Eurasian Journal of Soil Science 2026, 15(2), 291 - 301

Eurasian Journal of Soil Science

Journal homepage : http://ejss.fesss.org

Evaluation of soil quality in a cornell-based polye ecosystem in a Karstic Area: Integration of geostatistics and machine learning

Hüseyin Alper Aktop [a], Sevda Altunbaş Beritanli [a,*], Bayram Çağdaş Demirel [b]

[a] Akdeniz University, Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Antalya, Türkiye

[b] Akdeniz University, Vocational School of Technical Sciences, Antalya, Türkiye

Article Info

Received : 21.10.2025

Accepted : 23.02.2026

Available online: 03.03.2026

Author(s)

H.A.Aktop

S.A.Beritanli *

B.Ç.Demirel

* Corresponding author

Abstract

Assessing soil quality and establishing processes to control quality is critically important in environmentally sensitive karst systems. This study aimed to analyze the spatial behavior of soil quality and the dominant soil properties controlling it in the karstic polje system of Gembos Plain (Southern Turkey) using the Cornell Soil Quality Index (SQI). SQI values were calculated using physical and chemical data obtained from 72 soil samples (0–30 cm). According to analysis results SQI values ranged from 53 to 59, with an average value of approximately 56. Semivariogram analysis showed that SQI exhibited moderate spatial dependence; the Gaussian variogram model was selected as the most appropriate model due to lowest error values and RMSSE close to 1. The spatial distribution obtained by ordinary kriging revealed that low-to-medium quality classes were dominant across the area and that geochemical micro-scale patterns prevailed over distinct morphological zoning. In addition, Random Forest and XGBoost models showed the highest performance in machine learning analyses (R^2 = 0.74 and 0.71). Variable importance analysis has shown that soil reaction (pH) is the primary control factor determining SQI variability, while P, Fe, and K are effective at a secondary level. Also soil organic matter plays a secondary role in explaining SQI variability due to its narrow range of variation. The results reveal that soil quality in karstic polje conditions is controlled by chemical constraints rather than physical properties, and that the combined use of geostatistical modeling and machine learning significantly improves the interpretability of soil quality analyses.

Keywords: Soil quality, Karstic polje, Cornell Soil Quality Index, Geostatistical modeling, Machine learning.

© 2026 Federation of Eurasian Soil Science Societies.

Introduction

Soil quality is a comprehensive concept that refers to a soil's capacity to sustain biological productivity in natural or managed ecosystems, preserve environmental quality, and support plant and animal health (Karlen et al., 2008; Bünemann et al., 2018; Lehmann et al., 2020). This concept is not limited to the evaluation of individual soil properties; it requires a holistic approach to physical, chemical, and biological indicators (Bünemann et al., 2018; Yertayeva et al., 2019; Heepngoen et al., 2021; Adyanova et al., 2023). Soil Quality Indices (SQI) developed within this scope combine multidimensional soil properties under a single numerical value, providing important tools for monitoring ecosystem functionality and supporting sustainable land management decisions (Raiesi and Salek-Gilani, 2020; Demirağ Turan, 2021). In addition, Demir and Gülser (2015) indicated that improving and sustaining of soil quality reduces fertilizer and pesticide use, improves the air and water quality and helps to prevent the release of greenhouse gases into the atmosphere as well as increased the agricultural productivity of country.

Among these approaches, the Cornell Agricultural Health Score (CASH) is widely used due to its indicator-based structure and adaptability to different ecosystem types (Fine et al., 2017). The CASH-based SQI

: https://doi.org/10.18393/ejss.1901800

: https://ejss.fesss.org/10.18393/ejss.1901800

Publisher : Federation of Eurasian Soil Science Societies

e-ISSN : 2147-4249

methodology stands out by focusing not only on the nutrient content of soils but also on their functional properties such as hydraulic conductivity, aggregate stability, and biological activity. However, the sensitivity of these indices to local environmental conditions directly affects the representativeness and interpretability of the index (Amsili et al., 2023). Karstic polyhedrons are special ecosystems with unique environmental dynamics in terms of soil quality assessments.

These areas are defined as physiographies characterized by closed drainage systems, high calcium carbonate content, alkaline reactions, and pronounced lithology–hydrogeology–soil formation interactions (Gutiérrez and Gutiérrez, 2016; De Waele and Gutiérrez, 2022). In these environments, limited topographic variability causes soil properties to be controlled more by chemical and hydraulic processes. Therefore, examining how SQI approaches developed in different ecosystems behave in karst systems is of particular importance in terms of the index's sensitivity and the interactions between indicators (Yu et al., 2023).

In the context of spatial analysis, geostatistical methods (e.g., kriging) are widely used to convert soil quality indices into continuous surfaces. However, these methods do not provide direct information on which soil properties control the resulting spatial patterns (Sanad et al., 2024). This shortcoming makes it difficult to understand the internal behavior of the index, especially in heterogeneous ecosystems.

At this point, machine learning (ML) approaches stand out as a powerful complement. ML algorithms provide effective tools for modeling complex and nonlinear relationships, determining variable importance levels, and identifying the dominant factors controlling soil quality (Bandyopadhyay and Maiti, 2021; Dengiz et al., 2023). The combination of geostatistical methods' spatial mapping power and machine learning's capacity to explain variable dependencies adds a quantitative and interpretative dimension to soil quality assessments (De Caires et al., 2025).

In this context, the main objectives of this study are as follows: (i) to determine the spatial distribution of a Cornell-based soil quality index in a karstic polje environment using geostatistical methods; (ii) to evaluate the reliability of the created spatial model through cross-validation; (iii) to identify the dominant soil properties controlling the variability of the soil quality index using machine learning algorithms.

This study evaluates the spatial behavior of a Cornell-based soil quality index under karstic polje conditions and goes beyond spatial mapping to question the role of soil chemical constraints—particularly pH-related processes—underlying variability in soil quality in these environments where organic matter inputs and topographic slopes are naturally limited.

Material and Methods

Study Area

This study was conducted in the Gembos Plain, located in southern Türkiye. The plain exhibits typical karstic polje characteristics; it has a closed or semi-closed drainage system and a low-sloped (0–2%) depression morphology. This morphology creates unique environmental conditions where surface drainage is limited and soil formation is primarily determined by lithology, groundwater dynamics, and the calcium carbonate cycle (Gökkaya and Gutiérrez, 2022). The limited topographic variability causes soil properties in the Gembos Plain to be controlled more by chemical and hydraulic processes.

The area is characterized by a Mediterranean climatic regime with cold, wet winters and dry summers. Based on data obtained from the Turkish State Meteorological Service (Derebucak station), the mean annual precipitation is approximately 800 mm, with the majority occurring between November and March, while summer months receive minimal rainfall. The mean annual air temperature is approximately 11–12 °C, with average monthly temperatures ranging from near 0 °C in winter to over 20 °C in late summer.

Land use within the polje is predominantly agricultural. Crop production is dominated by dry bean (*Phaseolus vulgaris*), wheat (*Triticum* spp.), silage maize, sunflower, sugar beet, and strawberry cultivation, reflecting intensive seasonal cropping systems on the polje floor.

Geologically, the Gembos Polje developed within neritic limestones and rudist-bearing carbonate formations and represents a structurally controlled polje bounded by normal faults. Borehole data indicate approximately 11 m of alluvial deposits overlying flysch units, followed by limestone bedrock (Gökkaya, 2016). The general location and physiography of the study area are shown in Figure 1.

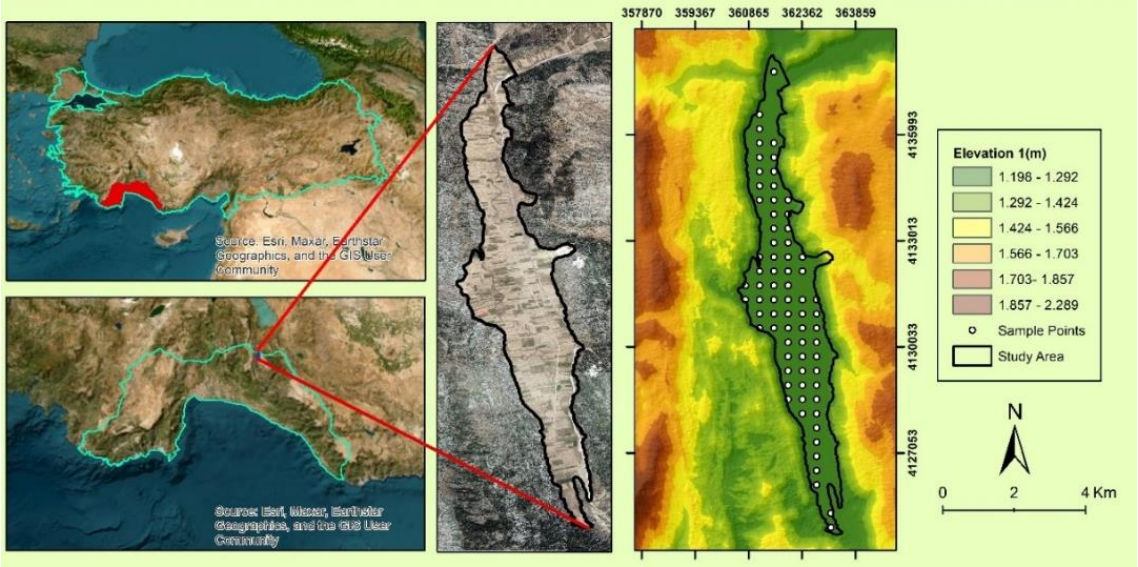


Figure 1. Location of the study area in Türkiye and the Gembos Plain (karst polje), showing the study boundary and the spatial distribution of soil sampling points over the elevation (DEM) map

Soil Sampling and Laboratory Analyses

Soil sampling was carried out according to a systematic grid pattern established at approximately 400-meter intervals across the study area. To represent the areas under agricultural production, disturbed samples representing the surface soil at a depth of 0–30 cm were collected from a total of 72 sampling points. Standard soil science methods were followed in laboratory analyses. The procedures used in the literature to determine the physical, and chemical properties of soils are shown in Table 1.

Table 1. Analytical methods and references used for determining soil physical and chemical properties (0–30 cm).

Parameters	Protocol	Reference
pH	(w:v) soil–water suspension	Soil Survey Staff (1992)
EC	(w:v) soil–water suspension	Soil Survey Staff (1992)
Texture	Hydrometer method	Bouyoucos (1951)
Lime	Scheibler calsimeter	Soil Survey Staff (1992)
DTPA–Fe, Cu, Mn, Zn	DTPA extraction	Lindsay and Norvell (1978)
NH4OAC–K, Ca, Mg, Na	Ammonium acetate extraction	Soil Survey Staff (1992)
Total N	Kjeldahl	Bremner and Mulvaney (1982)
OM	Walkley-Black with $K_2Cr_2O_7$	Nelson and Sommers (1982)
$NaHCO_3$–P	The molybdophosphoricblue method	Olsen et al. (1954)

OM: Organic matter, EC: Electrical conductivity

Prior to geostatistical and machine learning analyses, descriptive statistical analyses were performed to characterize the distribution of measured soil properties. Minimum, maximum, mean, standard deviation, and coefficient of variation (CV) values were calculated using R software (version 3.40.8). These preliminary analyses were conducted to evaluate data structure and variability patterns before subsequent spatial and modeling procedures.

Cornell-Based Soil Quality Index (SQI)

Soil quality was assessed using the Soil Quality Index (SQI), calculated based on the Cornell Soil Health Assessment (CASH) approach. This approach is based on converting different physical, chemical, and biological indicators into dimensionless values through scoring functions that reflect soil functionality (Moebius-Clune et al., 2016). Each indicator has been evaluated according to the principles of "more is better," "less is better," or "optimal range." For example, higher values for organic matter and total nitrogen represent better soil quality, while lower values for parameters such as salinity and sodium content are considered advantageous. Indicators such as pH and C/N ratio, on the other hand, have been evaluated within an optimal range.

In this study, while the standard framework of the Cornell approach was maintained, only reliable and directly measured physical and chemical indicators were included in the index, based on the availability of data for the study area. Some physical indicators, such as aggregate stability (WSA) and available water capacity (AWC), were not included in the SQI calculation because they were not directly measured in the study area. This choice

aims to avoid the use of hypothetical or indirect parameters in the calculation of the index and to interpret the results using a data-driven approach.

After scoring, all indicators were combined using an equal weighting approach, yielding continuous SQI values ranging from 0 to 100 (Moebius-Clune et al., 2016; Fine et al., 2017). The equal weighting approach avoided defining a prior importance hierarchy among the indicators, allowing the relative impact of each parameter on the index to be evaluated later through statistical and machine learning-based analyses (Becker et al., 2017). These index values were used as continuous data in statistical analyses and modeling, and were divided into "low" and "medium" qualitative classes for mapping and visualization purposes (Gugino et al., 2009). All indicators used in the SQI calculation, the type of scoring functions applied, and the scoring ranges are summarized in Table 2.

Table 2. Indicators, units, scoring direction, and scoring functions used to compute the Cornell-based Soil Quality Index (SQI)

Indicator Group	Indicator	Unit	Scoring Direction	Scoring Principle	Soil Function
Physical	Soil Texture	%	Optimum	Soil structure becomes more suitable for root development as texture approaches the optimum	Water retention
Chemical	Soil Reaction (pH)	–	Optimum	Nutrient availability increases between pH 6.5–7.5; below or above this range availability decreases	Nutrient availability
Chemical	Electrical Conductivity (EC)	dS m^{-1}	Less is better	Excessive salt content reduces plant growth	Salt stress
Chemical	Organic Matter	%	More is better	Higher organic matter content increases soil quality	Nutrient cycling
Chemical	Total Nitrogen	%	More is better	Higher nitrogen content increases soil quality	Nutrient supply
Chemical	Available Phosphorus (Olsen P)	mg kg^{-1}	Optimum	Phosphorus deficiency or excess negatively affects plant growth	Energy transfer
Chemical	Exchangeable Potassium (K)	mg kg^{-1}	Optimum	Potassium deficiency or excess reduces nutrient balance	Nutrient balance
Chemical (Restrictive)	Lime ($CaCO_3$)	%	Less is better	Excessive lime content reduces nutrient availability	Nutrient retention
Chemical (Restrictive)	Exchangeable Sodium Percentage (ESP)	%	Less is better	High ESP values deteriorate soil structure	Structural degradation

Geostatistical Analysis

The spatial variability of the soil quality index (SQI) was evaluated based on geostatistical variogram analysis. In this context, experimental semi-variograms were created using point-based SQI values, and the spatial dependency structure of the data was examined (Sanad et al., 2024). Prior to the analysis, the distribution characteristics of the SQI data were checked, and the disruptive effect of extreme values on the spatial dependency structure was evaluated.

Different theoretical models such as Spherical, Exponential, and Gaussian were applied to experimental semivariograms used in ordinary kriging. The suitability of the models was compared by considering the combination of variogram parameters, namely range, sill, and nugget values. The most suitable model was selected as the one that best represented the spatial continuity of the data and provided the lowest error values in terms of cross-validation criteria.

Based on the selected variogram model, spatial interpolation of SQI values was performed using the ordinary kriging method. The kriging method allows SQI values to be estimated at unobserved points by taking into account the spatial correlation between measurement points. This approach was used to identify spatial patterns, and the resulting surfaces were evaluated as a continuous spatial representation of measured data rather than a statistical estimation model.

The model's performance was evaluated using the cross-validation procedure available in the ArcGIS Geostatistical Analyst module. In this context, the differences between observed and predicted SQI values were analyzed; model accuracy was interpreted using the mean error (ME), root mean square error (RMSE), and standardized RMSE (RMSSE) metrics. The statistics obtained showed that the selected variogram model adequately represented spatial dependence and that prediction uncertainties were balanced (Webster and Oliver, 2007; Sanad et al., 2024).

Machine Learning Analysis

Machine learning (ML) analysis was applied in this study not to directly estimate the Cornell-based soil quality index (SQI), but rather to determine the dominant soil properties controlling the variability of the index. In this context, the calculated SQI values were defined as the dependent variable, while the physical and chemical soil properties obtained from laboratory analyses were used as independent variables.

To protect data integrity and prevent model bias, the sub-score components directly used in the Cornell SQI calculation (e.g., pH score, organic matter score, and similar scoring variables) have been deliberately excluded from the inputs of machine learning models (Zeraatpisheh et al., 2019). This approach was adopted to prevent data leakage and to prevent the model from establishing artificial relationships that replicate the calculation logic of the SQI (Kapoor and Narayanan, 2023). Instead, other soil properties representing the chemical and physical environment of the soil were included in the model as independent variables.

Prior to modeling, the dataset was examined and samples containing missing observations were excluded from the analysis. To prevent differences in variable magnitudes from affecting model performance, scale-sensitive methods such as Support Vector Regression and Artificial Neural Networks were applied to centered and scaled data using z-score standardization. Tree-based Random Forest and XGBoost algorithms were trained on raw data as they are scale-independent.

Based on the assumption that the relationships between soil properties and the soil quality index may have a non-linear structure, ensemble-based machine learning algorithms were primarily preferred in the analyses. In this context, the Random Forest regression model was used as the primary method; Multiple Linear Regression, Support Vector Regression, and Artificial Neural Network models were also included in the analyses for comparison purposes.

The generalizability of the models was evaluated using a k-fold cross-validation approach applied to the training data; the final model performances were compared using the R^2, RMSE, and MAE metrics calculated on an independent test dataset. The variable importance analysis performed on the best-performing model enabled the identification of dominant soil properties controlling SQI variability. These analyses clearly demonstrate that the machine learning approach was used in this study not as a prediction tool, but as a complementary analytical tool explaining the behavior of the Cornell-based soil quality index.

Results

Descriptive Statistics of Soil Properties

The descriptive statistics of the physical and chemical properties of the 72 surface soil samples summarized in Table 3. The textural composition of the examined soils is predominantly concentrated in the loamy and clay loam classes, with the variation ranges of the sand, silt, and clay fractions exhibiting a relatively narrow distribution. This situation indicates a texturally homogeneous soil structure consistent with accumulation processes developing under low topographic energy conditions on the karstic polje base.

Soil reaction (pH) values showed alkaline characteristics across the entire sampling area; the narrow range of pH distribution revealed the dominant effect of the calcareous structure of the parent material on the chemical environment. Electrical conductivity (EC) values were generally low, indicating that there was no significant salinity problem in the study area.

Calcium carbonate (CaCO3) contents exhibit a wide distribution range, but the high average values reflect the direct influence of karstic parent material. Organic matter (OM) contents ranged from 0.78 to 3.05%, with a mean value of 1.56%. According to commonly applied soil organic matter classification thresholds (<1% very low, 1–2% low, 2–4% moderate), the soils generally correspond to the low class, with some samples approaching moderate levels (Hazelton and Murphy, 2007). Descriptive statistics for macro (P, K, Ca, Mg) and micronutrients (Fe, Mn, Zn, Cu) indicate that variability in soil properties within the study area is largely related to chemical processes. These findings indicate that soil formation and differentiation under karstic polje conditions is controlled by geochemical equilibrium processes rather than topographic factors.

Cornell-Based Soil Quality Index (SQI) and Spatial Distribution

Soil Quality Index (SQI) values calculated based on the Cornell Soil Health Assessment approach ranged from 53 to 59 in the study area, with an average value of approximately 56. These results indicate that soils of the Gembos polje generally fall within the low to moderate soil quality class. The narrow range of SQI values is related to the low topographic slope (0–2%) across the area and the largely homogeneous physical environmental conditions.

Table 3. Descriptive statistics of measured soil properties from 72 surface soil samples (0–30 cm) in the Gembos Plain

Variable	Min	Max	Average	Std. Deviation	CV (%)
Sand (%)	1.00	71.00	29.19	14.13	48.39
Clay (%)	5.00	65.00	30.83	11.05	35.82
Silt (%)	14.00	58.00	39.83	7.71	19.36
pH	6.60	7.99	7.33	0.32	4.34
EC (dS m^{-1})	0.03	0.60	0.13	0.08	61.01
$CaCO_3$ (%)	1.50	16.80	3.03	2.13	70.25
OM (%)	0.78	3.05	1.56	0.42	26.91
Total N (%)	0.04	0.15	0.08	0.02	26.74
P (mg kg^{-1})	2.19	42.61	9.25	5.93	64.08
K (mg kg^{-1})	57.57	395.00	116.06	51.43	44.31
Ca (mg kg^{-1})	1520	6539	4264.79	1032.21	24.20
Mg (mg kg^{-1})	222	1413	590.76	214.74	36.35
Na (mg kg^{-1})	19.93	114.00	38.21	10.73	28.08
Fe (mg kg^{-1})	8.04	37.91	17.33	6.26	36.14
Mn (mg kg^{-1})	3.77	19.58	9.17	2.94	32.11
Zn (mg kg^{-1})	0.46	3.05	1.22	0.56	45.83
Cu (mg kg^{-1})	1.82	4.58	2.71	0.53	19.45
Exchangeable Na (%)	0.36	1.19	0.64	0.15	23.27

The experimental semivariogram analysis of the soil quality index revealed that SQI values exhibit moderate spatial dependence (Table 4). The calculation of the nugget/sill ratio as 0.42 indicates that SQI variability is influenced by both local environmental conditions and larger-scale geochemical processes. Different theoretical semivariogram models (Gaussian, exponential, and spherical) were tested; the Gaussian model was selected as the model that best represented the data structure due to its lower error values and standardized root mean square error (RMSSE) value close to 1.

Table 4. Parameters of the fitted Gaussian semivariogram model and cross-validation statistics for ordinary kriging of SQI

Parameter	Model	Sample (n)	Nugget/sill	Range (m)	RMSE	ME	RMSSE
Value	Gaussian	72	0.42	2770	4.43	0.01	0.99

The SQI distribution map obtained as a result of spatial interpolation performed using the ordinary kriging method based on the selected Gaussian variogram model (Figure 2) shows that low and medium quality classes are dominant across the area. The fact that the mean error is close to zero (ME ≈ 0) and the standardized root mean square error (RMSSE ≈ 1) is close to unity in the cross-validation results indicates that the model does not contain systematic bias and that estimation uncertainties are appropriately represented (Webster and Oliver, 2007; Şenol et al., 2020).

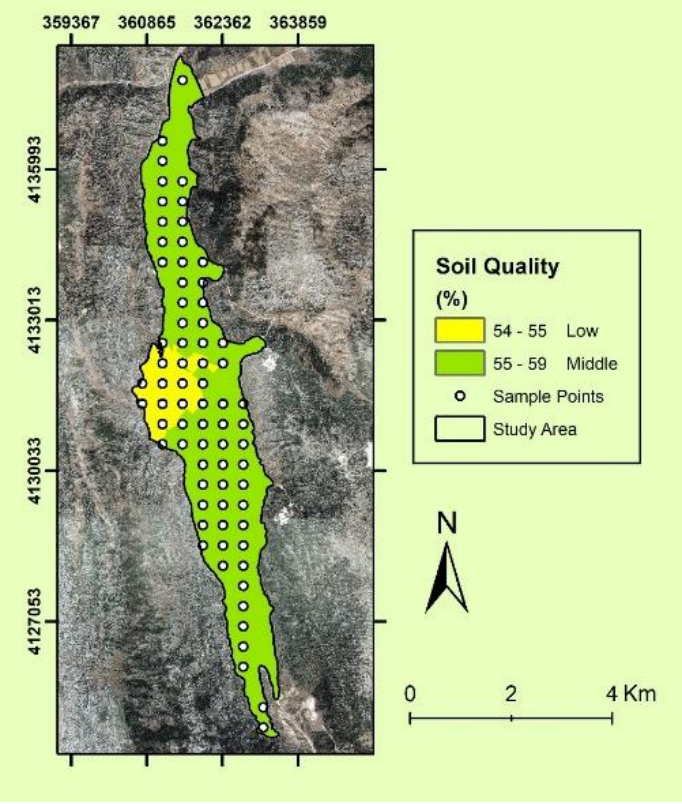


Figure 2. Ordinary kriging prediction map of the soil quality index (SQI) using the Gaussian variogram model, showing the spatial distribution of SQI classes (Low: 54–55%; Middle: 55–59%) and sampling locations within the study area

The spatial pattern obtained indicates that, rather than distinct morphological gradients within the SQI study area, microzones shaped by geochemical processes are distributed. Relatively lower SQI values were observed in the central sections of the polje base, while higher index values were determined in the peripheral areas due to more stable chemical conditions. This situation reveals that soil quality in karstic polje conditions is controlled by chemical limitations rather than physical heterogeneity.

Machine Learning Model Performance

The performance of machine learning algorithms used to model the soil quality index (SQI) based on raw soil properties was evaluated using the determination coefficient (R^2), root mean square error (RMSE), and mean absolute error (MAE) metrics. The comparative performance results of the models are presented in Table 5.

Table 5. Predictive performance (R^2, RMSE, MAE) of machine-learning models used to explain SQI variability

Model	R^2	RMSE	MAE
Random Forest Regressor (RF)	0,737	2,87	2,27
XGBoost Regressor	0,711	2,89	2,95
Support Vector Regression (SVR)	0,434	4,10	2,95
Artificial Neural Network (ANN)	0,167	6,42	4,76
Multiple Linear Regression (MLR)	0,103	4,81	3,49

According to the analysis results, the Random Forest (RF) model emerged as the algorithm with the highest explanatory power (R^2 = 0.737), followed by the XGBoost model (R^2 = 0.711). Both models produced lower error values compared to other methods and represented SQI variability with higher accuracy. The Support Vector Regression (SVR) and Artificial Neural Network (ANN) models showed moderate to low performance, while the Multiple Linear Regression (MLR) model had the lowest explanatory power.

The fact that ensemble-based models produce more successful results than linear regression approaches indicates that the relationship between SQI and soil properties exhibits a non-linear and multivariate structure. Therefore, variable importance analyses and advanced evaluations were performed using the Random Forest and XGBoost models, which showed the highest performance.

Machine Learning-Based Variable Importance Analysis

Within the scope of machine learning analyses applied to determine the dominant soil properties controlling the variability of the soil quality index (SQI), variable importance analysis was performed using the Random Forest model, which demonstrated the highest performance. The contribution of the variables to the model was evaluated using the %IncMSE criterion, which is based on the increase in prediction error.

The results revealed that soil pH was by far the most dominant variable controlling SQI variability (Figure 3). Excluding the pH variable from the model resulted in a significant increase in prediction error, indicating that SQI in the study area was largely driven by soil reaction. pH was followed by phosphorus (P), iron (Fe), potassium (K), and calcium carbonate ($CaCO_3$) contents. Although these variables had a more limited impact compared to pH, they provided secondary-level meaningful contributions to SQI.

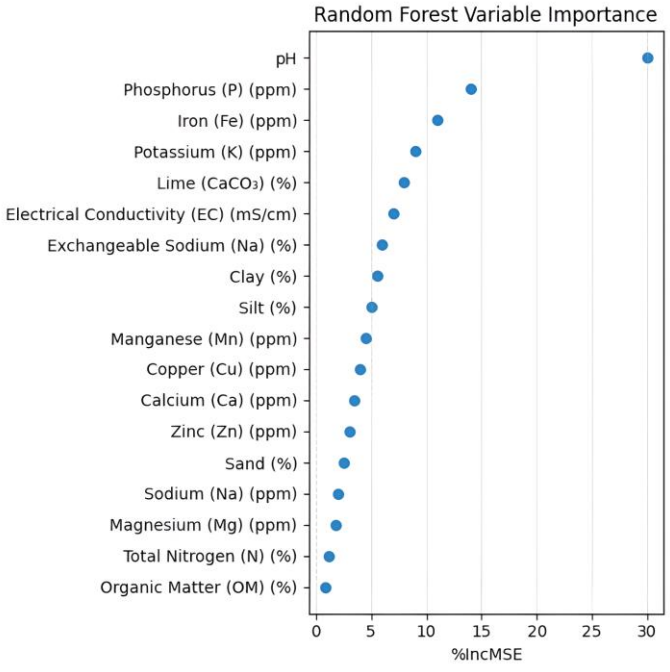


Figure 3. Relative importance of soil properties in explaining the variability of the Cornellbased soil quality index (SQI), as determined by the Random Forest model (%IncMSE)

The relative importance of parameters such as textural components (sand, silt, clay), electrical conductivity (EC), and exchangeable sodium remained low; organic matter (OM) and total nitrogen (N) variables contributed only marginally to the explanatory power of the model. These results indicate that SQI variability in the study area is primarily controlled by variables representing chemical equilibrium and ionic environment, with biological inputs playing a secondary role.

The variable importance rankings for the Random Forest model are presented in Figure 3. Comparative results regarding the SQI prediction performance of different machine learning models are summarized in Table 5. These findings in the study demonstrate that the machine learning approach offers an explanatory analytical framework that quantitatively reveals the environmental sensitivity of the Cornell-based soil quality index, beyond being merely a prediction tool.

Discussion

The Dominant Role of pH and Karstic Polje Conditions

In this study, machine learning-based variable importance analysis clearly demonstrated that soil pH is the most dominant factor controlling the variability of the Cornell-based soil quality index (SQI). This result is directly related to the Gembos Polje being a karstic system developed on carbonate parent material. In such environments where carbonate buffering is strong, pH is central to soil chemical equilibrium and indirectly controls the bioavailability of nutrients in particular (Yu et al., 2023).

Under alkaline conditions, the fixation of phosphorus in calcium phosphate forms and the reduced solubility of micronutrients (particularly Fe, Zn, and Cu) emerge as the primary chemical mechanisms limiting SQI scores (Alloway, 2008; Prasad and Shivay, 2021). In this context, the fact that even relatively small changes in pH have significant effects on SQI indicates that chemical thresholds are critical for soil functions in karstic polygenetic environments. Therefore, the dominance of pH is not merely a statistical finding but an environmental outcome consistent with the pedochemical nature of karstic systems.

The Secondary Role of Organic Matter and Preventing Misinterpretation Risks

Although soil organic matter (OM) is generally accepted as a fundamental indicator in determining soil quality, in this study it contributed at a secondary level in variable importance analyses. This does not imply that organic matter is insignificant; rather, it is explained by the fact that OM contents in the study area varied within a narrow range. It is methodologically expected that a variable with low variance will show limited explanatory power in machine learning models (Hastie et al., 2009; Kuhn and Johnson, 2013; Panigrahi et al., 2025).

The generally low to moderate and spatially homogeneous distribution of organic matter contents in the Gembos Polje resulted in SQI variability being controlled not by organic matter, but by indicators representing chemical equilibrium. Given its relatively narrow range of variation in the study area, organic matter (OM) appears to function as a baseline soil quality component influencing the overall SQI level, rather than as a primary driver of its spatial variability, a pattern consistent with the statistical behavior of low-variance predictors in multivariate models (Hastie et al., 2009; Bünemann et al., 2018). Making this distinction clearly prevents machine learning results from being misinterpreted as indicating that organic matter is insignificant.

Joint Interpretation of Geostatistical Patterns and Machine Learning Findings

Geostatistical analyses have shown that SQI values exhibit weak to moderate spatial continuity across the area and have a low-contrast distribution structure. Spatial patterns obtained through Kriging interpolation revealed that SQI is distributed in the form of geochemically defined microzones rather than distinct morphological zones. However, this spatial representation is limited in explaining why these patterns occur.

At this stage, machine learning analysis has served as an explanatory tool complementing geostatistical results. While Kriging describes the spatial pattern, machine learning quantitatively identified the dominant variables controlling this pattern. The combined use of both approaches enabled a comprehensive assessment of both “where” and “why” soil quality varies. This integration significantly increases the interpretability of soil quality analyses, particularly in karstic polje environments where physical heterogeneity is limited (Wang et al., 2025; Qiu et al., 2025).

Methodological Evaluation and Limitations of the Study

The methodological approach adopted in this study is based on the evaluation of a Cornell-based soil quality index using geostatistical and machine learning methods. However, several limitations should be considered. First, sampling was limited to surface soil (0–30 cm) and variability in the vertical profile was not assessed.

This situation has resulted in chemical processes that can vary with depth, particularly in karst systems, not being reflected in the analysis (Qin et al., 2022).

Additionally, machine learning analyses were applied on a point-by-point basis, and spatial dependency structures were not directly integrated into the models. Therefore, spatial patterns were evaluated separately using geostatistical methods, while variable relationships were assessed using machine learning. Nevertheless, this approach clearly demonstrates that machine learning is used as an explanatory rather than a causal tool. Future studies incorporating different depth layers, temporal variability, and spatially aware machine learning models (spatial ML) would contribute to a more comprehensive assessment of soil quality in karstic polje environments.

Conclusion

This study evaluated the spatial behavior of the Cornell-based soil quality index (SQI) in the Gembos Plain, a karstic polje system, and the fundamental soil properties controlling this behavior using a combination of geostatistical and machine learning approaches. The results indicated that the soils in the study area generally exhibited medium soil quality and that index variability was largely controlled by chemical constraints.

Machine learning-based analyses have clearly demonstrated that soil reaction (pH) is the primary control factor determining SQI variability. This result is consistent with the decisive role of chemical buffering processes on soil functions in karstic polje conditions dominated by carbonate parent material. The indirect control of pH over the bioavailability of nutrients has created a dominant effect on SQI scores, particularly through phosphorus and micronutrients. Therefore, the assessment and management of soil quality in this environment requires a focus on chemical balance rather than physical indicators.

Although soil organic matter (OM) is an important component that limits the overall level of the index, it has played a secondary role in explaining SQI variability due to the narrow range of variation in the study area. This situation does not imply that organic matter is insignificant; on the contrary, it shows that OM functions as a factor that determines the background conditions more in karstic polje environments. In this context, while organic matter management is necessary in practices aimed at improving soil quality, it is not sufficient on its own and must be addressed in conjunction with approaches that target chemical balance.

Geostatistical analyses have shown that the SQI exhibits weak to moderate spatial continuity across the area and has a low-contrast distribution structure. While the ordinary kriging method reliably represents the spatial pattern of the index, the underlying control mechanisms of these patterns have been quantitatively revealed through machine learning analysis. This demonstrates that geostatistical methods answer the "where" question, while machine learning answers the "why" question, and that using both approaches together significantly increases interpretive power.

In this study, machine learning methods are positioned not as a prediction tool but as an explanatory tool that analyzes the environmental sensitivity and internal behavior of the soil quality index. The high and consistent performance of the Random Forest and XGBoost models revealed that the relationships between SQI and soil properties have a nonlinear and multivariate structure. This finding indicates that classical statistical approaches alone may not be sufficient in index-based soil quality assessments.

In conclusion, the Gembos Plain example demonstrates that soil quality in karstic polje environments is largely determined by chemical constraints. It shows that the combined use of geostatistical and machine learning approaches provides an important methodological advantage for quantitatively identifying these constraints. Rather than proposing a new index, the study demonstrates how an existing index can be analyzed in an environmentally sensitive manner, offering a more cautious and analytical framework for soil quality assessments in karstic systems.

References

Adyanova, A., Buluktaev, A., Mukabenova, R., Mandzhieva, S., Rajput, V., Sayanov, V., Djimbee,N., Sushkova, S., 2023. Characterization of arid soil quality: Physical and chemical parameters. *Eurasian Journal of Soil Science* 12(2): 151-158.

Alloway, B.J., 2008. Micronutrients and Crop Production: An Introduction. In: Micronutrient Deficiencies in Global Crop Production. Alloway, B.J. (Ed.). Springer, Dordrecht, pp. 1–39.

Amoakwah, E., Rahman, M.A., Nketia, K.A., Djouaka, R., Didenko, N.O., Islam, K.R., 2021. Impact of deforestation and subsequent land-use change on soil quality. *Eurasian Journal of Soil Science* 10(2): 150-160.

Amsili, J.P., van Es, H.M., Aller, D.M., Schindelbeck, R.R., 2023. Empirical approach for developing production environment soil health benchmarks. *Geoderma Regional* 32: e00688.

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* 159: 106124.

Becker, W., Saisana, M., Paruolo, P., Vandecasteele, I., 2017. Weights and importance in composite indicators: Closing the gap. *Ecological Indicators* 80: 12–22.

Bouyoucos, G.J., 1951. A recalibration of the hydrometer method for making mechanical analysis of soils. *Agronomy Journal* 43(9): 434–438.

Bremner, J.M., Mulvaney, C.S., 1982. Nitrogen-Total. In: Methods of soil analysis. Part 2. Chemical and microbiological properties. Page, A.L., Miller, R.H., Keeney, D.R., (Eds.). American Society of Agronomy, Soil Science Society of America, Madison, Wisconsin, USA. pp. 595-624.

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.

De Caires, S.A., Martin, C.S., Atwell, M.A., Kaya, F., 2025. Advancing soil mapping and management using geostatistics and integrated machine learning and remote sensing techniques: A synoptic review. *Discover Soil* 5: 82.

De Waele, J., Gutiérrez, F., 2022. Karst Hydrogeology, Geomorphology and Caves. John Wiley & Sons Ltd. 912p.

Demir, Z., Gülser, C. 2015. Effects of rice husk compost application on soil quality parameters in greenhouse conditions. *Eurasian Journal of Soil Science* 4(3): 185-190.

Demirağ Turan, İ., 2021. Spatio analysis of soil quality assessment in semi-arid ecosystem using a minimum data set. *Eurasian Journal of Soil Science* 10(3): 222-235.

Dengiz, O., Alaboz, P., Saygın, F., Adem, K., Yüksek, E., 2024. Evaluation of soil quality of cultivated lands with classification and regression-based machine learning algorithms optimization under humid environmental condition. *Advances in Space Research* 74: 5514–5529.

Fine, A.K., van Es, H.M., Schindelbeck, R.R., 2017. Statistics, scoring functions, and regional analysis of a comprehensive soil health database. *Soil Science Society of America Journal* 81: 589–601.

Gökkaya, E., 2016. Geomorphological evolution of the upper Manavgat River basin: Geomorphology of the Kembos and Eynif poljes. Unpublished M.Sc. Thesis, Ankara University, Graduate School of Social Sciences, Ankara, Türkiye.

Gökkaya, G., Gutiérrez, F., 2022. Poljes in the Sivas gypsum karst, Turkey. *Geomorphology* 398: 108024.

Gugino, B.K., Abawi, G.S., Idowu, O.J., Schindelbeck, R.R., Smith, L.L., Thies, J.E., Wolfe, D.W., van Es, H.M., 2009. Cornell Soil Health Assessment Training Manual. Cornell University College of Agriculture and Life Sciences, Ithaca, NY. 58p.

Gutiérrez, F., Gutiérrez, M., 2016. Karst landforms. In: Landforms of the Earth. Gutiérrez, F., Gutiérrez, M. (Eds.). Springer, Cham, pp. 63–86.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer, New York, USA. 745p.

Hazelton, P., Murphy, B., 2007. Interpreting soil test results. What do the numbers mean? CSIRO Publishing, Melbourne. Australia. 152p.

Heepngoen, P., Thoumazeau, A., Renevier, M.S., Sajjaphan, K., Gay, F., Brauman, A., 2021. Relationships between physico-chemical, biological and functional approaches for soil quality assessment: A case study along a gradient of disturbance. *European Journal of Soil Biology* 104: 103300.

Kapoor, S., Narayanan, A., 2023. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* 4: 100804.

Karlen, D.L., Andrews, S.S., Wienhold, B.J., Zobeck, T.M., 2008. Soil quality assessment: Past, present and future. *Journal of Integrative Biosciences* 6: 3–14.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York, USA. 600p.

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020. The concept and future prospects of soil health. *Nature Reviews Earth & Environment* 1: 544–553.

Lindsay, W.L., Norvell, W.A., 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. *Soil Science Society of America Journal* 42(3): 421–428.

Moebius-Clune, B.N., Moebius-Clune, D.J., Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., Ristow, A.J., van Es, H.M., 2016. Comprehensive assessment of soil health – The Cornell framework manual. 3rd ed. Cornell University, Geneva, NY, USA. 123p.

Nelson, D.W., Sommers, L.E., 1982. Total carbon, organic carbon, and organic matter. In: Methods of Soil Analysis, Part 2, Chemical and Microbiological Properties, Second Edition. Number 9, Page, A.L., Keeney, D. R., Baker, D.E., Miller, R.H., Ellis, R. Jr., Rhoades, J.D. (Eds.). ASA-SSSA, Madison, Wisconsin, USA. pp. 539–580.

Olsen, S.R., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate, Circular No. 939. US Department of Agriculture, Washington, D.C. USA. 19p.

Panigrahi, B., Razavi, S., Doig, L., Cordell, B., Gupta, H., Liber, K., 2025. On robustness of the explanatory power of machine learning models: Insights from a new explainable AI approach using sensitivity analysis. *Water Resources Research* 61(1): e2024WR037398.

Prasad, R., Shivay, Y.S., 2021. Phosphorus × Other plant nutrient ınteractions, reaction products, anion exchange and phosphate fixation in soil and strategies to increase availability of the native and applied p to crop plants-A mini review and critique. *Agricultural Reviews* 42(2): 220–224.

Qin, C., Li, S.L., Yu, G.H., Bass, A.M., Yue, F.J., Xu, S., 2022. Vertical variations of soil carbon under different land uses in a karst critical zone observatory, SW China. *Geoderma* 412: 115741.

Qiu, J., Liu, F., Wang, D., Yan, K., Guo, J., Huang, W., Feng, Y., 2025. Mapping key soil properties in low relief areas using integrated machine learning and geostatistics. *Ecological Indicators* 171: 113228.

Raiesi, F., Salek-Gilani, S., 2020. Development of a soil quality index for characterizing effects of land-use changes on degradation and ecological restoration of rangeland soils in a semi-arid ecosystem. *Land Degradation and Development* 31(17): 2382–2397.

Sanad, H., Moussadek, R., Mouhir, L., Oueld Lhaj, M., Dakak, H., El Azhari, H., Yachou, H., Ghanimi, A., Zouahri, A., 2024. Assessment of soil spatial variability in agricultural ecosystems using multivariate analysis, soil quality index (SQI), and geostatistical approach: A case study of the Mnasra region, Gharb Plain, Morocco. *Agronomy* 14(6): 1112.

Soil Survey Staff, 1992. Procedures for collecting soil samples and methods of analysis for soil survey. Soil Survey Investigations Reports U.S. Govermentan Print Office, Washington D.C., USA.

Şenol, H., Alaboz, P., Demir, S., Dengiz, O., 2020. Computational intelligence applied to soil quality index using GIS and geostatistical approaches in semiarid ecosystem. *Arabian Journal of Geosciences* 13(23): 1235.

Vasu, D., Tiwary, P., Chandran, P., 2024. A novel and comprehensive soil quality index integrating soil morphological, physical, chemical, and biological properties. *Soil and Tillage Research* 244: 106246.

Wang, Y., Yuan, F., Cammarano, D., Liu, X., Tian, Y., Zhu, Y., Cao, W., Cao, Q., 2025. Integrating machine learning with spatial analysis for enhanced soil interpolation: Balancing accuracy and visualization. *Smart Agricultural Technology* 11: 101032.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. 2nd ed. John Wiley & Sons, Chichester, UK. 336p.

Yertayeva, Z., Kızılkaya, R., Kaldybayev, S., Seitkali, N., Abdraimova, N., Zhamangarayeva, A., 2019. Changes in biological soil quality indicators under saline soil condition after amelioration with alfalfa (Medicago sativa L.) cultivation in meadow Solonchak. *Eurasian Journal of Soil Science* 8(3): 189-195.

Yu, P., Zhang, K., Xiao, D., Li, C., Wang, S., 2023. Establishing a soil quality index to evaluate soil quality after afforestation in a karst region of Southwest China. *Catena* 220: 106675.

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* 338: 445–452.