Prepublished copy

Nature Environment & Pollution Technology

This is a peer-reviewed prepublished version of the paper to be published in Vol. 25, No. 3, September 2026 issue

Type of the Paper: Original Research

Application of Principal Component Analysis (PCA) and Interpolated Spatial Distribution Maps in Soil Quality Control of Dragon Fruit Farms in Tay Ninh Province, Vietnam

Nguyen Van Phuong†, Tong Hai Minh and Le Hong Thia

Institute of Environmental Science, Engineering & Management, Industrial University of Ho Chi Minh City, 70000, Vietnam

† Corresponding author: Nguyen Van Phuong; nguyenvanphuong@iuh.edu.vn

ORCID IDs of Authors: Nguyen Van Phuong, https://orcid.org/0000-0001-7709-388

Key Words	Dragon fruit, Soil quality index (SQI), PCA, GIS, Minimum dataset (MDS), Soil quality control
DOI	https://doi.org/10.46488/NEPT.2026.v25i03.D1893 (DOI will be active only after the final publication of the paper)
Citation for the Paper	Phuong, N.V., Minh, T.H. and Thia, L.H., 2026. Application of Principal Component Analysis (PCA) and interpolated spatial distribution maps in soil quality control of dragon fruit farms in Tay Ninh Province, Vietnam. *Nature Environment and Pollution Technology*, 25(3), D1893. https://doi.org/10.46488/NEPT.2026.v25i03.D1893

Abstract: Accurate and timely assessment of soil quality is one of the core activities in sustainable farming and agricultural management. Soil quality index (SQI) is greatly affected by fluctuations in physical chemistry indicators of soil. 12 composite soil samples were collected in the dragon fruit growing area, Tay Ninh, Vietnam. The total data set (TDS) has 11 soil parameters, including pH, EC, TOC, CEC, P_av, NH_4^+, bulk density, particle density, clay, silt, and sand, were determined. Principal Component Analysis (PCA) is used to determine the Minimum Dataset (MDS) and weights. These are then combined with scores from land index scoring functions to estimate the Soil Quality Index (SQI). Furthermore, a spatial distribution map of SQI based on the IDW method in ArcGIS 10.8 was also constructed. The results indicate that the MDS depends of four Principal Components (PC), explaining 89.02 of the total data variance, with representative parameters CEC, P_av, EC, and pH corresponding to weights of 0.37, 0.26, 0.22, and 0.15, respectively. The average soil quality assessment accounted for 41.6% and the degraded type accounted for 58.3%. The study demonstrated that the combination of PCA and GIS provided a comprehensive and intuitive SQI evaluation. Furthermore, constructing agronomic maps with large sample sizes to serve sustainable agricultural management based on research results is feasible.

1. INTRODUCTION

Soil quality determines the ability to support plant health, maintain productivity, and stabilize the economy for farmers (Rajput et al., 2023). Comprehensively monitoring soil health and fully understanding the soil indicators and their essential functions are critical steps for developing effective solutions and policies aimed at sustainable land management (Țopa et al., 2025).

Various approaches have been established to quantify and monitor soil quality based on the calculation of indices, including soil quality index (SQI), soil health index (SHI), or soil sustainability index (SSI) as decision support tools (Damiba et al., 2024). However, none of them is widely used/accepted for soil quality assessment. The reason may be due to the intricacy and diversity of soil systems. Total data sets (TDS), "expert" data sets (EDS), and MDS are commonly used for soil quality assessment. (Damiba et al., 2024). The SQI of TDS is the average score of all indices, of EDS is based on weights collected from experts (data not enough for all soil indices), of MDS is based on PCA. In which MDS is considered to require less intensive field and laboratory work, thereby saving time and cost (Damiba et al., 2024)

In addition, many popular studies have focused on using individual physical and chemical indicators, although these are often ambiguous, as many soil properties are the result of interactions between several processes. In some studies, biological indicators have been ignored or excluded (Damiba et al., 2024; Țopa et al., 2025). Priority soil indicators were selected based on the main criteria: easy to measure, quick and low cost (applicability of field techniques), sensitive to changes in management processes (Țopa et al., 2025). Common soil indices used in soil quality assessment include soil physical (bulk density, specific gravity, water retention, clay, sand, and silt), chemical parameters (pH, EC, organic matter, cation exchange capacity, NH_4^+, available phosphorus, copper, and zinc) (Abdel-Fattah et al., 2021; Hanif et al., 2025; Rajput et al., 2023; Van, 2025).

The MDS-based SQI calculation method is currently the most widely used method due to its flexibility, quantitative nature, and ease of use. Statistical tools such as PAC and Pearson analysis are used in many studies (Rajput et al., 2023; Țopa et al., 2025; Van, 2025). This method offers a only SQI index and a set of minimum soil parameters for monitoring soil health (Rajput et al., 2023). PCA is an invaluable tool for data simplification and modeling efficiency. It excels by removing correlated features, thereby enhancing the statistical significance of variables and minimizing multicollinearity. PCA achieves this by converting the dataset from a high-dimensional space to dimensionality reduction, which drastically reduces the number of variables in the model, improving algorithm performance and speeding up analysis. Ultimately, this dimensionality reduction makes data visualization and exploration far more effective (Abdel-Fattah et al., 2021). The IDW Interpolation method in ArcGIS 10.8 is an important technique used to estimate SQI values at unsampled locations based on data from existing sample points that have been used to map the spatial distribution of SQI (Abdel-Fattah et al., 2021; Kahsay et al., 2025).

Tay Ninh province is the second-largest dragon fruit-growing province in Vietnam, more than 7,000 ha (Mai et al., 2025). The cultivated land area represents the characteristics of acid sulfate soils of the Vam Co river basin. However, soil quality management in this dragon fruit growing region still lacks solutions to maintain sustainability in production (precision farming, modern management). The study combines PCA and GIS to identify and simulate the fluctuations of SQI values in dragon fruit farms in Tay Ninh province, Vietnam based on a synthetic dataset of 11 soil parameters, including pH, EC, TOC, P_av, NH_4^+, d, D, % clay, % silt, % sand, CEC.

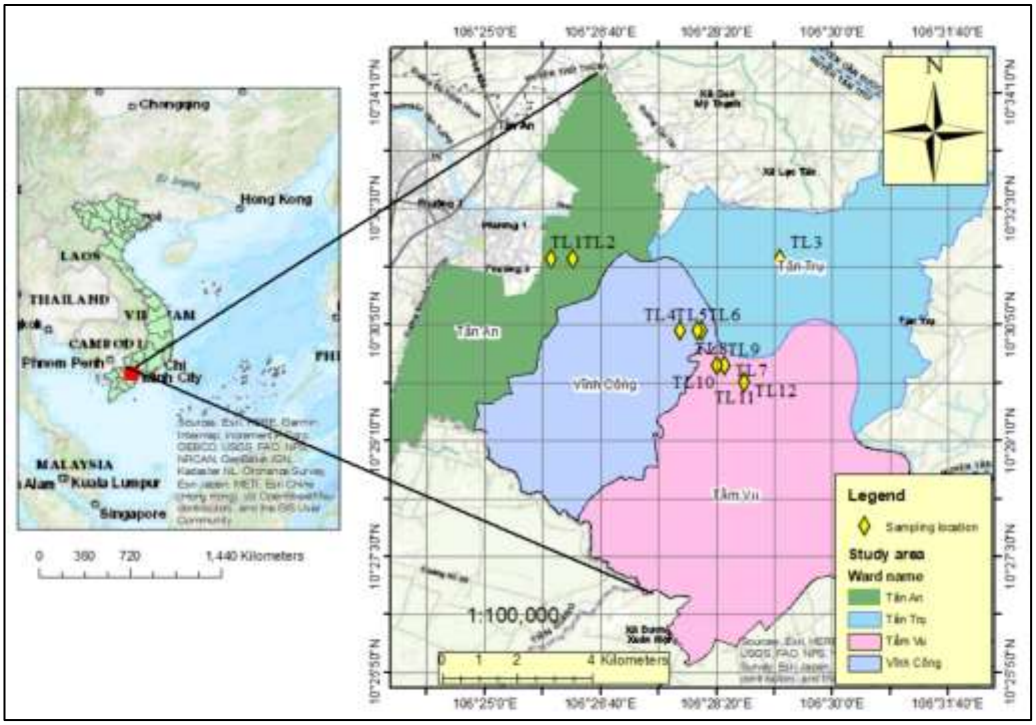


Fig. 1: Map of sampling locations in dragon fruit farms in Tay Ninh, Vietnam (TL1 to TL12 are 12 composite samples collected in the dragon fruit farms, Tay Ninh, Vietnam)

2. MATERIALS AND METHODS

2.1 Field sampling

In January 2024, 60 soil samples in the dragon fruit farms of Tay Ninh province were collected to create 12 composite samples (Fig. 1). Each area, approximately 4 km2 (400 ha), exhibits the highest homogeneity in dragon fruit tree age, as well as in color observation. Within each selected area, a zone of no less than 1.5 ha will be chosen for sampling. The soil samples were taken at a depth of 0-30 cm, this is the soil layer that best represents the dynamic soil quality. The sampling procedure was carried out using the composite sampling method: in each selected growing area (>1.5 ha), composite sampling was conducted at the four corners and the diagonal center. The composite soil sample was naturally dried, finely ground, and sieved through a 2mm sieve to analyze 11 soil parameters: pH, EC, TOC, CEC, P_av, NH_4^+, bulk density, density, clay, silt, sand.

2.2 Analytical methods

Determination of Soil particle density (d) according to TCVN 6863:2001; pH, soil electrical conductivity (EC) according to ISO 10390:1993, and TCVN 6650:2000; bulk density (D) according to TCVN 8305:2009; Total organic carbon (TOC) according to the Walkley Black method; available phosphorus (P_av) according to Olsen method; NH_4^+ according to Baethgen & Alley (Baethgen & Alley, 1989); Soil texture by hydrometer method (George, 1962); Cation exchange capacity (CEC) by the ammonium acetate method according to TCVN 8568:2010

The chemicals used were of analytical purity including H_2SO_4, $K_2Cr_2O_7$, $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$, Na_3PO_3, $SnCl_2$, $(NH_4)_6Mo_7O_{24}$ $4H_2O$, sourced from China and Merck.

2.3 Soil quality assessment

Principal Component Analysis (PCA) is a data reduction technique and simplifies the data structure by extracting principal components (PCs) and analyzing orthogonal variable correlations. Following PCA, the soil quality index (SQI) is calculated through a series of steps: standardizing the variables, creating a correlation matrix, identifying principal components and their explained variance, calculating weights, analyzing factor loadings, and then determining the final SQI (Aisah et al., 2025; Damiba et al., 2024).

Because the measurement indicators have different units, they are standardized on a 0-1 scale to evaluate the overall soil quality, avoiding missing important data (Damiba et al., 2024). Based on reference values and expert judgment, three types of scoring functions, linear and non-linear, categorized as "more is better," "optimal range," and "less is better", are established (Bünemann et al., 2018). For the soil indices: “the more is better” is shown in formula (Eq. 1) (Bandyopadhyay & Maiti, 2021).

$$S_i = \frac{X}{X_{max}} \qquad \textbf{(Eq. 1)}$$

Where X, Xmax, and Xmin are the measured, the maximum value, and the minimum value. Si is the score of indicator i. Non-linear functions including pH, EC, clay, silt, and sand, and linear (NH4+, P_av, TOC, CEC, the more the better), the less the better (bulk density, particle density) presented in detail in Table 1 (Damiba et al., 2024; Kahsay et al., 2025)

For the soil indices: “less is better” is described in formula (Eq. 2):

$$S_i = \frac{X_{min}}{X} \qquad \textbf{(Eq. 2)}$$

For the non-linear scoring function, the soil parameters are converted according to the S-shaped curve equation, formula (Eq. 3) is as follows:

$$S_i = \frac{a}{\left[1 + \left(\frac{X_i}{X_{imean}}\right)^b\right]} \qquad \textbf{(Eq. 3)}$$

Where Xi is the measured value, Ximean is the average value and Xopt is the optimal value (Table 1). Slope b = -2.5 (more is better), b = +2.5 (less is better) và Xo (Damiba et al., 2024; Kahsay et al., 2025)

Table 1. Components of the scoring function

Indices	Scoring curve	Xmax	Xmin	Xopt
Sand (%)	Optimum	90.00	6.00	36
Silt (%)	Optimum	50.00	4.00	30
Clay (%)	More is better	84.00		
D (g cm^{-3})	Less is better	66.00	0.89	
pH (H_2O)	Optimum	9.62	5.04	7.00
EC (dS m^{-1})	Less is better	2	0.001	1
TOC (%)	More is better	5.96		
P_av (mg Kg^{-1})	More is better	58.89		
CEC (cmol Kg^{-1})	More is better	61.98		

A Pearson correlation matrix was constructed to assess the relationships between variables. Correlation coefficients range from -1 to +1, with the magnitude indicating the strength and the sign indicating the direction (positive or negative) of the relationship. These correlations inform the subsequent PCA. PCA reduces the dimensionality of the dataset by transforming the original variables into a new set of uncorrelated principal components (PCs) that capture the majority of the original variance. The indicators with the highest loadings on these PCs represent the parameters with the greatest influence on agricultural land quality (Rangel-Peraza et al., 2017). Communalities are calculated as the percentage of variance explained by each variable in the PC. PCs have eigenvalues greater than 1 and explain at least 10% of the variation in the data. From the selected PCs, a Minimum Data Set (MDS) is constructed by retaining only the variables with significant factor loadings. These "high factor loading" variables are defined as those with the greatest weight on a particular PC, or those with absolute factor loadings within 10% of the highest value variable (Abdu et al., 2023; Damiba et al., 2024). If multiple indicators are in the same principal component (PC), the Pearson correlation matrix is used to test the correlation ($p < 0.05$). The indicator with the highest factor loading is selected if the correlation is strong ($r > 0.5$); if the correlation is weak ($r < 0.5$), all indicators are retained (Damiba et al., 2024).

2.4 SQI calculation and Classification

Calculate the Soil Quality Index Plus (SQI_Plus), formula 4 (Eq. 4) (Damiba et al., 2024)

$$SQI\_Plus = \frac{\sum_{i=1}^{n} S_i}{n} \quad \textbf{(Eq. 4)}$$

SQI is calculated according to formula (Eq. 5):

$$SQI = \sum_{i=1}^{n} W_i . S_i \quad \textbf{(Eq. 5)}$$

In which: n is the total number of related indicators, Wi is the weight of PCi, and Wi is calculated according to formula (Eq. 6) (Damiba et al., 2024). Where PCi is the rotated load of principal component i and ΣPC is the total rotational load.

$$W_i = \frac{PC_i}{\sum PC} \quad \textbf{(Eq. 6)}$$

Classification for SQI levels was assessed according to the scale used by Li et al, where SQI>0.8 (very good), 0.7-0.8 (good), 0.6-0.7 (moderate), 0.4-0.6 (degraded), and <0.4 (poor) (Li et al., 2024)

2.5 Data processing

To ensure reliability, each analysis was performed in triplicate, with the mean, standard deviation, and range calculated using Excel. Data suitability for PCA was assessed via Bartlett's test and the Kaiser-Meyer-Olkin (KMO) measure, where values above 0.5 are acceptable, and Pearson correlation analyzed soil index relationships using SPSS 23.0 software. (Abdel-Fattah et al., 2021; Plonsky, 2015). The spatial distribution map of SQI was made using ArcGIS 10.8 using the inverse distance weighted mean (IDW) interpolation method.

3. RESULTS AND DISCUSSION

3.1 Some soil properties in the study area

The results of determining 11 soil physicochemical indices of the dragon fruit growing area in Tay Ninh, Vietnam, are presented in Table 2. Specifically, the pH values ranged from 6.07 to 7.21, with an average value of 6.48, showing that the pH of the study area is very suitable for plant growth (Mukherjee & Lal, 2014). This showed that farmers were aware of the importance of controlling soil pH, although previous research showed that the soil characteristics of the area were generally alum soil (pH 4.0-5.4) (Mai et al., 2025; Van, 2025). The results showed that the study area was characterized by low EC of soils, with an average EC value of 0.31, ranging from 0.16 to 0.45 mS cm^{-1}. According to Mukherjee & Lal, EC values of 0.2-0.5 mS cm^{-1} were considered to be the most suitable for crop cultivation (Mukherjee & Lal, 2014). The average TOC content is 3.28, ranging from 1.80 to 5.04%, belonging to the soil type with medium to high organic content (Mukherjee & Lal, 2014). However, because dragon fruit is a perennial plant, if the organic supply is not given due attention by farmers, organic matter degradation is likely to occur. The CEC of the study area varied over a wide range, from 11.1 to 33.7 cmol kg^{-1} soil, with an average value of 22.5 cmol kg^{-1}. The results were similar to soil samples in Ben Tre, Vietnam, where CEC ranged from 19.0 to 28.8 cmol kg^{-1}

(Van, 2025). The average available P content was 37, ranging from 12 to 66 mg kg^{-1}, which was lower than that of durian growing area soils with available P ranging from 25.4-117.1 mg kg^{-1} (Van, 2025). This may be due to the different phosphorus requirements of each crop as well as the level of nutritional investment, which is related to the cost of agricultural products. The average NH_4^+ content was 114, ranging from 44 to 148 mg kg^{-1}, indicating that the NH_4^+ content in the soil was of the same high type as the neighboring soil samples in the previous study (Van, 2025). The lowest NH_4^+ value was in soil sample TL6, the area just harvested, and the highest was in sample TL4, the farm was in the fruit-growing stage. Soil particle density ranges from 2.04 to 2.75 g cm^{-3}, average of 2.43 g cm^{-3}. Soil bulk density ranges from 0.87 to 1.19 g cm^{-3}, belonging to a light to medium soil type, which is common in the Southern Delta, Vietnam (Van, 2025). The proportion of clay, silt, and sand ranges from 13 to 38, 13 to 28 and 48 to 65%, respectively. The soil texture is loam, suitable for cultivation.

Table 2: Soil physical and chemical indices in the study area

	pH	EC mS cm^{-1}	TOC %	CEC cmol kg^{-1}	P_av mg kg^{-1}	NH_4^+ mg kg^{-1}	d g cm^{-3}	D g cm^{-3}	Clay %	Silt %	Sand %
TL1	6.17	0.35	3.62	11.1	57	125	2.35	1.06	13	28	60
SD	0.01	0.03	0.14	0.5	8	1	0.13	0.01	1	1	2
TL2	7.15	0.28	3.16	15.7	20	137	2.67	0.91	18	23	60
SD	0.06	0.03	0.16	0.7	1	2	0.07	0.01	1	1	3
TL3	6.92	0.29	3.48	22.5	28	104	2.55	0.98	25	18	58
SD	0.04	0.00	0.24	0.8	6	6	0.02	0.01	1	1	3
TL4	6.77	0.32	4.88	24.7	57	148	2.33	1.03	28	18	55
SD	0.02	0.01	0.16	0.9	2	9	0.16	0.01	1	1	2
TL5	6.60	0.38	4.00	22.5	51	143	2.21	1.01	25	18	58
SD	0.00	0.01	0.28	0.8	3	3	0.09	0.00	1	1	2
TL6	6.46	0.16	3.02	20.4	53	44	2.73	1.15	23	13	65
SD	0.06	0.02	0.02	0.0	2	7	0.02	0.03	1	1	3
TL7	6.32	0.22	2.78	33.7	23	63	2.43	1.10	38	15	48
SD	0.05	0.01	0.10	0.8	1	2	0.04	0.01	1	1	2
TL8	6.25	0.31	3.27	29.3	42	90	2.08	1.01	33	13	55
SD	0.01	0.01	0.13	0.8	6	9	0.01	0.08	1	1	2
TL9	6.45	0.44	3.18	22.5	37	121	2.57	0.92	25	18	58
SD	0.04	0.02	0.18	0.1	2	4	0.02	0.05	1	1	3
TL10	6.22	0.35	3.18	24.7	36	135	2.22	1.05	28	18	55
SD	0.09	0.03	0.18	0.2	1	5	0.11	0.03	1	1	2
TL11	6.31	0.33	2.90	22.5	16	116	2.26	0.91	25	18	58
SD	0.04	0.02	0.10	0.3	1	4	0.22	0.01	1	1	3
TL12	6.12	0.36	1.83	20.4	20	138	2.72	0.93	23	13	65

SD	0.04	0.01	0.03	0.5	8	2	0.01	0.01	1	1	3
Min	6.07	0.16	1.80	11.1	16	44	2.04	0.87	13	13	48
Max	7.21	0.45	5.04	33.7	57	148	2.75	1.19	38	28	65
Average	6.48	0.31	3.28	22.5	37	114	2.43	1.01	25	17	58

SD: standard deviation.

3.2 Pearson correlation analysis

The correlation coefficients between soil parameters are listed in Table 3. Soil pH had low correlations with all other soil parameters. This shows that the use of alkaline substances such as lime (CaO, $CaCO_3$), in pH control and regular control has little impact on other soil parameters, consistent with the soil pH results (Table 3), showing low pH fluctuations. Soil EC had a significant positive correlation ($p < 0.05$) with NH_4^+ ($r = 0.80$). The results showed that the amount of nitrogen mineral fertilizer used contributed to increasing the amount of soluble salts in the soil. Soil TOC was strongly correlated with available P ($r=0.65$, $p<0.01$), suggesting that mineralization of organic matter contributed to the increase in available P through desorption. The significant variation in TOC was attributed to changes in organic matter use strategies in crop management (Ţopa et al., 2025). CEC had a significant positive correlation ($p < 0.01$) with clay ($r=1.00$), silt ($r=0.95$), and sand ($r=0.64$). The results showed that soil texture greatly influenced soil CEC. Similar findings were also found in the study by Kawaguchi & Kyuma, suggesting that clay minerals explain more than 80% of the variation in CEC, while the contribution of organic matter is not statistically significant in paddy soils in tropical Asia (Kawaguchi & Kyuma, 1975)

Table 3: Correlation coefficients of soil parameters

	pH	**EC**	**TOC**	**P_av**	**NH_4^+**	**d**	**D**	**% clay**	**% silt**	**% sand**	**CEC**
pH	1										
EC	-0.165	1									
TOC	0.454^{**}	0.150	1								
P_av	-0.036	0.002	0.654^{**}	1							
NH_4^+	0.176	0.799^{**}	0.308	0.012	1						
d	-0.259	0.286	0.373^{*}	0.254	0.171	1					
D	0.244	0.580^{**}	-0.210	-0.559^{**}	0.563^{**}	-0.117	1				
% clay	-0.162	-0.212	-0.039	-0.202	-0.376^{*}	0.344^{*}	-0.229	1			
% silt	-0.073	-0.125	-0.030	-0.225	-0.275	0.283	-0.087	0.945^{**}	1		
% sand	-0.323	-0.343^{*}	-0.417^{*}	-0.131	-0.507^{**}	-0.052	-0.197	0.621^{**}	0.698^{**}	1	
CEC	-0.163	-0.220	-0.055	-0.210	-0.386^{*}	0.331^{*}	-0.224	0.999^{**}	0.949^{**}	0.639^{**}	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed)

3.3 Principal component analysis

Table 4 presents the Bartlett and KMO test results of the total data set (TDS). Bartlett test shows a significance level <0.0001 and the observed chi-square value (607.619), p-value = 0.000 < 0.05, which indicates that the variables are significantly correlated with each other, and the data is suitable for PCA analysis. In addition, the KMO value is also above 0.538 (>0.50), confirming that the sample size is reliable enough for factor analysis. These results confirm that the data set meets the conditions for performing PCA (Abdel-Fattah et al., 2021).

Table 4: Bartlett and KMO test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.538
Bartlett's Test of Sphericity	Approx. Chi-Square	607.619
	df	55
	Sig.	0.000

The relationship between eigenvalues and PC is shown in Fig. 2. The results showed that four PCs with eigenvalues >1, 4.12, 2.19, 2.14 and 1.35 respectively and they explained more than 89.02% of the variance in the original data were selected (Table 5). The contribution levels of the PCs were 33.2, 22.7, 19.9, and 13.2 % for PC1, PC2, PC3, and PC4, respectively (Table 5).

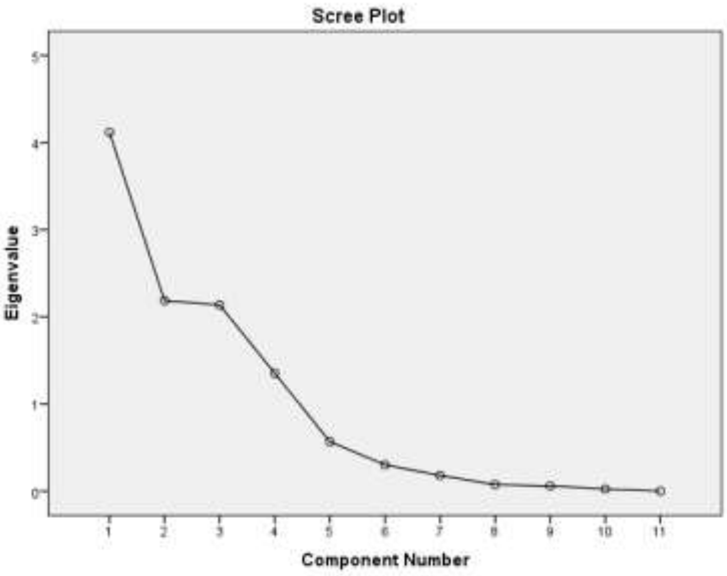


Fig. 2: Eigenvalues by Principal Components

Table 5: PCs Explain Total Variance

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.118	37.439	37.439	3.655	33.223	33.223
2	2.185	19.867	57.306	2.497	22.703	55.926
3	2.136	19.420	76.725	2.184	19.850	75.776
4	1.352	12.295	89.020	1.457	13.244	89.020
5	0.568	5.159	94.179			
6	0.302	2.744	96.924	W_PC1	0.37	
7	0.180	1.637	98.560	W_PC2	0.26	
8	0.075	0.681	99.241	W_PC3	0.22	
9	0.059	0.538	99.780	W_PC4	0.15	
10	0.024	0.220	100.000			
11	2.588E-05	.000	100.000			
Extraction Method: Principal Component Analysis.						

W_PC1, W_PC2, W_PC3, and W_PC4 are the weights of the respective PCs, calculated according to Eq. 6.

Table 6: Component matrix

	Component			
	1	2	3	4
pH	-0.294			**0.932**
EC	-0.474		**0.762**	-0.294
TOC	-0.268	0.830	0.241	0.369
P_av	-0.188	**0.861**	-0.181	-0.197
d	0.174	0.526	0.588	-0.294
% silt	0.864		0.416	0.210
CEC	**0.916**		0.334	0.135
% sand	0.817	-0.193		-0.116
D	-0.393	-0.643	0.535	0.204
% clay	0.909		0.343	0.135
NH_4^+	-0.659		0.652	
Extraction Method: Principal Component Analysis.				

PC1 explained 33.2% of the total variance. Parameters with absolute loading factors >0.6 included CEC, clay, silt, sand, and NH_4^+, with loadings of 0.916, 0.909, 0.864, 0.817, and -0.659, respectively, indicating that these were the parameters with large contributions to PC1 (Table 6). Among them, the loading value of CEC was the highest; the loading values of clay and silt were within 10% of the CEC loading value, which were considered (Abdu et al., 2023; Damiba et al., 2024). However, the significant correlation ($P < 0.01$) between these two indices compared with CEC was r of 1.00 and 0.95 (Table 3), which should be removed to avoid multicollinearity (Damiba et al., 2024). Therefore, only CEC was retained as representative for PC1. The results showed that PC1 represents the soil texture component, which characterizes the weathering process of the parent soil and the impact of soil erosion (which may be caused by rain or irrigation water).

PC2 explained 22.7% of the total variance. Parameters with loadings >0.6 included P_av, TOC, and bulk density D, with loadings of 0.861, 0.830, and 0.643, respectively. Of these, P_av had the largest loading, and the loading value of TOC was within 10% of P_av, which were the two parameters considered in PC2. However, there was a strong correlation between P_av and TOC $r>0.6$ (0.654 with $p<0.01$), Table 3, so only P_av was retained as a representative for PC2 (Damiba et al., 2024). The results showed that the source of available P came from fertilizer use; however, because the study soils were acidic and has a very high phosphorus fixation capacity, the presence of organic matter can contribute to increasing available P. Therefore, PC2 can be considered a representative of the combination of phosphorus and organic matter. PC3 accounted for 19.9% of the total variance. Parameters with loadings >0.6 included EC and NH_4^+, with loadings of 0.762 and 0.652, respectively. Among them, EC had the largest loading, and the loading value of NH_4^+ was within 10% of EC, which were the two parameters considered in PC3. However, there was a strong correlation between EC and NH_4^+ $r>0.6$ ($r=0.799$) with $p<0.01$, Table 3; therefore, only EC was retained as a representative for PC3 (Damiba et al., 2024). PC3 is considered to be representative of soil mineralization and ammonium nitrogen application. PC4 explained 13.2 % of the total variance. pH was the only parameter with a loading of 0.932 (Table 6), and it was retained to represent PC4 (Damiba et al., 2024). PC4 is considered to represent the use of amendments to control soil pH, mainly lime (according to research from the farmers). PC1, PC2, PC3, and PC4, with parameters including CEC, P_av, EC, and pH, were respectively selected as representative components of MDS with a contribution ratio of 89.02%.

The weighted values (W_PC), representing the contribution of each PC, were calculated based on Eq. 6. The results were 0.37, 0.26, 0.22, and 0.15 for PC1, PC2, PC3, and PC4, respectively (Table 6). The general formula for calculating SQI (Eq. 5) for the sampling locations of the growing area is presented in Eq. 7:

$$SQI = \sum_{i=1}^{n} W_i . S_i = 0.37 * S_{CEC} + 0.26 * S_{P_{av}} + 0.22 * S_{EC} + 0.15 * S_{pH} \qquad \textbf{(Eq. 7)}$$

Calculate Si scores for soil indices: The results of the Si scoring based on linear, non-linear, and optimized scoring functions (Eq. 1, Eq. 2, and Eq. 3) for soil indices are presented in Table 7. The results show that, except for Si of soil particle density and bulk density, both >0.5, ranging from 0.75 to 0.98 and 0.76 to 0.96 respectively, most Si of other soil indices show large fluctuations and signs of degradation (<0.5). pH ranging from 0.46 to 0.56 is alarming and requires monitoring. Specifically, indices with large fluctuations include EC (0.16 to 0.7), P_av (0.25 -

0.87), NH_4^+ (0.28-0.94), CEC (0.31-0.87), and clay (0.33-1.00). This is also considered a basis for detecting soil index degradation.

Table 7: Si calculation results based on soil index scoring functions

	pH	**EC**	**TOC**	**P_av**	**NH_4^+**	**d**	**D**	**clay**	**silt**
TL 1	0.47	0.56	0.72	0.87	0.80	0.87	0.82	0.33	0.31
TL 2	0.56	0.43	0.63	0.31	0.87	0.76	0.96	0.47	0.51
TL 3	0.54	0.44	0.69	0.42	0.67	0.80	0.89	0.67	0.72
TL 4	0.53	0.52	0.97	0.87	0.94	0.88	0.85	0.73	0.76
TL 5	0.51	0.61	0.79	0.77	0.91	0.92	0.87	0.67	0.72
TL 6	0.50	0.16	0.60	0.81	0.28	0.75	0.76	0.60	0.66
TL 7	0.48	0.28	0.55	0.34	0.40	0.84	0.80	1.00	0.87
TL 8	0.48	0.49	0.65	0.64	0.58	0.98	0.86	0.87	0.83
TL 9	0.50	0.70	0.63	0.56	0.77	0.80	0.95	0.67	0.72
TL 10	0.47	0.56	0.63	0.54	0.86	0.92	0.83	0.73	0.76
TL 11	0.48	0.52	0.58	0.25	0.74	0.91	0.96	0.67	0.72
TL 12	0.46	0.58	0.36	0.30	0.88	0.75	0.94	0.60	0.66
Min	0.46	0.16	0.36	0.25	0.28	0.75	0.76	0.33	0.31
Max	0.56	0.70	0.97	0.87	0.94	0.98	0.96	1.00	0.87
Average	0.50	0.49	0.65	0.56	0.72	0.85	0.87	0.67	0.68

In this study, the SQI of 12 Dragon Fruit growing areas was estimated based on two calculation methods: Average addition SQI_Plus (calculated from Eq. 4) and PCA analysis SQI_PCA (from Eq. 7) were presented Fig. 3. SQI_Plus has an average value of 0.65, ranging from 0.57 to 0.75, lowest in sample TL 2 and highest in TL 4. SQI_PCA ranges from 0.43 to 0.69, average of 0.57, also the lowest in sample TL 2, highest in TL 4 (Fig. 3). The SQI from these two methods is highly consistent in its variation, although the SQI values vary depending on the method of calculation. Both methods provide the same justification for a given SQI. The SQI_PCA values are consistently lower than the SQI_Plus, which may be due to the contribution of highly correlated parameters to the SQI calculation (multicollinearity).

According to the classification scale applied in the study of Li et al, (Li et al., 2024), SQI_Plus has 3 samples TL 4, TL 5 and TL 8 in good condition (SQI>0.7) accounting for 25%, 6 samples in average condition (SQI 0.6 to 0.7) accounting for 50% and 3 samples (TL 1, TL 2 and TL 6) in degraded condition (SQI 0.4 to 0.6) accounting for 25%. While SQI_PCA shows that there is no soil sample in good condition, 5 samples are in average condition, accounting for 41.6% and 7 samples are in degraded condition, accounting for 58.3%. The results showed that the use of SQI_PCA is more suitable for soil quality degradation risk management. The soil quality in the study farms showed a tendency to degrade, which may be due to the impact of the price of dragon fruit in recent years, which has led to a decrease in investment in care, improvement, and nutrition for the soil. Moreover, according to field observations, most of the dragon fruit farms have been harvested for more than 10 years, and the fragmented cultivation has led to a mixture of good and bad soil quality (Fig. 3). This is also considered a risk in sustainable agriculture.

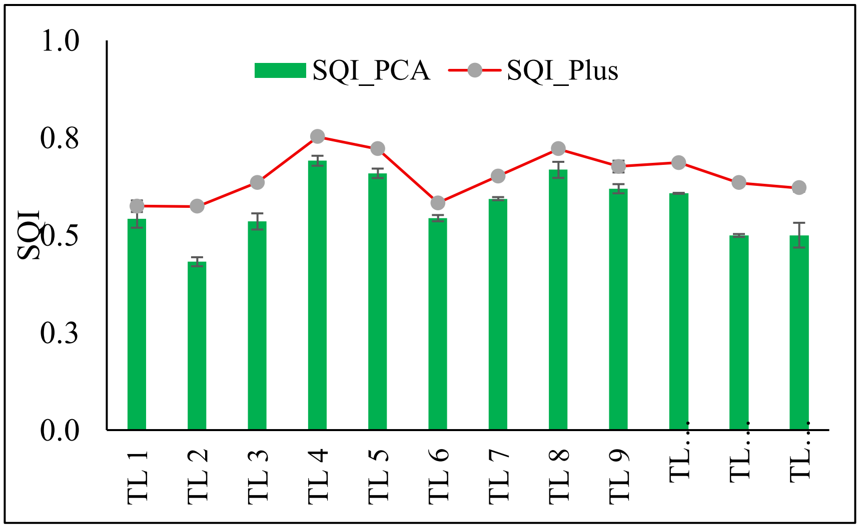
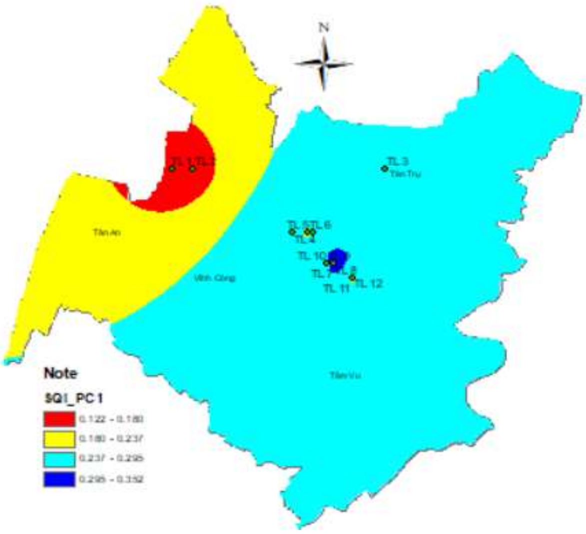
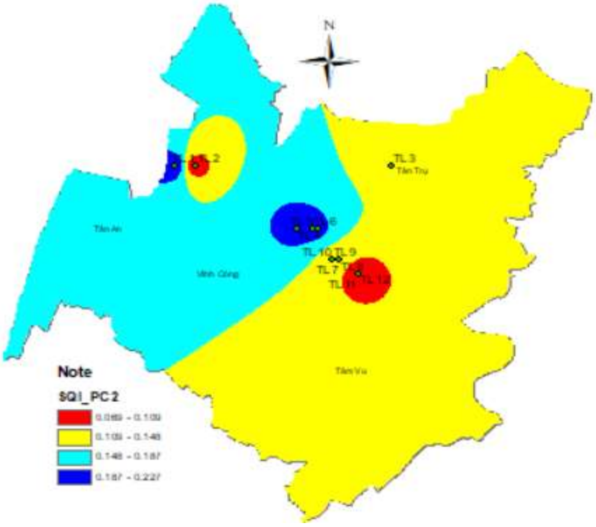


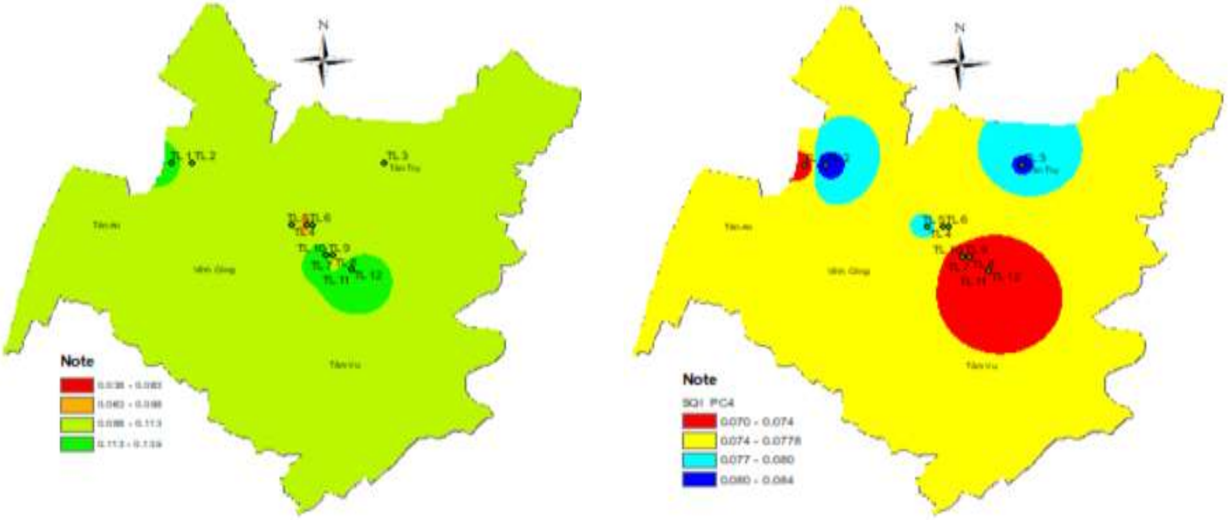
Fig. 3: SQI estimation results of soil samples by 2 methods (Additive method and PCA)



a. SQI for PC1



b. SQI for PC2



c. SQI for PC3

d. SQI for PC4

Fig. 4: SQI for each Principal Component

3.4 Mapping: Spatial Interpolation Analysis

Soil Quality Index Mapping, applying the IDW interpolation method on ArcGIS 10.8, was used to interpolate the spatial variation of soil quality in the study area based on the results of SQI estimation. Fig. 4a, SQI_PC1 showed that samples TL1 and TL2 west of the dragon fruit core farms were smaller than other farms with low clay content, and had erosion due to inappropriate irrigation and using too strong irrigation hoses (according to field observations). The SQI values for PC2 (Fig. 4b) showed that the central zone (TL11, TL12 samples) and the western zone (TL1) needed control, which was the addition of P and organic matter to increase the SQI value of the soil. Fig. 4c, the SQI value of PC3 of the central (TL10, TL11, and TL12) and western (TL1) regions of the study area is low. Therefore, activities due to excessive use of mineral fertilizers (ammonium nitrogen) need to be controlled. Instead, it is advisable to use more organic matter to control nitrogen. Finally, Fig. 4d showed that the central and western areas of the study area showed low SQI values, indicating poor soil pH control, which requires more attention.

Fig. 5 shows that the soil quality in the central growing area is mainly in the medium group (signs of degradation), while the western area has poor soil quality (TL1, TL2). The western samples are areas where farmers have recently converted to dragon fruit cultivation and are also trying to control soil quality. The central growing area has areas with poorer quality, such as sample TL6 (Fig. 5). The results show that SQI depends on soil properties, crop age, and the farmer's farming experience. The application of a soil quality index distribution map can be considered a supportive solution to help managers detect soil degradation trends early.

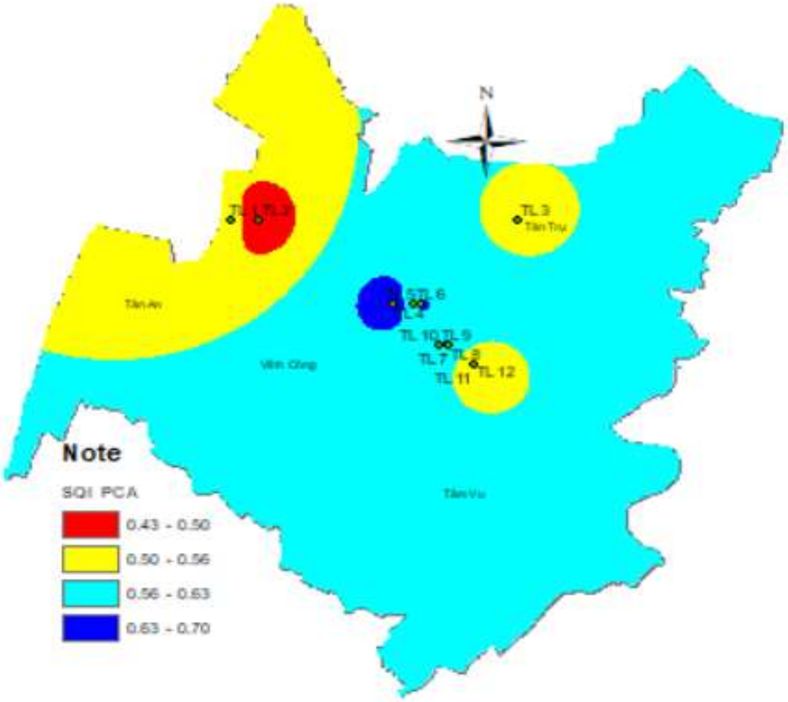


Fig. 5: Spatial distribution map of SQI values according to PCA in dragon fruit farms

4. CONCLUSIONS

Agricultural practices significantly influence soil health by affecting its physical, chemical, and fertility properties. This study employs soil physico-chemical properties for SQI assessment the dragon fruit farms in Tây Ninh, Vietnam. PCA was employed alongside GIS mapping to illustrate the spatial distribution of SQI. Minimum data sets (MDS) from PCA analysis explained 89.0% of the total variance, with the contributions of the four selected principal components being 33.2, 22.7, 19.9, and 13.2%, respectively. The representative parameters, including CEC, P_av, EC, and pH, were selected as representative components of MDS with weights of 0.37, 0.26, 0.22, and 0.15, respectively. The SQI values calculated by PCA method were determined to be better risk management. Accordingly, the average quality soil type accounted for 41.6% and the degraded type accounted for 58.3%. The results demonstrate that the combination of PCA and GIS is a more effective toolkit in detecting soil degradation trends in areas of intensive perennial crop cultivation.

ACKNOWLEDGEMENTS

This research was conducted with the support of students majoring in Land Management, Course 16. We also express our deep gratitude to the reviewers for their in-depth comments, which contributed significantly to improving the manuscript.

REFERENCES

Abdel-Fattah, M. K., Mohamed, E. S., Wagdi, E. M., Shahin, S. A., Aldosari, A. A., Lasaponara, R., & Alnaimy, M. A. (2021). Quantitative evaluation of soil quality using Principal Component Analysis: The case study of El-Fayoum depression Egypt. *Sustainability*, *13*(4), 1824.

Abdu, A., Laekemariam, F., Gidago, G., & Getaneh, L. (2023). Explaining the Soil Quality Using Different Assessment Techniques. *Applied and Environmental Soil Science*, *2023*. https://doi.org/https://doi.org/10.1155/2023/6699154

Aisah, B. N., Ikhwani, A. T., Setiawan, A. B., Purwantoro, A., Respatie, D. W., Ambarwati, E., Anggraeni, L., Teo, C. H., Setyawan, C., & Khamidah, A. (2025). Assessment of F_5 Mungbean Genotypes from Intergeneric Hybridization with Common Bean for Agronomic Performance, Heritability, Variance Components, and Genetic Uniformity. *Caraka Tani: Journal of Sustainable Agriculture*, *40*(3), 347-360. https://doi.org/https://doi.org/10.20961/carakatani.v40i3.94563

Baethgen, W., & Alley, M. (1989). A manual colorimetric procedure for measuring ammonium nitrogen in soil and plant Kjeldahl digests. *Communications In Soil Science and Plant Analysis*, *20*(9-10), 961-969.

Bandyopadhyay, S., & Maiti, S. K. (2021). Application of statistical and machine learning approach for prediction of soil quality index formulated to evaluate trajectory of ecosystem recovery in coal mine degraded land. *Ecological Engineering*, *170*, 106351. https://doi.org/https://doi.org/10.1016/j.ecoleng.2021.106351

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T. W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J. W., & Brussaard, L. (2018). Soil quality – A critical review. *Soil Biology and Biochemistry*, *120*, 105-125. https://doi.org/https://doi.org/10.1016/j.soilbio.2018.01.030

Damiba, W. A. F., Gathenya, J. M., Raude, J. M., & Home, P. G. (2024). Soil quality index (SQI) for evaluating the sustainability status of Kakia-Esamburmbur catchment under three different land use types in Narok County, Kenya. *Heliyon*, *10*(5).

George, J. (1962). Hydrometer method improved for making particle size analyses of soils. *Agron J*, *54*, 464-465.

Hanif, S., Ali, S., Chaudhry, A. H., Sial, N., Marium, A., & Mehmood, T. (2025). A Review on Soil Metal Contamination and its Environmental Implications. *Nature Environment and Pollution Technology*, *25*(1), Article ID D1684. https://doi.org/10.46488/NEPT.2025.v24i01.D1684

Kahsay, A., Haile, M., Gebresamuel, G., & Mohammed, M. (2025). Developing soil quality indices to investigate degradation impacts of different land use types in Northern Ethiopia. *Heliyon*, *11*(1).

Kawaguchi, K., & Kyuma, K. (1975). Paddy Soils in Tropical Asia: Part 3. Correlation and Regression Analyses of the Soil Data. *South East Studies*, *13*(1).

Li, Y., Zhao, H., Liu, J., Chaonan, C., & Yuxuan, G. (2024). A framework for selecting and assessing soil quality indicators for sustainable soil management in waste dumps. *Scientific Reports*, *14*(1), 8491.

Mai, T. D., Dinh, H. H. D., Ha, N. S., & Nguyen, C. K. (2025). *Expanding Export Markets for Vietnamese Dragon Fruit*. International Journal of Scientific Research and Management (IJSRM).

Mukherjee, A., & Lal, R. L. (2014). Comparison of Soil Quality Index Using Three Methods. *PLoS ONE*, *9*(8). https://doi.org/10.1371/journal.pone.0105981

Plonsky, L. (2015). Advancing quantitative methods in second language research. In: Routledge.

Rajput, R., Patley, S., Choudhary, B., & Gupta, S. (2023). Assessment and Comparison of Soil Quality Using Principal Components Analysis (PCA) & Expert Opinion (EO) Methods in different Rice-based Cropping Systems in Alfisol.

Rangel-Peraza, J. G., Padilla-Gasca, E., López-Corrales, R., Medina, J. R., Bustos-Terrones, Y., Amabilis-Sosa, L. E., Rodríguez-Mata, A. E., & Osuna-Enciso, T. (2017). Robust Soil Quality Index for Tropical Soils Influenced by Agricultural Activities. *Journal of Agricultural Chemistry and Environment*, *6*, 199-221.

Țopa, D.-C., Căpșună, S., Calistru, A.-E., & Ailincăi, C. (2025). Sustainable Practices for Enhancing Soil Health and Crop Quality in Modern Agriculture: A Review. *Agriculture; Basel*, *15*(9).

Van, P. N. (2025). Assessing Soil Health Through Multivariate Analysis: A Focus On Durian Cultivation In Cho Lach, Ben Tre Province, Vietnam.