pointing both upwards, showing the positive correlation, higher the K value, will have higher Ca in the soil samples. They show an angle of 90° to each other showing no correlation between each other. The K and sand have negative correlation with each other as vector of these are pointing in opposite direction, tends to have higher potassium level with lower sand value.

The length of each vector represents the strength of the variable's contribution to the principal components. Longer vectors indicate that the variable has a strong influence on the PCs, capturing significant variance. A long vector with a high cos2 value indicates that the variable is well represented by the principal components plotted. If "K" has a high cos2 and a long vector, it means potassium content is both influential and accurately represented by PC1 and PC2 (22, 24).

Points near the direction of a vector have higher values for that variable. For example, points near the "OC (%)" vector have higher organic carbon content (19). The sum of the variance explained by PC1 and PC2 gives an idea of how much of the total variance in the data is captured by these two components (24).

Cluster analysis and soil zonation

Cluster analysis

Cluster analysis was performed to group soils with similar characteristics and to delineate management zones within the micro watershed. Both k-means clustering and hierarchical clustering were applied to the standardized dataset using the selected MDS indicators (sand, silt, clay, BS % and CEC) (25, 26).

K-Means clustering

The optimum number of clusters was determined by evaluating the within-cluster sum of squares and inspecting the elbow of the scree plot, which suggested that three clusters (Fig. 3) adequately captured the major patterns in the dataset (27). The k-means algorithm partitioned the 63 soil samples into three distinct clusters (Fig. 4 & Table 7).

Cluster 1 contained soils with moderately high clay and CEC values, intermediate BS % and moderate OC content. These soils generally exhibited good soil quality and were classified as moderately fertile. Cluster 2 grouped soils with the highest clay content, highest CEC and highest BS % values, coupled with adequate OC levels. This cluster represented the most fertile soils in the study area and corresponded to the "Very Good" CSQI class, indicating optimal conditions for sustainable crop production. Cluster 3 consisted primarily of coarse-textured soils (higher sand content), with low CEC, low total cations and relatively low OC. These soils exhibited poor nutrient retention capacity, lower CSQI scores and were classified as "poor to very poor" quality, indicating the need for immediate soil improvement measures.

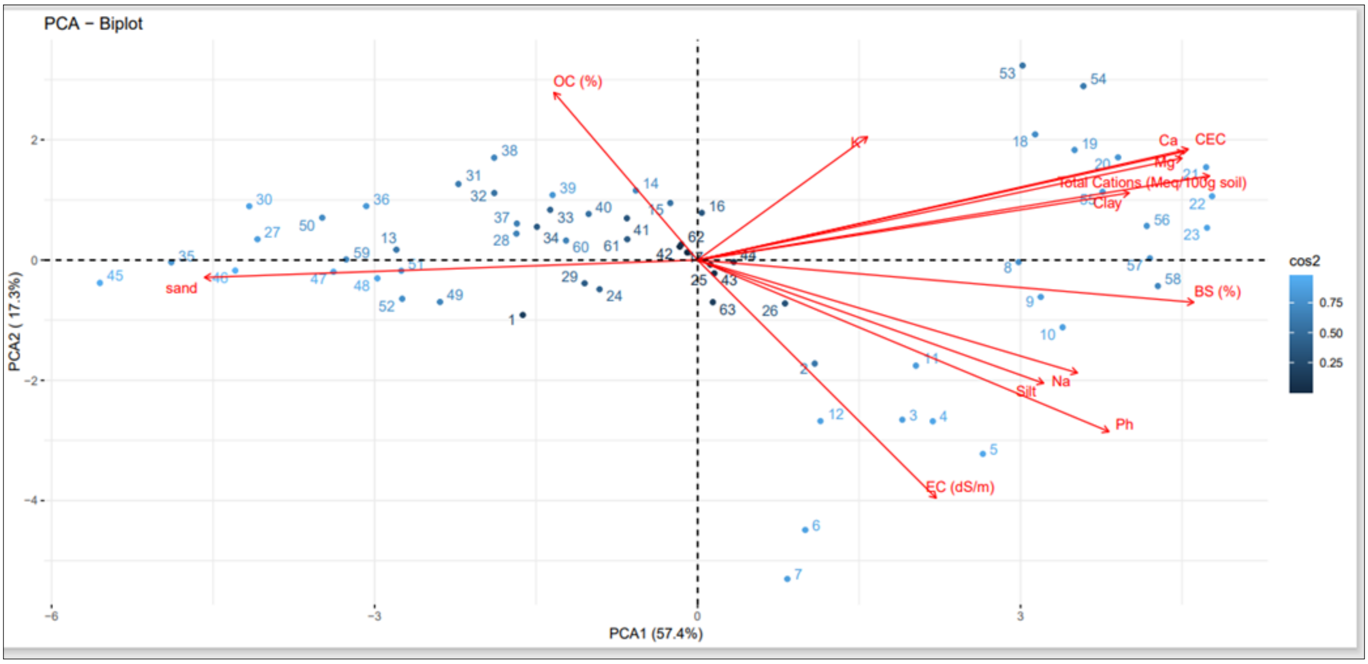

Fig. 2. PCA biplot showing the contribution from both the principal components.

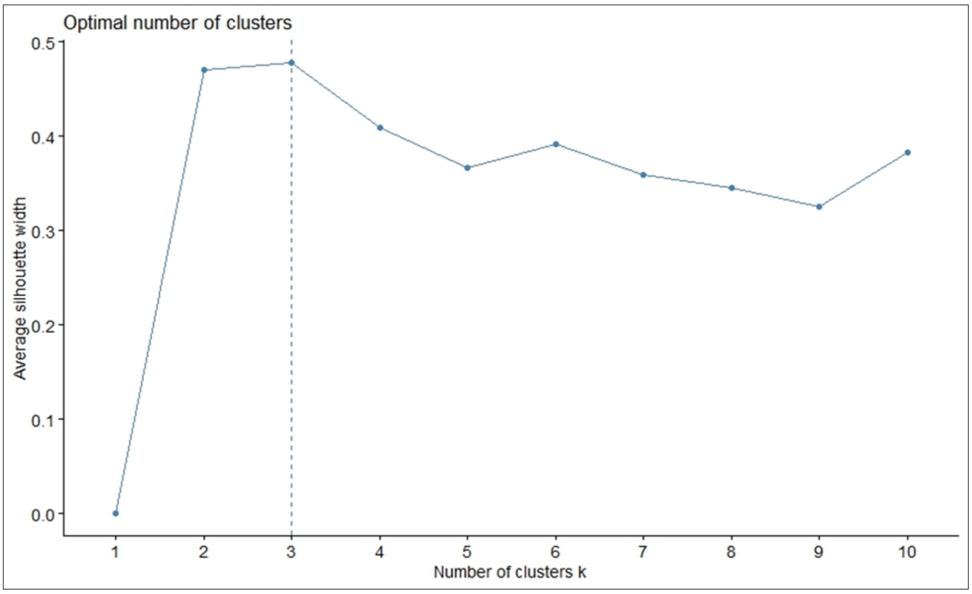

Fig. 3. Indicates the numbers of cluster need to be considered for grouping the data.

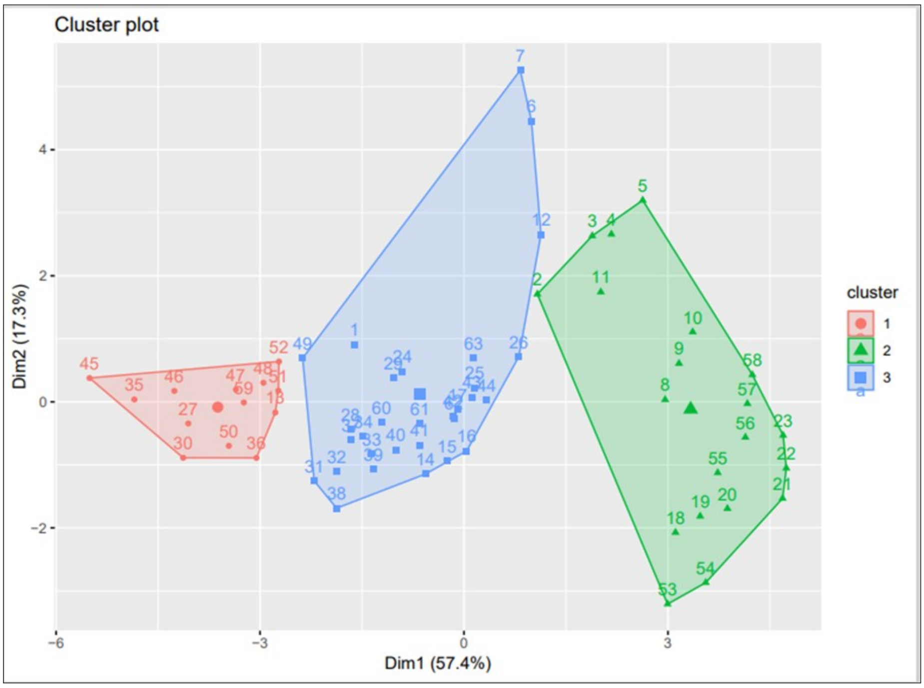

Fig. 4. The clusters of the data showing 3 clusters.

Table 7. The classification soil samples by 3 clusters by k- means clustering

Observation	Cluster	Observation	Cluster	Observation	Cluster
1	3	22	2	43	3
2	2	23	2	44	3
3	2	24	3	45	1
4	2	25	3	46	1
5	2	26	3	47	1
6	3	27	1	48	1
7	3	28	3	49	3
8	2	29	3	50	1
9	2	30	1	51	1
10	2	31	3	52	1
11	2	32	3	53	2
12	3	33	3	54	2
13	1	34	3	55	2
14	3	35	1	56	2
15	3	36	1	57	2
16	3	37	3	58	2
17	3	38	3	59	1
18	2	39	3	60	3
19	2	40	3	61	3
20	2	41	3	62	3
21	2	42	3	63	3

The spatial distribution of the three k-means clusters revealed clear zonation across the micro watershed, with Cluster 2 concentrated in the lower landscape positions (likely alluvial or depositional zones) and Cluster 3 dominating the upland and eroded areas. The observations are in line with previously reported studies that emphasize the utility of k-means clustering for delineating soil management zones based on physical and chemical properties (25, 26).

Hierarchical clustering

Hierarchical agglomerative clustering was performed using Ward's minimum variance method to confirm the grouping obtained from k-means analysis. The resulting dendrogram (Fig. 5) displayed a clear separation into three major clusters (Table 8), supporting the k-means solution (25). The cluster centroids (Table 9) showed similar trends, with Cluster 2 having the highest mean CEC (20.71 cmol(+)/kg) and BS % (90.41 %), while Cluster 3 recorded the lowest mean CEC (9.70 cmol(+)/kg) and BS % (64.61 %). Hierarchical clustering additionally revealed the relative similarity among individual samples, helping to visualize transitional soils

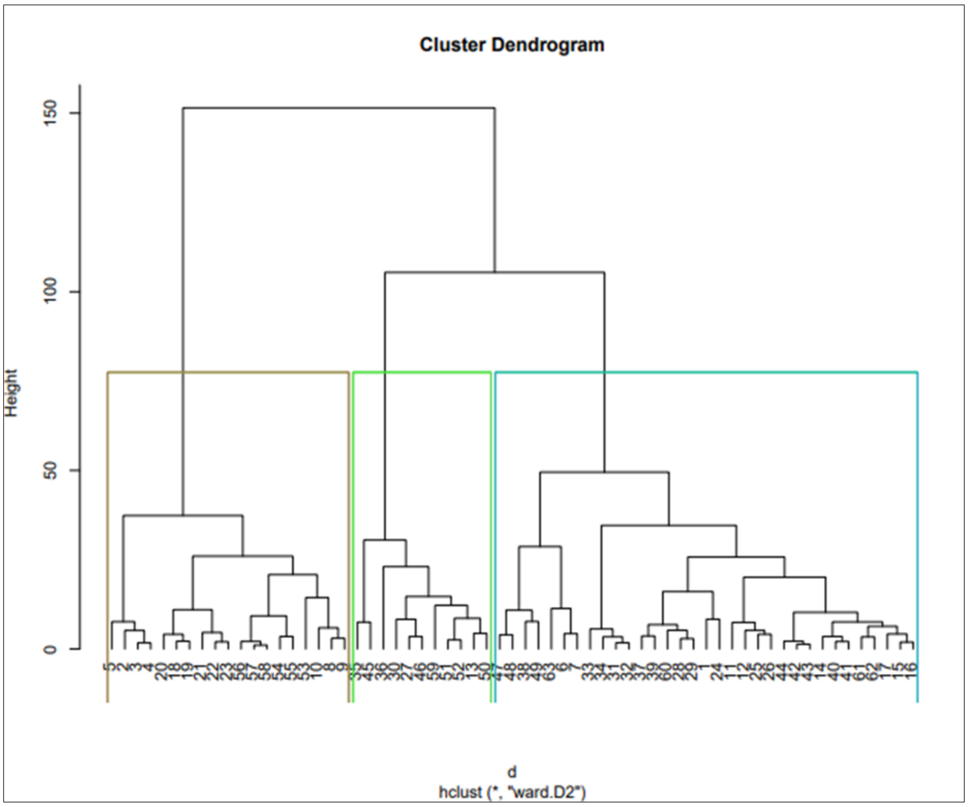

Fig. 5. Dendrogram for Hierarchical clustering.

Table 8. Clustering by means of hierarchical clustering

Observation	Cluster	Observation	Cluster	Observation	Cluster	Observation	Cluster
1	1	19	2	37	1	55	2
2	2	20	2	38	1	56	2
3	2	21	2	39	1	57	2
4	2	22	2	40	1	58	2
5	2	23	2	41	1	59	3
6	1	24	1	42	1	60	1
7	1	25	1	43	1	61	1
8	2	26	1	44	1	62	1
9	2	27	3	45	3	63	1
10	2	28	1	46	3		
11	1	29	1	47	1		
12	1	30	3	48	1		
13	3	31	1	49	1		
14	1	32	1	50	3		
15	1	33	1	51	3		
16	1	34	1	52	3		
17	1	35	3	53	2		
18	2	36	3	54	2		

Table 9. Indicates the centres of the 3 clusters with respective soil parameters through hierarchical clustering

Clusters	1	2	3	4	5	6	7	8	9	10	11	12	13
C1	56.16	8.43	34.97	7.67	0.27	0.31	6.76	3.09	0.62	0.12	10.82	13.31	80.85
C2	41.89	13.86	43.67	9.05	0.38	0.34	11.61	5.13	1.78	0.33	18.85	20.83	90.41
C3	68.91	8.77	20.74	6.64	0.10	0.45	3.77	2.00	0.22	0.16	6.33	9.70	64.61

Note: 1-13 are soil parameters C is clusters

lying near the cluster boundaries. For example, several samples in Clusters 1 and 3 formed sub-groups that may require site-specific management strategies distinct from the core cluster populations, which is a benefit highlighted in earlier cluster analysis studies (26).

The integrated interpretation of the cluster analysis and CSQI classification clearly delineated the micro watershed into distinct fertility zones, with Cluster 2 soils exhibiting the highest clay and CEC as well as the highest soil quality indices and Cluster 3 soils showing coarse texture and poor nutrient status with the lowest CSQI scores. The observations are in line with published research demonstrating that combining clustering methods for soil zonation is highly useful for precision agriculture and land-use planning, enabling targeted, site-specific soil management (27, 28). For instance, low-fertility zones such as Cluster 3 may benefit from organic amendments, gypsum or lime application where pH limits exist and balanced fertilizer regimes, while high-fertility zones like Cluster 2 can be maintained with conservation nutrient inputs to sustain productivity (27). Such targeted management facilitates efficient resource allocation, risk reduction and sustainable crop production, as reinforced by multiple studies applying clustering for agricultural management zone delineation (27, 28).

The sample 1 is considered in cluster 3 in K-means clustering while, same is considered under cluster 1 in hierarchical clustering, this might be due to the k means assigns the positions based on the proximity to centroids, which can shift positions with each iteration while hierarchical cluster groups are based on the pairwise distances and linkage, producing clusters that may have irregular shapes and sizes.

Thematic Mapping of Soil Quality Index (SQI)

A thematic map of the Soil Quality Index (SQI) was prepared to visualize the spatial distribution of soil quality across the micro-watershed (Fig. 6). The CSQI values derived from PCA-based MDS indicators were interpolated using Inverse Distance Weighting (IDW) in GIS to create a continuous surface.

The resulting thematic map revealed a distinct spatial heterogeneity in soil quality. Very Good (1 ha) zones were concentrated in the lower and relatively flat parts of the micro-watershed (Table 10), which typically have finer textures, higher CEC and greater base saturation. These areas correspond to depositional sites where clay and nutrients accumulate, making them suitable for intensive agriculture with minimal intervention. Good quality (27 ha) soils were distributed across mid-slopes and gently undulating terrains, representing moderately fertile soils that require balanced nutrient management to sustain productivity.

In contrast, fair and poor-quality soils (123.4 and 87 ha, respectively) were predominantly located in the upper slopes and eroded uplands, where sand content was higher and OC and CEC were lower. These areas are more prone to runoff and nutrient loss, requiring soil conservation measures such as mulching, contour farming, cover cropping and organic matter addition. A few patches of very poor soils (175 ha) were observed in severely degraded areas, likely due to erosion, low base saturation and nutrient depletion. These zones (Table 10) are priority areas for rehabilitation through integrated nutrient management, organic amendments and soil structure restoration.

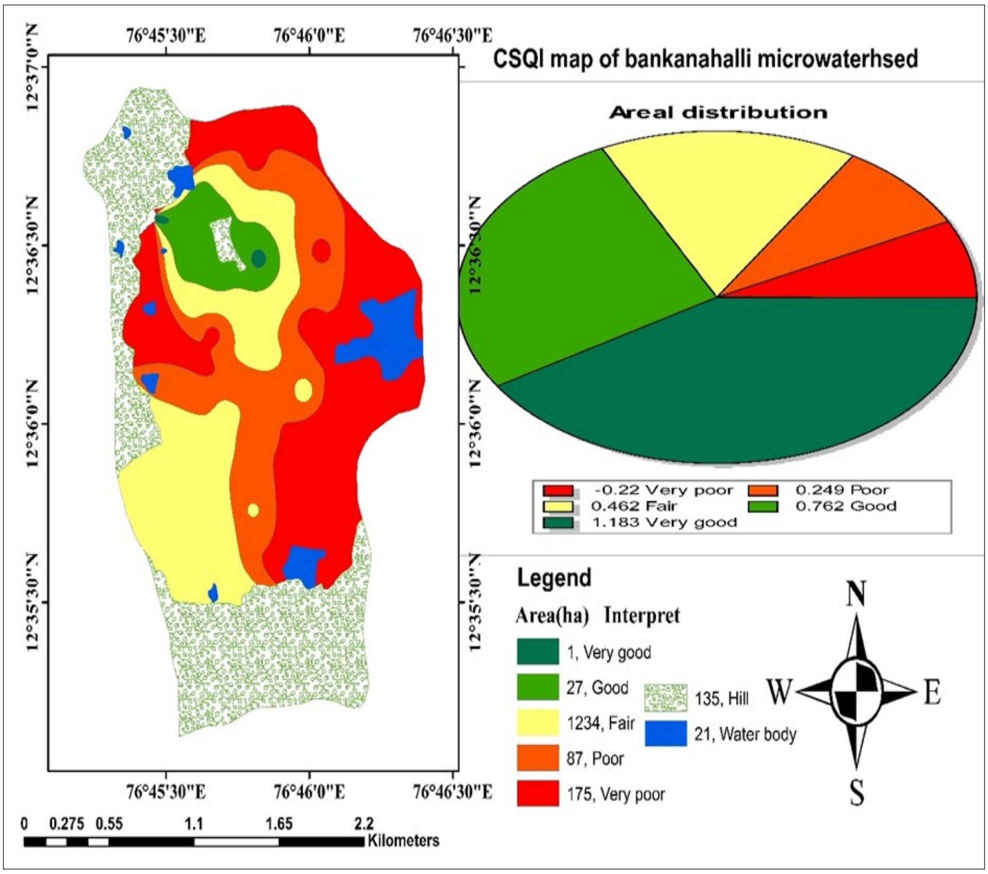

Fig. 6. Thematic map of CSQI.

Table 10. Statistical analysis of the zonation of soil parameters and quality index in the study area

Indicators	First zone Value	Second zone Value	Third zone Value	F value	Significance
Sand	55.23[ab]	42.12[b]	68.91[a]	82.96	S
Silt	8.88[a]	13.56[a]	8.77[a]	12.93	N
Clay	35.43[ab]	43.75[a]	20.74[b]	66.44	S
pH	7.75[a]	9.04[a]	6.64[a]	20.79	N
EC (dS/m)	0.28[a]	0.39[a]	0.10[a]	4.82	N
OC (%)	0.31[a]	0.34[a]	0.45[a]	3.76	N
Ca	7.07[ab]	11.53[a]	3.77[b]	38.80	S
Mg	3.24[ab]	5.06[a]	2.00[b]	36.69	S
Na	0.66[a]	1.83[a]	0.22[a]	21.78	N
K	0.12[a]	0.36[a]	0.16[a]	8.01	N
Total Cations (Meq/100g soil)	11.32[ab]	18.77[a]	6.33[b]	52.64	S
CEC	13.79[a]	20.71[ab]	9.70[b]	38.19	S
BS (%)	81.36[a]	90.50[a]	64.61[b]	97.65	S

Note: The letters are symbols of Duncan test that the means followed by the same letter in each row are not significantly different from one another at a 5 % probability level (Duncan Multiple Range Test).

Thematic mapping of SQI thus provided an effective visualization tool for identifying spatial patterns of soil health and guiding site-specific nutrient management (SSNM). By integrating this map with land use planning, farmers and policymakers can prioritize interventions, allocate resources efficiently and monitor changes in soil health over time (28-30).

Conclusion

This study demonstrates that multivariate statistical approaches effectively assess soil quality and support actionable management plans at the micro-watershed scale. PCA successfully reduced the dimensionality of the dataset and identified a robust MDS comprising key texture, fertility and biological (OC) indicators. The derived CSQI captured more than 90 % of the total variability, providing a reliable measure of soil health. The combination of k-means and hierarchical clustering revealed three distinct fertility zones, corroborating the SQI classification and enabling clear delineation of management units. Spatial analysis and thematic mapping highlighted that soils with higher clay, CEC and BS % were concentrated in depositional zones and exhibited superior quality, whereas coarse-textured upland soils were nutrient-poor and vulnerable to degradation. These findings underscore the need for site-specific interventions in low-SQI areas, including organic matter addition, nutrient balancing, erosion control and conservation agriculture practices. In contrast, high-quality zones can maintain current management to sustain productivity. Overall, the integration of PCA-based SQI, CSQI, clustering and GIS mapping provides a comprehensive framework for soil health monitoring and precision land management. This approach can be replicated in other agro-ecosystems to optimize resource allocation, reduce input costs and promote long-term soil sustainability.

Acknowledgements

The study was conducted as a part of the REWARD project (UAS, Bangalore), funded by the World Bank the Government of Karnataka and the Centre of Excellence on Watershed Management, UAS, GKVK, Bangalore. The authors greatly acknowledge the technical and financial support provided for this research.

Authors' contributions

AG carried out the studies and participated in drafting the manuscript. SBY and AMA planned and prepared the draft for research. MNT participated in the design of the study, TP, PP and SK performed the statistical analysis and mapping. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors declare that they don't have any conflict of interest.

Ethical issues: None

Declarations: The authors confirm that generative AI or AI-assisted technologies were not used in preparing this manuscript.

References

1. Klimkowicz-Pawlas A, Ukalska-Jaruga A, Smreczak B. Soil quality index for agricultural areas under different levels of anthropopressure. Int Agrophys 2019;33(4):455-69. https://doi.org/10.31545/intagr/113349
2. Lenka NK, Meena BP, Lal R, Biswas S, Patra AK, Mandal B. Comparing four indexing approaches to define soil quality in an intensively cropped region of Northern India. Front Environ Sci 2022;10:865473. https://doi.org/10.3389/fenvs.2022.865473
3. Mukherjee A, Lal R, Zimmerman AR. Comparison of soil quality index using three methods. PLoS ONE 2014;9(8):e105981. https://doi.org/10.1371/journal.pone.0105981
4. Celis RAO, Gutiérrez L, Díaz-Zorita M. Conceptual and practical challenges of assessing soil quality: A systematic review. Soil Use Manage 2024. https://doi.org/10.1111/sum.13137
5. Kahsay A, Desta B, Tesfaye D. Developing soil quality indices to investigate degradation under varied land use managements in Northern Ethiopia. Heliyon 2025;11(6):e024172167. https://doi.org/10.1016/j.heliyon.2024.e41185
6. Maguzu J, Paz-Kagan T, Meitasari D, Mesfin A, Mulat T, Adegbite KI, et al. Developing soil quality indices for predicting site classes in *Pinus patula* stands in Tanzania. Agric Ecosyst Environ 2024;358:2433330.
7. Vasu D, Naorem A, Khasi D, Mamabolo RR, Bouma J, Chang SE, et al. A novel and comprehensive soil quality index integrating morphological, physical, chemical and biological properties. Geoderma 2024;444:116247. https://doi.org/10.1016/j.still.2024.106246
8. Hauser E, Chorover J, Cook CW, Markewitz D, Rasmussen C, Richter DD, et al. Integrating decadal and century-scale root development with longer-term soil development to understand terrestrial nutrient cycling. Soil Sci 2023. https://doi.org/10.2139/ssrn.4202049
9. Nolan AL. Field and laboratory methods for assessing soil quality across landscapes. In: Lal R, Stewart BA, editors. Soil Quality and Ecosystem Services. CRC Press; 2006:123-35.
10. Mohamed KM, Abdel-Fattah MK, Mohamed AY. Quantitative evaluation of soil quality using principal component analysis and clustering approach. J Sustain Agric 2021;13(4):1824. https://doi.org/10.3390/su13041824
11. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53-65. https://doi.org/10.1016/0377-0427(87)90125-7
12. Grosjean P, Ibanez F, Etienne M. pastecs: Package for Analysis of Space-Time Ecological Series (version 1.4.2) [computer software]. CRAN 2024.
13. Dodge Y. The concise encyclopedia of statistics. Springer 2008.
14. Li N, Li Y, Li J, Wu X, Feng J, Wang S, et al. Construction and confirmatory factor analysis of the core competence scale for operating room nurses: A cross-sectional study. BMC Nurs 2020;19:75.
15. Parra-González MA, Rodríguez-Valenzuela M. Multivariate analysis for soil quality evaluation: A methodological approach. Environ Monit Assess 2017;189(8):382.
16. Menge DNL, Chu PD, Potani NA. Soil quality evaluation based on minimum data set: PCA and field testing. Soil Sci Soc J 2024;88(2):384-96.
17. Martín-Sanz J, Cobo JG, et al. Characterizing soil quality using principal component analysis and minimum data set approaches: Case study in Mediterranean agroecosystems. Soil Use Manage 2022;38(1):115-27.
18. Doran JW, Parkin TB. Defining and assessing soil quality. In: Lal R, Coleman DC, Bezdicek DF, Stewart BA, editors. Soil Quality for

Crop Production and Ecosystem Health. Soil Sci Soc Am Spec Publ 35; 1994:3-21.

19. Romano-Armada N, et al. Construction of a combined soil quality indicator to assess impacts of agricultural practices. Sci Total Environ 2019;665:70-8.

20. Osman KT, Mohamed HH, Ali SH. Quantitative evaluation of soil quality using principal component analysis: The case study of El Fayoum Depression, Egypt. Sustainability 2022;13(4):1824. https://doi.org/10.3390/su13041824

21. NRCS. Guidelines for soil quality assessment. Natural Resources Conservation Service, US Department of Agriculture 2023.

22. Abdel-Fattah MK, Mohamed KM. Quantitative evaluation of soil quality using principal component analysis and clustering approach. Sustainability 2021;13(4):1824.

23. Larson WE, Pierce FJ. The soil quality concept: Implications for finding and managing soil resources. Soil Sci Soc Am J 1994;58 (2):4-10.

24. Ferraz GAS. Principal components in the study of soil and plant attributes: Application in agricultural research. Rev Bras Eng Agríc Ambient 2019;23(2):424-31.

25. Maione C. A cluster analysis methodology for the categorization of watershed soils. Int J Cluster Sci 2023.

26. Aytaç E. Hydrological response unit based k-means clustering: Application in soil and landscape classification. Environ Softw 2020;136:157-65.

27. Miruthula S, Swathe M, Nivetha V. Crop yield prediction using clustering algorithms: Applications in precision agriculture. Int J Res Publ Rev 2025;6(9):1024-34.

28. Sharma M, Goel S, Elias AA. Predictive modeling of soil profiles for precision agriculture: A case study in safflower cultivation environments. Sci Rep 2025;15:89395. https://doi.org/10.1038/s41598-024-83551-9

29. Yadav MBN, Patil PL, Rundan V, Veda Vyas R, Nthebere K. Assessing the spatial variability of soil quality index of Ganjigatti Sub-Watershed using GIS-based geostatistical modeling. Indian J Ecol 2024;51(1):96-103.

30. Jalhoum MEM, Abou El-Magd MS, El-Sayed SH, El-Samman H. Multivariate analysis and GIS approaches for modeling soil quality index in arid zones. Sci Rep 2024;14:1762. https://doi.org/10.1016/j.heliyon.2024.e27577

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.