Analytical uncertainty of the statistical paradigm as a stochastic risk vector in soil quality assessment under long-term conservation management

Luiz Diego Vidal Santos
ldvsantos@uefs.br
State University of Feira de Santana
Alceu Pedrotti
Universidade Federal de Sergipe
Jusimara de Andrade Santos
Universidade Federal de Sergipe
Jaelson Souza Araújo
Universidade Federal de Sergipe
Sara Julliane de Araújo Santos
Universidade Federal de Sergipe
Emersson Guedes da Silva
Universidade Federal de Sergipe
Renisson Neponuceno de Araújo Filho
Federal Rural University of Pernambuco
Brisa Marina da Silva Andrade
Universidade Federal de Sergipe
Jémison Mattos dos Santos
State University of Santa Cruz
Francisco Sandro Rodrigues Holanda
Universidade Federal de Sergipe

Research Article

Keywords: soil quality index, fuzzy inference system, PLS-SEM, MARS, Random Forest, ANOVA, methodological comparison, chemical fertility, soil microbiology, cover crops, management systems

Posted Date: May 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9414180/v1

Additional Declarations: No competing interests reported.

Analytical uncertainty of the statistical paradigm as a stochastic risk vector in soil quality assessment under long-term conservation management

Luiz Diego Vidal Santos[2*], Alceu Pedrotti[1],
Jusimara de Andrade Santos[1], Jaelson Souza Araújo[4],
Sara Julliane de Araújo Santos[1], Emersson Guedes da Silva[4],
Renisson Neponuceno de Araújo Filho[3],
Brisa Marina da Silva Andrade[1], Jémison Mattos dos Santos[5],
Francisco Sandro Rodrigues Holanda[4]

[1]PRODEMA, Universidade Federal de Sergipe, São Cristóvão, Sergipe, Brazil.
[2*]DCHF, Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil.
[3]DTR, Universidade Federal Rural de Pernambuco, Recife, Pernambuco, Brazil.
[4]DEA, Universidade Federal de Sergipe, São Cristóvão, Sergipe, Brazil.
[5]Universidade Estadual de Santa Cruz, Ilhéus, Bahia, Brazil.

*Corresponding author(s). E-mail(s): ldvsantos@uefs.br;
Contributing authors: alceupedrotti@gmail.com;
jusimara.ufs@gmail.com; jaelsonagro1207@gmail.com;
sara.julliane@academico.ufs.br; academicoguedes@gmail.com;
renisson.neponuceno@ufrpe.br; brisamarina.andrade@gmail.com;
jmsantos4@uesc.br; fsrholanda@academico.ufs.br;

Abstract

Soil quality indices (SQI) are widely used to evaluate conservation management practices, yet the sensitivity of treatment rankings to the choice of aggregation method remains poorly quantified when structurally distinct statistical families are compared on the same dataset. In this study, we compared five paradigms (ANOVA/GLM, fuzzy inference, PLS-SEM, MARS, and Random Forest) applied to an Ultisol of the Coastal Tablelands monitored for 22 years under a factorial of three tillage systems and four cover crops ($\boldsymbol{n = 108}$, 34 indicators distributed across five domains: physical, chemical, microbiological, agronomic, and yield). Overall concordance was significant (Kendall $\boldsymbol{W = 0.73}$, $\boldsymbol{p < 10^{-33}}$), yet the nonlinear paradigms (fuzzy, MARS, RF) formed a practically interchangeable block ($\boldsymbol{\rho > 0.99}$; Bland–Altman limits $< \boldsymbol{\pm 0.40}$), whereas ANOVA and PLS-SEM diverged from the nonlinear block ($\boldsymbol{\bar{\rho}_{\text{Lin-NL}} = 0.50}$; limits of agreement $\boldsymbol{\approx \pm 1.80\text{–}2.40}$), with 67% of intermediate management combinations shifting rank depending on the paradigm. The divergence between families was confirmed by the Wilcoxon test ($\boldsymbol{W = 18}$, $\boldsymbol{p = 0.012}$), indicating that nonlinear methods exhibit higher mutual concordance than concordance with linear paradigms. Minimum tillage with *Cajanus cajan* ranked first across all paradigms. These results suggest that adopting at least one nonlinear method may reduce the risk of algorithm-dependent conservation recommendations.

Keywords: soil quality index, fuzzy inference system, PLS-SEM, MARS, Random Forest, ANOVA, methodological comparison, chemical fertility, soil microbiology, cover crops, management systems

1 Introduction

Soil degradation compromises the productive capacity of approximately 33% of global agricultural land, with annual losses estimated between 25 and 40 billion tonnes of fertile soil through water erosion (Lal 2015). In humid tropical regions, where accelerated organic matter mineralization and rainfall aggressiveness intensify these processes, the adoption of conservation systems (no-till, minimum tillage, and cover cropping) has become a central strategy for maintaining soil functions (Doran 2002; Cardoso et al. 2013). Quantitative assessment of the effectiveness of these practices requires tools that integrate physical, chemical, and biological attributes into auditable metrics, which has motivated the development of soil quality indices (SQI) (Karlen et al. 1997).

The SQI concept was consolidated from the framework proposed by Karlen et al. (1997) and the operational protocol of Andrews et al. (2002), which established the selection of a minimum data set (MDS) by principal component analysis (PCA) followed by additive scoring. Since then, comprehensive reviews have documented the proliferation of methodological variants, with more than 60 combinations of selection, weighting, and aggregation published by 2018 (Bünemann et al. 2018). This diversity, although a sign of disciplinary maturity, introduces a comparability problem among studies that adopt distinct paradigms for the same indicators.

In the classical approach, parametric models (ANOVA or GLM) produce linear scores from weights derived from PCA eigenvalues, assuming that the relationships between indicators and soil quality are monotonic and additive (Mukherjee and Lal 2014). The computational simplicity and interpretive transparency have made this paradigm predominant in conservation management assessments, including long-term experiments on Ultisols of the Coastal Tablelands in northeastern Brazil (Assunção et al. 2023). The limitation lies in the inability of the linear model to capture threshold effects and higher-order interactions among indicators from distinct domains (Dos Santos et al. 2021).

Fuzzy inference systems (FIS) were introduced as an alternative to accommodate the boundary uncertainty inherent in classifying continuous indicators into discrete quality categories (Torbert et al. 2008). By replacing rigid cutoffs with triangular or Gaussian membership functions, the FIS preserves the gradation of soil attributes and allows the incorporation of expert knowledge through linguistic rule bases (Nabiollahi et al. 2018). Recent applications in Brazilian tropical soils indicated that the FIS can detect management differences that the classical additive protocol does not distinguish (Dos Santos et al. 2021), although the calibration of membership functions remains dependent on regional data that are not always available.

Machine learning (ML) algorithms, such as Random Forest (RF) and Multivariate Adaptive Regression Splines (MARS), exploit nonlinearity and high-order interactions through recursive partitioning of the predictor space (Zeraatpisheh et al. 2020). RF, in particular, provides intrinsic variable importance metrics (by permutation or impurity) and out-of-bag estimates that dispense with explicit cross-validation, contributing to its widespread adoption in digital soil mapping (Wadoux et al. 2020). MARS, in turn, identifies adaptive knots that approximate discontinuities in the response without requiring prior discretization of indicators. The interpretability of ML, however, is lower than that of classical statistical models, and the risk of overfitting in small samples ($n < 50$) warrants consideration (Wadoux et al. 2020).

Partial least squares structural equation modeling (PLS-SEM) occupies an intermediate position among the preceding paradigms. By specifying latent constructs representing soil domains (physical, chemical, microbiological, agronomic, yield) and estimating path coefficients between them, PLS-SEM tests explicit causal hypotheses about the soil–plant–yield chain (Hair et al. 2019; Sarstedt et al. 2022). The mixed specification (reflective Mode A for indicators that manifest the construct and formative Mode B for indicators that cause the construct) provides theoretical flexibility (Diamantopoulos et al. 2008), although the linearity of structural paths limits the capture of nonlinear interactions among domains.

In practice, each environmental study opts for a single algorithmic paradigm and implicitly assumes that the classification of management combinations is deterministic and stable with respect to methodological choice. However, recent environmental modeling literature indicates that ignoring nonlinearities intrinsic to terrestrial ecosystems introduces analytical uncertainty that propagates to management decisions (Yoshioka and Yoshioka 2024; Brito et al. 2024; Granata et al. 2024).

Recent comparisons involving scoring variants (Mukherjee and Lal 2014), machine learning (Zeraatpisheh et al. 2020), and fuzzy–AHP logic (Pacci et al. 2026) addressed the problem marginally for SQI. Those efforts, however, remained restricted to subsets of the same analytical family and did not employ formal inter-rater concordance metrics (Venkateswarlu et al. 2025; Liu et al. 2025). In decision-support contexts under uncertainty, the absence of concordance protocols among statistical families may propagate algorithmic bias to management recommendations, compromising the reliability of the decision-making process at landscape scales (Angelini et al. 2024; Jarvis et al.

2023). Empirical evidence quantifying the risk of management ranking alteration when structurally distinct paradigms (linear and nonlinear) are simultaneously applied to the same soil dataset is therefore lacking.

In Ultisols of the Coastal Tablelands of northeastern Brazil, conservation systems improve soil indicators through continuous and often subtle gradients (Assunção et al. 2023; Oliveira et al. 2020). If statistical methods diverge precisely at the boundary ranges of these gradients, the practical utility of the SQI may be compromised, rendering management recommendations dependent on the algorithm employed (Rezaee et al. 2020; de Paul Obade and Lal 2016).

This study evaluates the analytical uncertainty in soil quality indices (SQI) imposed by the choice of aggregation method and quantifies the risk of altering conservation management rankings. To this end, we compared five analytically distinct approaches (ANOVA/GLM, FIS, PLS-SEM, MARS, and RF) applied to the same dataset ($n = 108$, 34 indicators distributed across five domains, three tillage systems, and four cover crop species in a 22-year experiment). The indicator vector combines physical, chemical, microbiological, agronomic, and yield attributes obtained from two complementary sampling campaigns at the same experimental site, providing representativeness of the domains required by international soil quality protocols (Karlen et al. 1997; Andrews et al. 2002).

If the five paradigms access compatible variance structures, ranking concordance is expected to be statistically different from zero ($W > 0$, $\alpha = 0.05$). This convergence, however, may not be homogeneous across analytical families, such that nonlinear methods (FIS, MARS, RF) would tend to exhibit higher mutual concordance than with linear paradigms. Additionally, concordance stability may depend on sample size, increasing log-linearly until stabilization beyond a critical sample threshold.

2 Material and Methods

2.1 Dataset description

The dataset used in this benchmarking was obtained from a long-term field experiment conducted at the Experimental Station of the Federal University of Sergipe, municipality of São Cristóvão, SE, Brazil (10°55′24″ S, 37°11′57″ W, 22 m asl), as illustrated in Fig. 1. The regional climate is classified as As′ (tropical rainy with dry summer) according to Köppen, with annual precipitation of approximately 1,600 mm concentrated between April and September. The soil is a Typic Hapludult (Acrisol, WRB; Argissolo Vermelho Amarelo Distrófico, Brazilian classification), derived from the Barreiras Formation, with a sandy loam A horizon and a subsurface cohesive layer. The experiment was established in 2001 in a strip-split plot design with subplots of $6\,\text{m} \times 10\,\text{m}$, and the analyzed data correspond to the 22nd year of continuous management (2022 season). The SQI evaluation by the classical additive method, referring to the 17th year of this experiment, was published by Assunção et al. (2023); the present study expands that baseline by applying five analytical paradigms to the 2022 season.

The observations ($n = 108$) were collected under a factorial design combining three tillage systems (no-till, NT; conventional tillage, CT; and minimum tillage, MT) with four cover crop species (cowpea, *Vigna unguiculata*; millet, *Pennisetum glaucum*;

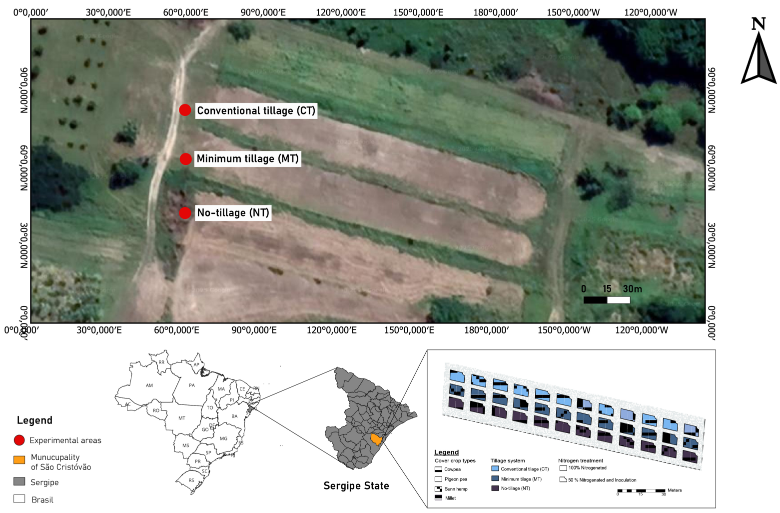

Fig. 1 Location and general view of the long-term experiment (established in 2001) at the Experimental Station of the Federal University of Sergipe, municipality of São Cristóvão, SE, on a Typic Hapludult of the Coastal Tablelands.

pigeon pea, *Cajanus cajan*; and sunn hemp, *Crotalaria juncea*), totaling 12 combinations with nine replicates each. The main crop *Zea mays* L. (hybrid SHS7990 PRO3) was sown at $0.80\,\mathrm{m} \times 0.20\,\mathrm{m}$ spacing ($\approx 54{,}000\,\mathrm{plants\,ha^{-1}}$), preceded by the cover crop manually sown 90 days before. Conventional tillage consisted of disk plowing to 25 cm followed by leveling harrow, minimum tillage employed only leveling harrow, and no-till excluded mechanical soil disturbance.

With nine replicates per factorial cell ($3 \times 4 = 12$ combinations), the design totals $n = 108$ observations. The *a posteriori* power analysis for the factorial ANOVA with $\alpha = 0.05$, medium effect size ($f = 0.25$), and $k = 12$ groups indicates statistical power > 0.90 for detecting differences among management combinations, meeting the threshold recommended by Mundfrom et al. (2005) for multivariate analyses.

Sampling was restricted to the 0–0.20 m layer for operational reasons, since the cohesive layer in this Ultisol lies between 0.25 and 0.45 m depth (Lima Neto et al. 2010) and sampling at greater depths would require trenches in all 108 plots, which was impractical in a long-term experiment with 60 m^2 plots. This restriction may underestimate the physical differentiation among treatments that act on the cohesive layer (e.g., the taproot of pigeon pea), constituting a recognized limitation of this study.

2.2 Soil and plant indicators

Thirty-four indicators distributed across five domains (soil physical, chemical, microbiological, agronomic, and yield) were measured (Table 1). Physical indicators (GMD, WMD, BD, RMH, WSA, SOC, VCI, MI, MA, BIR, PR, AW/TP), agronomic indicators (PH, ED, EL, STAND), and yield indicators (TE, ME, MEW, GY) were obtained in the 2022 season of the present experiment. Chemical indicators (pH, P, K, Ca, Mg, Al, CEC, SOM, TN) and microbiological indicators (qMIC, qCO2, BR, MBC, MBN) were obtained from a complementary evaluation conducted in the same experimental plots by Santos (2019), at the 0–10 cm depth, and integrated into the dataset as means per design cell (tillage × cover crop). The integration of two sampling campaigns at the same experimental site is justified by the temporal stability of soil attributes in long-term experiments (> 15 years) conducted on Ultisols of the Coastal Tablelands (Assunção et al. 2023), in which the differentiation among treatments results from cumulative effects that override interannual variability.

Soil pH was determined in a soil:water suspension (1:2.5) with a calibrated potentiometer (precision ±0.02), available P and K by Mehlich-1 with UV-Vis spectrophotometer reading at 725 nm and flame photometry, respectively (analytical CV $< 5\%$). Exchangeable Ca and Mg were extracted with $1\,\mathrm{mol\,L^{-1}}$ KCl and determined by atomic absorption spectrophotometry. Exchangeable Al was extracted with $1\,\mathrm{mol\,L^{-1}}$ KCl and titrated with $0.025\,\mathrm{mol\,L^{-1}}$ NaOH. CEC at pH 7 was calculated as the sum of bases plus potential acidity. Soil organic matter (SOM) was determined by the modified Walkley-Black method (CV $< 6\%$) and total nitrogen stock (TN) by Kjeldahl digestion. Soil organic carbon stock (SOC) was determined by the Walkley-Black method.

Microbiological indicators (MBC and MBN) were determined by the fumigation-extraction method (Vance et al. 1987), soil basal respiration (BR) by CO_2 capture in NaOH, and the metabolic (qCO2) and microbial (qMIC) quotients were calculated as derived ratios. Microporosity (MI) and macroporosity (MA) were obtained by tension table, basic infiltration rate (BIR) by double-ring infiltrometer, penetration resistance (PR) by impact penetrometer, and the available water/total porosity ratio (AW/TP) was calculated from the water retention curve.

The inclusion of 16 additional indicators from the chemical, microbiological, and physical domains required recalibration of FIS membership functions (expansion of the universe of discourse), readjustment of RF hyperparameters (`mtry` recalculated to $\lfloor\sqrt{34}\rfloor = 6$) and MARS (forward/backward reprocessing), as well as re-specification of the PLS-SEM model with five constructs (inclusion of the MICROBIOLOGICAL construct), so that all paradigms were re-estimated with the complete vector of 34 indicators. The $n/p = 108/34 \approx 3.2$ ratio falls below the $n/p = 5$ threshold recommended for PLS-SEM (Hair et al. 2019), a condition mitigated by multicollinearity screening (VIF > 5) that reduced the structural model to 25 indicators.

2.3 Paradigm 1: ANOVA/GLM with minimum data set selection

The classical approach was included because it represents the most widespread paradigm in SQI literature and operates under additivity and monotonicity assumptions that serve as a baseline for assessing the informational gain of nonlinear methods. The variance capture mechanism is restricted to the linear decomposition of variance by PCA eigenvalues, which implies that higher-order interactions and threshold effects among soil domains remain in the unexplained residual.

Table 1 Soil and plant indicators used in the benchmarking analysis.

Domain	Indicator	Abbreviation	Unit
Soil physical	Geometric mean diameter	GMD	mm
	Weighted mean diameter	WMD	mm
	Bulk density	BD	$\mathrm{Mg\,m^{-3}}$
	Residual moisture persistence	RMH	–
	Water-stable aggregates	WSA	%
	Soil organic carbon stock	SOC	$\mathrm{Mg\,ha^{-1}}$
	Vegetation cover index	VCI	%
	Microporosity	MI	$\mathrm{m^3\,m^{-3}}$
	Macroporosity	MA	$\mathrm{m^3\,m^{-3}}$
	Basic infiltration rate	BIR	$\mathrm{cm\,h^{-1}}$
	Penetration resistance	PR	MPa
	Available water / total porosity	AW/TP	–
Soil chemical	Soil pH in water (0–10 cm)	pH	–
	Available phosphorus (0–10 cm)	P	$\mathrm{mg\,dm^{-3}}$
	Available potassium (0–10 cm)	K	$\mathrm{mg\,dm^{-3}}$
	Exchangeable calcium	Ca	$\mathrm{cmol_c\,dm^{-3}}$
	Exchangeable magnesium	Mg	$\mathrm{cmol_c\,dm^{-3}}$
	Exchangeable aluminum	Al	$\mathrm{cmol_c\,dm^{-3}}$
	Cation exchange capacity	CEC	$\mathrm{cmol_c\,dm^{-3}}$
	Soil organic matter	SOM	$\mathrm{g\,kg^{-1}}$
	Total nitrogen stock	TN	$\mathrm{Mg\,ha^{-1}}$
Microbiological	Microbial biomass carbon	MBC	$\mathrm{mg\,C\,kg^{-1}}$
	Microbial biomass nitrogen	MBN	$\mathrm{mg\,N\,kg^{-1}}$
	Soil basal respiration	BR	$\mathrm{mg\,C\text{-}CO_2\,kg^{-1}\,day^{-1}}$
	Metabolic quotient	qCO2	$\mathrm{mg\,C\text{-}CO_2\,mg^{-1}\,MBC\,day^{-1}}$
	Microbial quotient	qMIC	%
Agronomic	Plant height	PH	cm
	Ear diameter	ED	mm
	Ear length	EL	cm
	Final stand	STAND	$\mathrm{plants\,ha^{-1}}$
Yield	Total ears per hectare	TE	$\mathrm{ears\,ha^{-1}}$
	Marketable ears per hectare	ME	$\mathrm{ears\,ha^{-1}}$
	Marketable ear weight per hectare	MEW	$\mathrm{kg\,ha^{-1}}$
	Grain yield	GY	$\mathrm{kg\,ha^{-1}}$

Note: All indicators were normalized to the 0–10 scale prior to applying the fuzzy and ML paradigms. Chemical and microbiological indicators were obtained from a complementary evaluation (Santos 2019) at the same experimental site and integrated as means per design cell. Physical indicators MI, MA, BIR, PR, and AW/TP also originate from the complementary campaign.

Generalized linear models (GLM) were fitted for each indicator following the protocol of Andrews et al. (2002), according to

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{1}$$

where α_i represents the effect of the tillage system ($i = 1, 2, 3$), β_j the effect of the cover crop species ($j = 1, \ldots, 4$), $(\alpha\beta)_{ij}$ their interaction, and ε_{ijk} the residual error.

Type II ANOVA was computed using the `car` package. Multiple comparisons were performed by estimated marginal means (`emmeans`) with Tukey adjustment.

The minimum data set (MDS) selection was conducted via PCA. Indicators with factor loadings exceeding 0.70 on each retained component were identified and, among correlated indicators within each component, the one with the highest communality was retained. The linear SQI was computed as

$$\mathrm{SQI}_{\mathrm{ANOVA}} = \sum_{i=1}^{k} w_i \cdot S_i \tag{2}$$

where w_i corresponds to the weight derived from PCA eigenvalues and S_i to the normalized score of each indicator. Normality of GLM residuals was assessed by the Shapiro-Wilk test and homoscedasticity by Levene's test (`car` package). Indicators whose residuals violated the normality assumption ($p < 0.05$) were Box-Cox transformed prior to ANOVA; when the transformation did not normalize the residuals, nonparametric ANOVA (Kruskal-Wallis) followed by Dunn's test with Bonferroni adjustment was employed.

2.4 Paradigm 2: Fuzzy inference system

Fuzzy inference was selected for its ability to accommodate the gradual transition between soil quality classes without imposing rigid boundaries on continuous indicators. The mechanism operates through the overlap of membership functions that assign partial degrees of association to each linguistic category, preserving the gradation of soil attributes and allowing expert knowledge incorporation through linguistic rule bases (Nabiollahi et al. 2018). Threshold effects that the linear combination compresses are thus captured by the rule base without prior parametric specification.

A Mamdani-type fuzzy inference system (Mamdani and Assilian 1975) was constructed following the framework of Torbert et al. (2008). Each of the 15 original indicators was fuzzified into three linguistic categories (low, medium, high) with triangular and, separately, Gaussian membership functions. The triangular functions were defined as

$$\mu_A(x) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \tag{3}$$

where a, b, and c delimit the support and peak of the fuzzy set. The Gaussian functions followed

$$\mu_A(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right) \tag{4}$$

The rule base was constructed from expert knowledge and defuzzification was performed by the centroid method. The output, termed Integrated Soil Performance and Conservation Index (ISPCI), was computed on the 0–10 scale using the `FuzzyR` package.

2.5 Paradigm 3: PLS-SEM structural equation modeling

PLS-SEM was included because it enables the specification of directional relationships among soil domains (physical, chemical, microbiological, agronomic, yield) as latent constructs, testing explicit causal hypotheses about the soil–plant–yield chain (Hair et al. 2019; Sarstedt et al. 2022). The capture mechanism resides in the estimation of path coefficients that weight the relative contribution of each domain to yield, while the formative specification (Mode B) allows indicators such as pH, P, Ca, and MBC to act as determinants of fertility and biological activity rather than as reflective manifestations of a latent factor. This flexibility distinguishes PLS-SEM from the other paradigms by incorporating a theoretical structure into the aggregation, although the linearity of structural paths limits the capture of higher-order nonlinear interactions among domains.

A hierarchical PLS path model was specified with five first-order latent constructs (PHYSICAL, CHEMICAL, MICROBIOLOGICAL, AGRONOMIC, and YIELD) and one second-order construct, SQI. The structural model specified directional paths according to

$$\mathrm{SQI}_{\mathrm{PLS}} = \gamma_1 \cdot \eta_{\mathrm{PHYS}} + \gamma_2 \cdot \eta_{\mathrm{CHEM}} + \gamma_3 \cdot \eta_{\mathrm{MICRO}} + \gamma_4 \cdot \eta_{\mathrm{AGRO}} + \gamma_5 \cdot \eta_{\mathrm{YIELD}} + \zeta \quad (5)$$

Indicators such as bulk density (BD) and ear production (MEW, ME) act as causes (rather than reflections) of their respective constructs, which implies formative relationships. For this reason, the model adopted formative specification (Mode B) for all constructs. A multicollinearity screening (VIF $>$ 5) excluded four indicators with excessive inflation: CEC (VIF = 37.5), TN (VIF = 13.7), MBC (VIF = 23.9), and GY (VIF = 20.1), resulting in 25 indicators distributed among the five first-order constructs: PHYSICAL (Mode B: GMD, WMD, BD, RMH, WSA, SOC, VCI, MI, MA), CHEMICAL (Mode B: pH, P, K, Ca, Mg, Al, SOM), MICRO (Mode B: qMIC, qCO2, BR, MBN), AGRO (Mode B: PH, ED, EL, STAND), and YIELD (Mode B: TE, ME, MEW). Validation of Mode B followed criteria of outer weight significance and absence of residual multicollinearity.

Estimation followed a two-stage approach with the `seminr` package (Ray et al. 2022). Overall model quality was assessed by the SRMR (Standardized Root Mean Square Residual), with an acceptance threshold < 0.08 (Hair et al. 2019), and by the Goodness of Fit (GoF) as the geometric mean of average communalities and mean R^2 of endogenous constructs. For formative constructs (Mode B), a redundancy test by correlation between the formative score and a global reflective indicator was applied, verifying convergent validity when $r > 0.70$. In the first stage, the five first-order constructs (PHYSICAL, CHEMICAL, MICRO, AGRO, YIELD) were estimated iteratively, with structural paths PHYSICAL→YIELD, CHEMICAL→YIELD, MICRO→YIELD, and AGRO→YIELD. In the second stage, the latent scores of the five constructs were submitted to PCA, and the first principal component was adopted as $\mathrm{SQI}_{\mathrm{PLS}}$, scaled to the 0–10 interval. Confidence intervals for path coefficients were obtained by bootstrap of 500 resamples.

2.6 Paradigm 4: MARS with ablation analysis

MARS was selected for identifying discontinuities in the response through hinge functions that segment the predictor space into locally linear regions, without requiring prior discretization or specification of the number or position of thresholds. The nonlinearity capture mechanism resides in the automatic selection of adaptive knots by the forward stepwise addition and backward pruning algorithm, which approximates relationships whose curvature varies along the range of indicator values.

MARS models (Friedman 1991) were fitted with the `earth` package (Milborrow 2023) to capture nonlinear relationships between indicators and a target variable (yield or composite SQI). The model selected basis functions by forward stepwise addition and backward pruning according to

$$\hat{f}(x) = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{K_m} h_{km}(x_{v(k,m)}) \tag{6}$$

where h_{km} are hinge functions of the form $\max(0, x - t)$ or $\max(0, t - x)$.

Variable importance was assessed by ablation analysis, in which each indicator was sequentially removed and the Mean Absolute Rank Shift quantified the impact on the predicted SQI.

Indicators whose ablation score exceeded a predefined threshold were identified as critical for the index. The adequacy of adaptive knots was verified by residuals-versus-fitted-values plots and the Breusch-Pagan test for heteroscedasticity (`lmtest` package); the absence of systematic patterns in the residuals and $p > 0.05$ in the BP test confirm that the hinge functions adequately capture the variance structure of the data. Additionally, global variable importance was estimated by SHAP (SHapley Additive exPlanations) values with the `DALEX` package, enabling direct comparison with RF importance metrics and interpretive alignment between paradigms.

2.7 Paradigm 5: Random Forest regression

RF (Breiman 2001) was included for its ability to capture high-order interactions and nonlinearities through recursive partitioning of the predictor space in an ensemble of independent decision trees. The mechanism operates by bootstrap aggregation (bagging) of multiple trees trained on random subsets of observations and variables, which reduces prediction variance and provides internal error estimates (OOB) that dispense with explicit train/test partitioning.

Random Forest models were trained with the `randomForest` package using 500 trees, $\lfloor\sqrt{q}\rfloor = 6$ variables randomly sampled at each split, and minimum terminal node size of 5. The choice of mtry $= \lfloor\sqrt{q}\rfloor$ instead of $\lfloor q/3\rfloor$ reduces inter-tree correlation and confers greater stability to the ensemble when the n/q ratio is moderate (≈ 3.2), a condition present in this dataset. Convergence of the out-of-bag error (OOB-MSE) was verified through an OOB-MSE $\times$ number of trees curve, and OOB error estimates served as an internal validation metric. Variable importance was quantified by two complementary routes, the percentage increase in mean squared error (%IncMSE) after permutation of each predictor and global SHAP values obtained with the `DALEX` package (500 permutations).

The 95% confidence intervals for SQI predictions were estimated by the infinitesimal jackknife method (Wager et al. 2014) implemented in the `randomForestCI` package. In addition to R^2_{OOB}, the OOB-RMSE in original units of the 0–10 scale was reported to allow direct interpretation of the magnitude of predictive error. The predicted SQI was defined as

$$\mathrm{SQI}_{\mathrm{RF}} = \frac{1}{B}\sum_{b=1}^{B} T_b(\mathbf{x}) \tag{7}$$

where T_b corresponds to the prediction of tree b for observation $\mathbf{x}$.

2.8 Inter-paradigm concordance assessment

The five paradigms generate SQI values on distinct scales. Prior to comparison, all SQI vectors were standardized by the rank-based inverse normal transformation (Blom 1958), according to

$$z_i = \Phi^{-1}\left(\frac{r_i - 0.5}{n}\right) \tag{8}$$

where r_i is the rank of observation i and Φ^{-1} the quantile function of the standard normal distribution. This transformation preserves the original ordering and places all paradigms on a standard Gaussian scale, eliminating distortions arising from domain differences (0–1 vs. 0–10 vs. factor scores).

The comparison among mathematical families aimed to isolate the variance attributable to paradigm choice and estimate the level of false positives that inadequate models may generate. Kendall's coefficient of concordance (W) assessed the consistency among the five simultaneous rankings (permutation test of 10,000 replicates). The intraclass correlation coefficient, ICC(2,1) (Shrout and Fleiss 1979), estimated absolute agreement. Additionally, Gwet's AC2 coefficient (Gwet 2008) corrected for chance-expected agreement (`irrCAC` package).

To delineate ranges in which paradigm switching would alter management classification, Bland–Altman analysis (Bland and Altman 1986) quantified the limits of agreement through systematic bias plots ($\mathrm{SQI}_A - \mathrm{SQI}_B$) with limits at ± 1.96 standard deviations. Additionally, a Spearman correlation matrix (ρ) quantified the correlations among standardized SQI vectors. A consensus meta-SQI was computed by Borda count (mean rank) of the five paradigms (Lin 2010).

The difference in concordance between nonlinear paradigms and concordance with linear paradigms was tested by paired Wilcoxon test on Fisher-Z transformed ρ values, comparing the subset of nonlinear$\times$nonlinear pairs with linear$\times$nonlinear pairs.

2.9 Sample size sensitivity

The sensitivity of concordance to sample size was assessed by stratified Monte Carlo simulation, aiming to establish the sample threshold beyond which inter-paradigm concordance stabilizes at acceptable levels. The complete set ($n = 108$) was subsampled at five levels ($n = 15$, 30, 50, 75, 100) with 200 iterations per level. In each iteration, sampling was stratified by design cell (tillage $\times$ cover crop), preserving the original proportionality and preventing chance imbalance from artificially inflating the variance between paradigms. The random seed of each iteration was stored for reproducibility.

In each subsample, the five paradigms were reprocessed and Kendall's W recalculated, allowing the assessment of concordance dependence on sample size. The relationship between concordance and sample size was tested by GLS regression of $\bar{W}$ on $\log(n)$ with a correlation structure for paired observations, identifying the threshold beyond which sample increments do not reduce concordance variability among rankings.

3 Results and Discussion

3.1 Predictive performance and variable importance

The Random Forest model explained 96.0% of SQI variance by the out-of-bag error ($R^2_{\mathrm{OOB}} = 0.960$; OOB-RMSE = 0.21 on the 0–10 scale), with OOB-MSE stabilization from approximately 200 trees onward and 95% confidence interval widths consistent with the within-cell variability among replicates.

Permutation importance (%IncMSE) identified yield and agronomic domain variables as the primary predictors (Fig. 2), with total ears per hectare (20.2%), final stand (19.5%), marketable ears (13.2%), and grain yield (13.2%) concentrating more than 66% of cumulative importance. Vegetation cover index (VCI, 11.2%) and ear diameter (ED, 12.3%) completed the high-importance block. Chemical indicators ranked at an intermediate level, with P (8.0%), Ca (7.6%), SOM (7.4%), CEC (6.7%), and K (6.6%) individually exceeding most physical indicators. Microbiological indicators (qMIC, 5.9%; MBC, 5.6%; BR, 5.2%; MBN, 5.1%; qCO2, 4.4%) clustered in the lower importance tier, above only the structural physical attributes (GMD, WMD, BD, WSA).

Global SHAP values, obtained by 500 permutations via DALEX, corroborated this hierarchy, with reduced dispersion for the most important variables (Lundberg and Lee 2017; Shahriari et al. 2023). The incorporation of chemical and microbiological indicators established two intermediate importance tiers between the productive/agronomic and physical-structural components, consistent with the mediating role of chemical fertility and biological activity in the soil–plant relationship documented by Nabiollahi et al. (2020) and Cardoso et al. (2013) in rainfed systems.

Chemical attributes responded more rapidly to management than physical indicators in a long-term experiment on black soil in China, where cover cropping required more than 15 years to produce statistically detectable physical differences (Zhao et al. 2022). In Ultisols of the Coastal Tablelands, the subsurface cohesive layer may amplify this asymmetry by restricting physical differentiation among treatments in the sampled layer (0–0.20 m), channeling most of the explainable variance to agronomic and chemical indicators located above the mechanical restriction.

Partial dependence curves (Fig. 3) reveal that total ears per hectare and final stand exhibit a sigmoidal response on the RF-predicted SQI, with a transition threshold between normalized values 4 and 6, beyond which the marginal effect stabilizes at an asymptotic plateau. Marketable ears and grain yield, in turn, exhibit a monotonically increasing response with lower saturation intensity, suggesting a quasi-linear contribution to the integrated score (Friedman 2001).

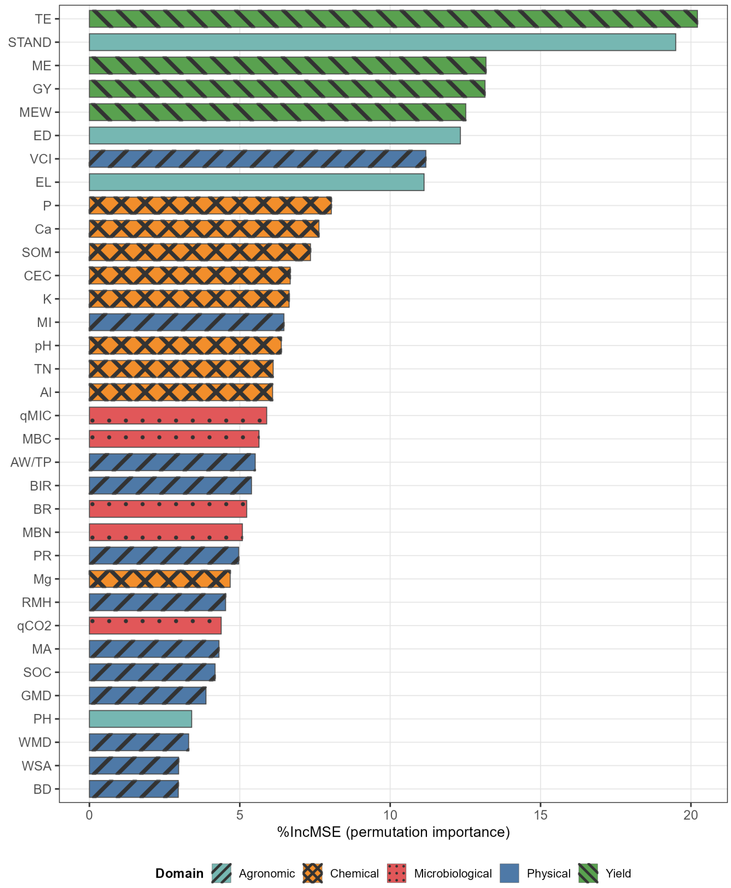

Fig. 2 Variable importance by permutation (%IncMSE) in the RF model.

Note: Bars colored by indicator domain.

This threshold pattern may reflect a compensation effect among cover crop species under an Ultisol with a cohesive layer, where productivity gains stabilize beyond a certain stand level, while morphological attributes maintain a gradual dose-response relationship (Assunção et al. 2023). Under no-till with maize/cover crop rotation on an Ultisol, pigeon pea and sunn hemp promoted greater spore abundance and aggregate stability than monoculture grasses (Moitinho et al. 2020), suggesting that rhizosphere biological mediation may contribute to the coupling between physical quality and productive response captured by the PDPs.

In the five-construct PLS-SEM model, AGRO explained the largest share of YIELD variance ($\gamma = 0.620$, $p < 0.001$), whereas the paths PHYSICAL→YIELD ($\gamma = -0.132$), CHEMICAL→YIELD ($\gamma = 0.086$), and MICRO→YIELD ($\gamma = -0.177$) did not reach statistical significance. The negative coefficient of the MICRO construct may reflect the inverse relationship between microbial activity and immediate yield

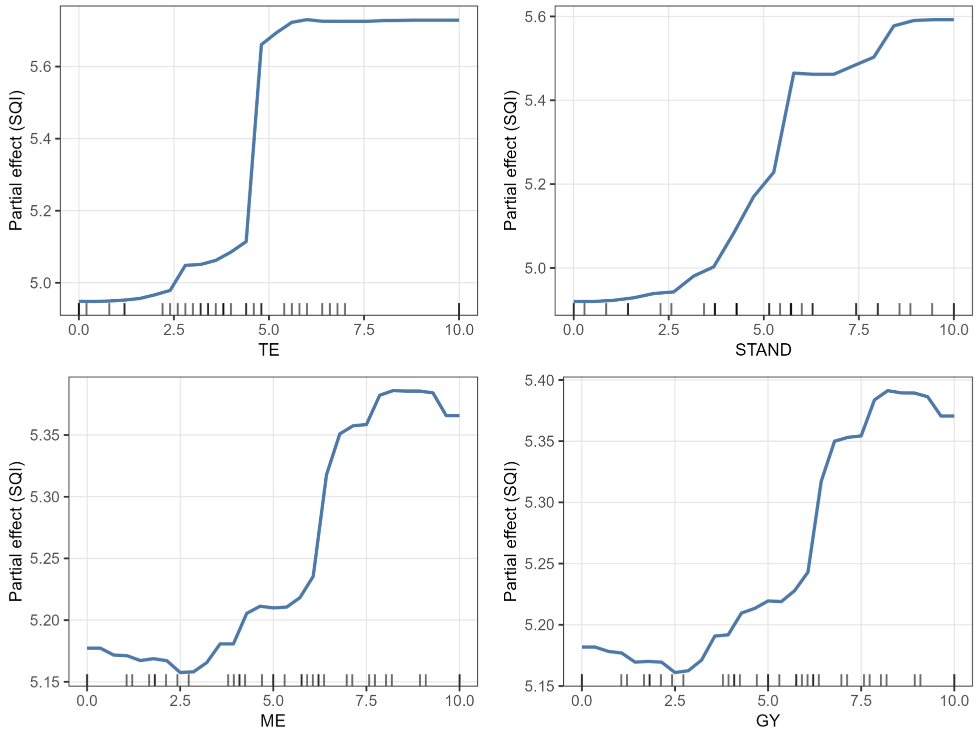

Fig. 3 Partial dependence curves for the four most important indicators in the RF model.

Note: Horizontal axis, normalized value (0–10); vertical axis, partial effect on predicted SQI. Rug plots indicate the marginal data distribution.

commonly observed in soils under continuous residue input, in which temporary nutrient immobilization by microbial biomass competes with root uptake (Cardoso et al. 2013). The negative PHYSICAL path is consistent with the restriction imposed by the cohesive layer between 0.25 and 0.45 m in this Ultisol, which concentrates absorption in the surface layer and may amplify apparent negative correlations between physical indicators (strongly conditioned by the subsurface restriction) and yield. Confirmation of this mechanism requires depth-stratified sampling and imaging techniques that allow visualization of the structural discontinuity of the cohesive layer.

The VIF screening excluded four indicators (CEC = 37.5; MBC = 23.9; GY = 20.1; TN = 13.7), reducing the model to 25 manifest variables distributed across five constructs. The structural model R^2 for the YIELD construct was 0.688 (Adj. $R^2 = 0.676$), indicating that the four exogenous constructs explain approximately 69% of yield variance. The second-order construct, obtained from the first principal component of the five latent scores, was scaled to the 0–10 interval and adopted as $\mathrm{SQI_{PLS}}$.

The five paradigms produced SQI distributions with distinct amplitudes and dispersions. FIS, MARS, and RF exhibited similar means (5.24 on the 0–10 scale, $s \approx 1.1$), whereas $\mathrm{SQI_{PLS}}$ (mean = 5.12, $s = 2.40$) and $\mathrm{SQI_{ANOVA}}$ (mean = 5.10, $s = 2.89$)

occupied broader ranges, reflecting the linearity and dimensionality reduction imposed by these paradigms.

3.2 Inter-paradigm concordance and meta-SQI

Kendall's coefficient of concordance ($W = 0.726$, $\chi^2 = 389$, $p = 1.3 \times 10^{-33}$) indicates significant overall concordance among the five simultaneous rankings (Legendre 2012). The permutation test (10,000 replicates) corroborated this result ($p < 0.0001$). The ICC(2,1) of 0.450 (95% CI: 0.360–0.544) indicates moderate absolute agreement, below the 0.75 threshold generally considered "good" in reliability applications (Koo and Li 2016), reflecting the greater divergence among analytical families when indicators from multiple domains are included. Gwet's AC1 coefficient, calculated on SQI quintiles, was 0.321 (95% CI: 0.27–0.37), a conservative value that reflects the penalization for chance-expected marginal agreement.

The Spearman correlation matrix revealed a two-block structure with sharp differentiation (Table 2). FIS, MARS, and RF formed a nearly indistinguishable block ($\rho > 0.992$ between any pairs), whereas ANOVA diverged markedly from both the nonlinear block ($\rho = 0.269$–0.286) and PLS-SEM ($\rho = 0.586$). PLS-SEM occupied an intermediate position, with moderate correlations with the nonlinear block ($\rho = 0.718$–0.743).

The Wilcoxon test confirmed that the mutual concordance among nonlinear paradigms exceeds concordance with linear paradigms ($W = 18$, $p = 0.012$), with the mean among nonlinear pairs ($\bar{\rho}_{\text{NL-NL}} = 0.994$) exceeding the mean among linear–nonlinear pairs ($\bar{\rho}_{\text{Lin-NL}} = 0.502$). This result indicates that the incorporation of indicators from the expanded chemical and microbiological domains amplified the divergence between analytical families, exposing high-order relationships and inter-domain interactions that nonlinear paradigms capture but that the linear PCA decomposition compresses to the residual (Zeraatpisheh et al. 2020).

Table 2 Spearman correlation matrix (ρ) among the five SQI paradigms after Blom transformation.

	ANOVA	Fuzzy	PLS-SEM	MARS	RF
ANOVA	1.000	0.269	0.586	0.276	0.286
Fuzzy		1.000	0.723	0.994	0.994
PLS-SEM			1.000	0.718	0.743
MARS				1.000	0.992
RF					1.000

The Bland–Altman limits of agreement (Bland and Altman 1986) (Table 3) quantitatively reinforced this two-block structure. The LoA between pairs within the FIS-MARS-RF block ranged from ±0.35 (Fuzzy vs. MARS) to ±0.41 (MARS vs. RF) on the Blom scale, suggesting practical substitutability. In contrast, pairs involving ANOVA exhibited LoA from ±1.80 (ANOVA vs. PLS) to ±2.40 (ANOVA vs. Fuzzy), whereas PLS-SEM versus the nonlinear block occupied an intermediate range

(LoA $\approx \pm 1.58$), indicating that substitution of the linear paradigm by any nonlinear alternative may alter individual ranking positions.

Table 3 Bland–Altman analysis summary between paradigm pairs (Blom scale).

Pair	$\overline{\Delta M}$	$s_{\Delta M}$	Lower LoA	Upper LoA
Fuzzy vs RF	0.00	0.206	−0.403	0.403
Fuzzy vs MARS	0.00	0.181	−0.353	0.357
MARS vs RF	0.00	0.206	−0.406	0.402
PLS vs Fuzzy	0.00	0.817	−1.601	1.601
PLS vs MARS	0.00	0.805	−1.576	1.580
PLS vs RF	0.00	0.804	−1.575	1.575
ANOVA vs PLS	0.00	0.918	−1.800	1.800
ANOVA vs Fuzzy	0.00	1.226	−2.403	2.403
ANOVA vs MARS	0.00	1.203	−2.356	2.360
ANOVA vs RF	0.00	1.207	−2.365	2.365

Note: $\overline{\Delta M}$, mean difference; $s_{\Delta M}$, standard deviation of the difference; LoA, limits of agreement ($\pm 1.96\, s_{\Delta M}$).

Practical substitutability within the nonlinear block is confirmed by the homogeneous dispersion of the Fuzzy–RF pair within the ±0.40 range along the mean axis, with points symmetrically distributed around the null difference and no proportional trend with SQI level. The ANOVA–Fuzzy pair, in contrast, exhibited a fan-shaped pattern suggesting proportional bias with SQI level (Giavarina 2015), with increasing discordance toward higher Blom-scale values and LoA of ±2.40, indicating that the linear compression diverges progressively from nonlinear paradigms as the integrated score increases.

The PLS–RF pair occupied an intermediate position (LoA ≈ ±1.58), with apparent heteroscedasticity at extreme values signaling more pronounced discordance for management combinations with lower quality (Fig. 4). This concordance gradient, from substitutability within the nonlinear block to systematic divergence of linear paradigms, reinforces that the choice of analytical family, rather than merely the internal parameterization of each algorithm, governs the degree of concordance among SQI rankings.

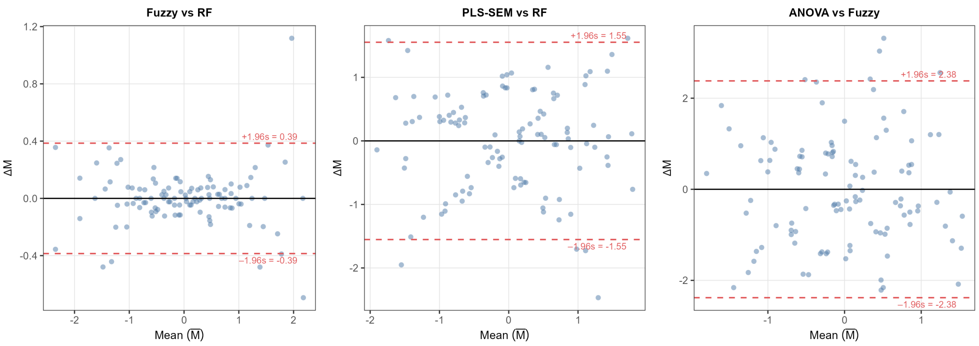

Fig. 4 Bland–Altman plots for three representative paradigm pairs (Blom scale).

Note: Solid line, mean difference ($\overline{\Delta M} \approx 0$); dashed lines, limits of agreement at $\pm 1.96\, s_{\Delta M}$.

Proximity to the identity line varies among paradigms in the 1:1 scatter plots (Fig. 5). Fuzzy ($\rho^2 = 0.989$) and MARS ($\rho^2 = 0.984$) align almost perfectly, whereas PLS-SEM ($\rho^2 = 0.552$) and ANOVA ($\rho^2 = 0.082$) exhibit marked dispersion, with individual observations deviating up to 2.0 Blom units from the 1:1 line (Venkateswarlu et al. 2025).

This magnitude of deviation implies that, for a given observation, the quality diagnosis "high" or "low" may be contingent on the paradigm when the conventional median threshold is adopted, a pattern that reinforces the need for concordance metrics in SQI evaluation (Liu et al. 2025). A comparison of eight SQI models (TDS/MDS × linear/nonlinear scoring × SQI/NQI) in degraded lands of Ethiopia showed that the nonlinear scoring function produced indices systematically 12–18% higher than the linear score, with amplified divergence in low-quality soils (Kahsay et al. 2025), a pattern consistent with the fan-shaped bias between ANOVA and Fuzzy observed in this

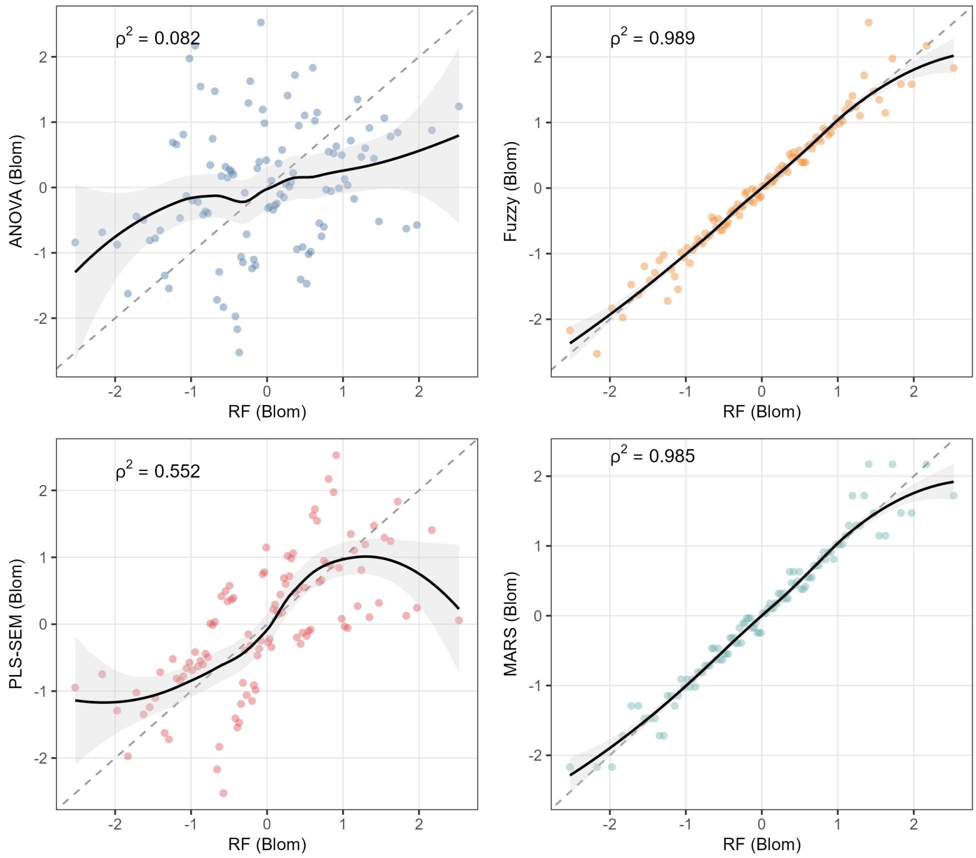

Fig. 5 1:1 scatter plots of SQI (Blom scale) between each paradigm and RF as the reference.

Note: Dashed line, identity; solid line, local regression. ρ^2 values annotated in each panel.

study (Fig. 4). In wheat–maize rotation under contrasting tillage regimes in northern China, the nonlinear-scored SQI detected conservation management benefits that the linear score did not distinguish (Zhang et al. 2024), reinforcing the hypothesis that the analytical family, rather than merely the indicator set, conditions index sensitivity.

The convergence among FIS, MARS, and RF is consistent with the hypothesis that these paradigms access the same variability subspace, defined by nonlinear relationships and high-order interactions among indicators that the linear model compresses or ignores (Zeraatpisheh et al. 2020; Wadoux et al. 2020). Dos Santos et al. (2021) observed an analogous pattern in a Ferralsol under integrated production systems, with the fuzzy SQI detecting management differences that the classical additive protocol did not distinguish, supporting the generality of nonlinear convergence in tropical soils.

The expansion of the indicator vector from 18 to 34, with the incorporation of the expanded chemical and microbiological domains, intensified the differentiation between analytical families, reducing $\bar{\rho}_{\text{Lin-NL}}$ from 0.71 to 0.50 and rendering the divergence between analytical families significant ($p = 0.012$). This result indicates that the additional microbiological and chemical indicators introduced between-plot variance with predominantly nonlinear structure, which adaptive algorithms capture as high-order inter-domain interactions, while PCA projects this information onto orthogonal components that compress inter-domain relationships (Rezaee et al. 2020).

The magnitude of this discordance varies with soil type and indexing method. Divergences of up to 30% between weighted additive and cumulative indices were reported in soils with heterogeneous indicator distribution in central Iran (Emami et al. 2024), a magnitude consistent with the LoA of ± 1.80 to ± 2.40 observed in this study for pairs involving ANOVA. In saline coastal soils of India, concordance among three methods (additive SQI_w, Nemoro SQI_n, and cumulative SQI_c) varied with salinity severity, being higher in intermediate-quality soils and lower at the extremes (Mahajan et al. 2021), a pattern resembling the heteroscedasticity documented in the Bland–Altman plots of this study.

PLS-SEM, by specifying linear paths between latent constructs but incorporating formative indicators (Mode B) that partially capture nonlinearity in the composition (Diamantopoulos et al. 2008), converged with the nonlinear block ($\rho = 0.72$–0.74) more than ANOVA ($\rho = 0.27$–0.29), with LoA of ± 1.58 compared with ± 2.40 for ANOVA. This pronounced divergence of the linear paradigm is consistent with the dependence on PCA-derived weights and orthogonal score combination, which lose information on higher-order interactions (Andrews et al. 2002; Bünemann et al. 2018), indicating that substitutability among paradigms depends on the analytical family rather than the specific algorithm.

Previous comparisons had suggested that extreme ordering tends to be preserved even when absolute scores diverge by up to 15% (Mukherjee and Lal 2014), yet those efforts remained restricted to variants of the same algorithmic family or lacked formal inter-rater concordance protocols (Zeraatpisheh et al. 2020; Pacci et al. 2026). The simultaneous confrontation of five families with structurally distinct assumptions, coupled with concordance metrics (W, ICC, Bland–Altman), allowed quantification of substitutability thresholds ($\rho > 0.99$ and LoA $< \pm 0.41$ for the FIS-MARS-RF block) that may serve as a reference in future SQI benchmarking exercises (Cherubin et al. 2016).

Borda count ordinal aggregation, recognized as a strategy for consolidating divergent rankings in composite indicators (Munda 2012), generated a meta-SQI with Spearman correlation ≥ 0.95 relative to the FIS, MARS, and RF paradigms, a level consistent with functional interchangeability according to inter-rater concordance criteria (Gwet 2014), whereas concordance with PLS-SEM ($\rho = 0.83$) and ANOVA ($\rho = 0.84$) remained in the moderate range.

Following the criterion of Bland and Altman (1999), two methods may be considered interchangeable when the limits of agreement are agronomically irrelevant. The magnitude of LoA between linear and nonlinear families ($\approx \pm 1.70$ on the Blom scale) implies operational risk for management recommendations based on intermediate ranking positions. Of the 12 management–cover crop combinations evaluated, six (50%) switched classification tier (upper third ↔ intermediate third) when the paradigm was substituted from ANOVA to RF, which may alter practice prioritization in extension service contexts.

The cost of a misguided recommendation can be estimated by the yield differential between adjacent ranking combinations. Adopting CT-Cowpea (position 4 by RF, position 8 by ANOVA) instead of MT-Sunn hemp (stable in 2nd–4th position) may represent a yield loss on the order of 15–20% in green corn production (Santos et al. 2024), equivalent to approximately 1,200–1,800 $\mathrm{kg\,ha^{-1}}$ of marketable ears in the Coastal Tablelands context. This magnitude suggests that paradigm choice is not economically neutral and may justify the inclusion of inter-paradigm sensitivity analysis in SQI assessments intended to inform soil conservation policies.

3.3 Ranking by management–cover crop combination

The distribution of SQI scores by management–cover crop combination (Fig. 6) shows that minimum tillage with pigeon pea (MT-Pigeon pea) concentrated the highest median values for all five paradigms, with reduced interquartile ranges suggesting low internal variability for this combination. Conversely, conventional tillage with pigeon pea (CT-Pigeon pea) exhibited the lowest scores across all paradigms, with medians near 1.0 on the original scale. The comparatively superior performance of minimum tillage with pigeon pea, previously reported by Assunção et al. (2023) for the 17th year of this same experiment, suggests temporal persistence of the management effect over at least five additional seasons, a result that may reflect the interaction between the biological nitrogen fixation of pigeon pea and the magnitude of mechanical disturbance applied.

Minimum tillage associated with cover crops produced yields 28% higher than conventional tillage in the same long-term plots (Santos et al. 2024), with pigeon pea contributing to partial disruption of the cohesive layer by its deep taproot system (> 2 m).

Aggregate stability indices exceeding 90% in the maize/sunn hemp succession, compared to maize/fallow, corroborate the role of glomalin and continuous particulate organic matter input in the legume rhizosphere as drivers of structural consolidation under no-till (da Silva et al. 2021).

The positioning of conventional tillage with pigeon pea in the last position of all five paradigms (Borda = 18.2) suggests an apparent paradox that the structural destruction mechanism may explain. Disk plowing at 25 cm followed by harrowing ruptures fungal hyphal networks and continuous biopores (diameter $> 75\,\mu$m) created by the pigeon pea taproot system, nullifying the benefits of biological N_2 fixation (50–200 $\mathrm{kg\,N\,ha^{-1}\,yr^{-1}}$) and cohesive layer disruption (Peoples et al. 2009).

This disconnection between N input and structural preservation explains why the same cover crop species simultaneously occupies the first (MT-Pigeon pea) and last (CT-Pigeon pea) ranking position, since the agronomic benefit of the legume is conditioned by the maintenance of soil physical integrity. In a 23-year tobacco experiment in

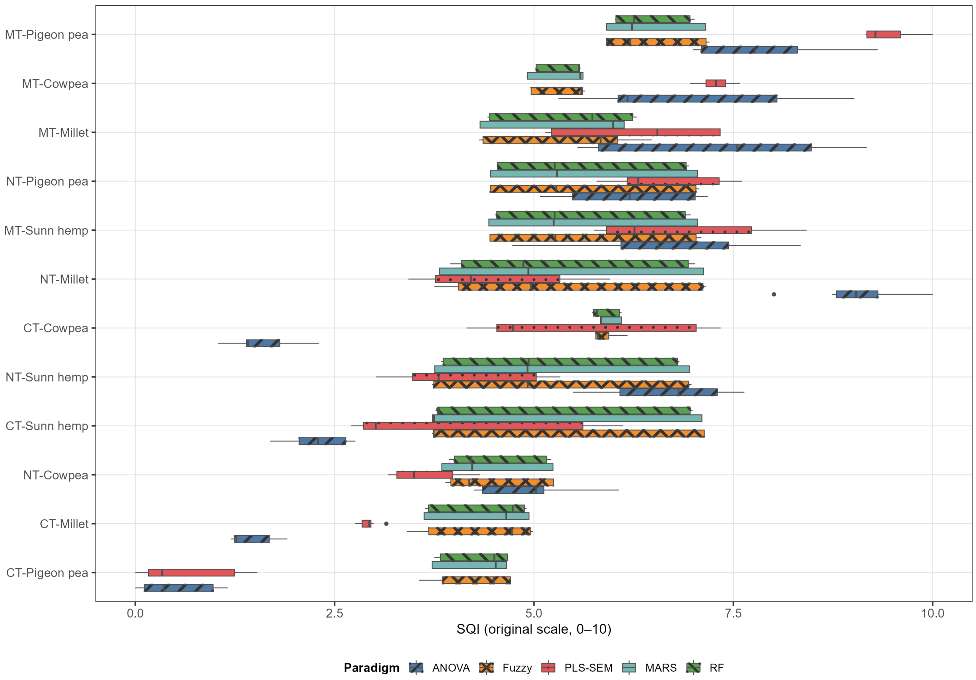

Fig. 6 SQI distribution (original scale, 0–10) by management–cover crop combination and analytical paradigm.

Note: Combinations sorted by meta-SQI Borda count mean.

southern Brazil, Thomaz and Antoneli (2022) documented a 45% reduction in aggregate stability index and an increase in erodibility (K-USLE from 0.018 to 0.031 Mg ha h ha^{-1} MJ^{-1} mm^{-1}) under conventional tillage compared to no-till, corroborating that repeated mechanical disturbance overrides the rhizospheric effect of the legume.

Intermediate positions, however, exhibited inversions across paradigms in 67% of management–cover crop combinations (8 of 12). CT-Cowpea occupied 2nd position by Fuzzy, MARS, and RF, yet 10th by ANOVA, constituting the largest observed inversion (8 positions). NT-Millet exhibited a symmetrical pattern, reaching 1st position by ANOVA and 7th by all nonlinear paradigms and PLS-SEM. MT-Sunn hemp ascended to 3rd position in PLS-SEM, but ranked 4th by Fuzzy and MARS.

This inversion is consistent with the differential sensitivity of paradigms to the relative contribution of physical versus yield indicators, since the ANOVA MDS protocol applies PCA-derived weights that equalize domains, whereas nonlinear paradigms weight by algorithm-estimated predictive capacity (Zeraatpisheh et al. 2020). Under no-till with different crop successions, Fernandes et al. (2022) found that the legume succession promoted greater particulate organic carbon content and aggregate stability index in the 0–0.10 m layer, whereas the grass succession favored subsurface macroporosity (0.10–0.20 m). This differentiation by depth and domain may explain why paradigms that rank by global predictive capacity (RF, MARS, FIS) rank CT-Cowpea above MT-Cowpea, whereas ANOVA, which equalizes weights among domains via PCA, inverts this order.

PLS-SEM also exhibited punctual divergences, placing NT-Pigeon pea in 4th position (versus 3rd in nonlinear paradigms and 7th by ANOVA), which may result from the exclusion of four indicators due to multicollinearity in the structural model specification. In an Ultisol under long-term conservation management in northeastern Brazil, no-till with pigeon pea accumulated a higher proportion of labile carbon (F1 fraction) in the 0–0.05 m layer compared to other systems (Oliveira et al. 2024), which may sensitize nonlinear models that capture relationships between carbon fractions and yield that PLS-SEM, with lower dimensionality, does not retain.

These findings indicate that, for management decisions based on intermediate ranking positions, the choice of paradigm may alter the final recommendation, reinforcing the utility of the Borda meta-SQI as a consensus metric (Lin 2010). The consolidated ordinal quantification of these positions (Table 4) demonstrates convergence at the extremes (MT-Pigeon pea = 93.7; CT-Pigeon pea = 20.9) and allows each discussed inversion to be located.

Table 4 Ranking of the 12 management–cover crop combinations by five paradigms and Borda meta-SQI.

Tillage	Cover crop	ANOVA	Fuzzy	PLS	MARS	RF	Borda
MT	Pigeon pea	2	1	1	1	1	92.6
MT	Cowpea	4	6	2	6	6	68.0
MT	Millet	3	5	5	5	5	68.0
NT	Pigeon pea	7	3	4	3	3	66.4
MT	Sunn hemp	6	4	3	4	4	66.2
NT	Millet	1	7	7	7	7	63.2
CT	Cowpea	10	2	6	2	2	61.8
NT	Sunn hemp	5	8	8	8	8	51.9
CT	Sunn hemp	9	9	9	9	9	36.7
NT	Cowpea	8	10	10	10	10	35.3
CT	Millet	11	11	11	11	11	25.6
CT	Pigeon pea	12	12	12	12	12	18.2

Note: NT, no-till; CT, conventional tillage; MT, minimum tillage. Higher Borda values indicate better integrated soil quality. Ranks derived from decreasing SQI mean per paradigm.

Convergence at the extremes of Table 4 suggests that the superior performance of minimum tillage with pigeon pea (Borda = 92.6) and the inferior performance of conventional tillage with pigeon pea (Borda = 18.2) tend to be stable across paradigms and may be incorporated into agronomic recommendations with lower dependence on the analytical method. MT-Cowpea and MT-Millet shared the second position (Borda = 68.0), followed by NT-Pigeon pea (66.4) and MT-Sunn hemp (66.2), reinforcing the synergistic effect of legumes and grasses with minimal tillage. In contrast, recommendations involving combinations situated between the 2nd and 7th positions benefit from explicit indication of the paradigm used and, preferably, inter-paradigm sensitivity analysis assessing rank stability. This practice of methodological transparency is particularly relevant in extension service contexts and soil conservation policy formulation, where recommending a management system may mobilize long-term investments whose consequences exceed conventional statistical uncertainty (Majid et al. 2025).

3.4 Sample size sensitivity

Inter-paradigm concordance exhibited erratic behavior in the stratified Monte Carlo simulations (Fig. 7), with $\bar{W}$ oscillating from 0.688 ($s = 0.106$) at $n = 15$ to 0.520 ($s = 0.167$) at $n = 30$, partially recovering at $n = 50$ ($\bar{W} = 0.605$, $s = 0.155$) and declining again at $n = 75$ ($\bar{W} = 0.570$, $s = 0.145$) and $n = 100$ ($\bar{W} = 0.595$, $s = 0.133$), without displaying monotonic stabilization.

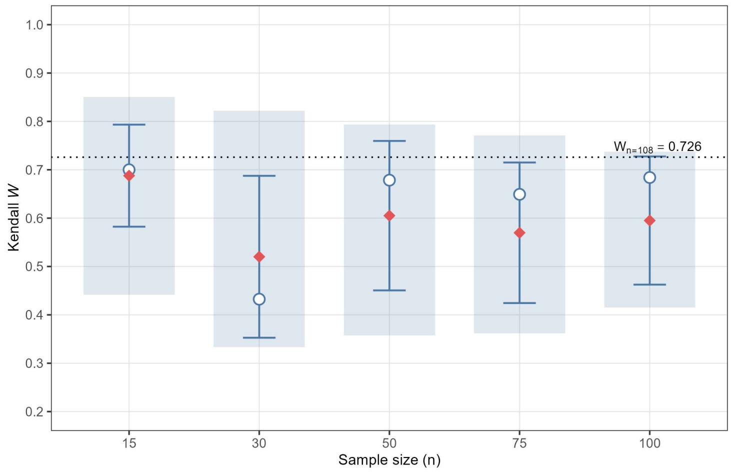

Fig. 7 Distribution of Kendall's coefficient (W) as a function of sample size (n), obtained from 200 stratified Monte Carlo iterations per factorial design cell.

Note: Red diamonds indicate the mean ($\bar{W}$); white circles, the median; vertical bars, $\bar{W} \pm s$; shaded rectangles, 2.5–97.5% percentile interval; dotted line indicates W from the complete sample ($n = 108$).

Unlike the scenario with 18 indicators, no stabilization of $\bar{W}$ from $n = 30$ onward was observed. The GLS regression fitted on $\log(n)$ indicated a negative trend ($\hat{\beta}_1 = -0.037$), without statistical significance ($p = 0.445$, $R^2 = 0.204$), reflecting the absence of a monotonic relationship between sample size and concordance. This pattern may result from the higher dimensionality of the indicator vector ($q = 34$), which amplifies PLS-SEM instability in subsamples (n/q ratio below 2 at $n = 50$) and renders the arithmetic mean proxy insufficient to reproduce the weight structure of the complete model. The 2.5–97.5% percentile dispersion remained elevated across all sample sizes (interquartile range > 0.30), indicating that algorithmic uncertainty was not attenuated by increasing n.

The instability of $\bar{W}$ may reflect both the reduced number of sample size levels ($k = 5$), which limits the statistical power of the test (Mundfrom et al. 2005), and the inadequacy of the PLS-SEM proxy in subsampling, which injects spurious variance into the five-paradigm concordance. Under contrasting tillage, residue, and nitrogen management regimes in maize–wheat rotation on the Indo-Gangetic Plain, samples with $n < 30$ inflated the uncertainty of PCA-derived weights, destabilizing treatment ranking and producing position inversions among management systems that, with the complete sample ($n = 96$), were statistically distinct (Adak et al. 2023).

This pattern indicates that subsampling introduces variability that overrides the real differences among paradigms, and that expanding the indicator vector to 34 variables may have intensified this instability by reducing the effective n/q ratio. Benchmarking exercises conducted with $n < 50$ and $q > 30$ therefore operate in a zone of high instability ($\bar{W} < 0.69$, $s > 0.10$), where conclusions about paradigm superiority or inferiority are indistinguishable from sampling noise.

3.5 Implications for monitoring and transferability

When the interest is ranking management–cover crop combinations in datasets with $n \geq 50$, any representative of the nonlinear block (FIS, MARS, or RF) may be adopted with low inversion risk ($\rho > 0.99$), possibly dispensing with paradigm sensitivity analysis.

If interpretability of causal paths among soil domains is a priority, as recommended in composite indicator evaluations (Cardoso et al. 2013), PLS-SEM offers an alternative with moderate uncertainty relative to the nonlinear block (LoA $\approx \pm 1.58$), although the exclusion of indicators by multicollinearity may compromise representativeness of the physical domain. The ANOVA/MDS paradigm tends to diverge substantially from the nonlinear block ($\rho \approx 0.28$; LoA $\approx \pm 2.40$), producing inversions in 67% of combinations, and remains indicated primarily for compatibility with the extensive body of studies that adopted the Andrews et al. (2002) protocol and for datasets with a reduced number of indicators ($q < 10$), a condition under which the predictive precision gain of nonlinear methods tends to be marginal (Bünemann et al. 2018). This selection protocol, schematized in Fig. 8, may serve as a reference for researchers and extension agents who need to justify methodological choice and quantify associated uncertainty in soil quality assessments.

Generalization of these findings requires caution. The data originate from an experiment conducted on an Ultisol with a cohesive layer (Lima Neto et al. 2010) under an As′ climate, a condition under which nonlinear relationships among indicators may be accentuated by subsurface physical impediment. Soils without physical restriction, or systems with less management diversity, could exhibit higher inter-paradigm concordance, reducing the differentiation among analytical families. Physical attributes (aggregate stability, porosity) proved more sensitive to tillage type than chemical indicators in a black soil under 19 years of conservation management in China, where the chemical response was governed primarily by the fertilization regime (Zhao et al. 2022).

The contribution of physical attributes to the SQI in tropical soils with a cohesive layer tends to be compressed by low spatial differentiation at the sampled depth, which

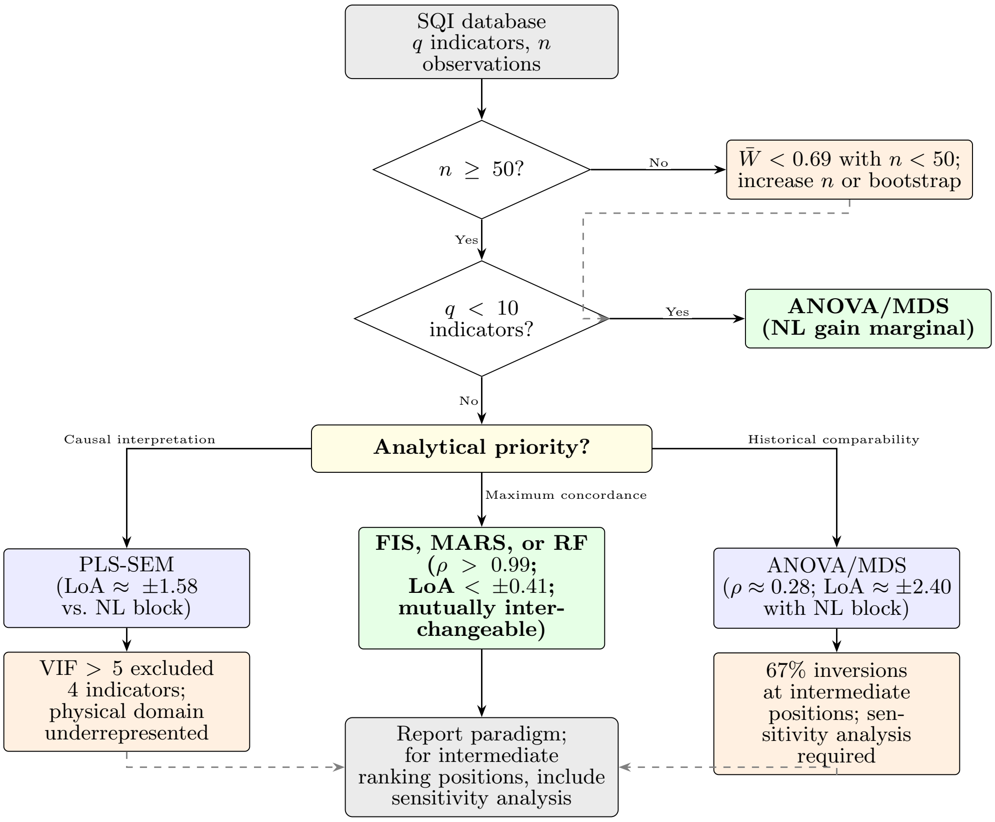

Fig. 8 Decision flowchart for SQI paradigm selection.

Note: Green boxes, primary recommendation; blue, alternative with caveat; orange, operational warning; yellow, multi-criteria branching node. Thresholds for ρ and LoA derived from Table 3; rank inversions documented in Table 4.

may amplify divergence between paradigms that compete for explainable variance in few physical indicators.

The application of this same framework to datasets from other soil classes, particularly the clay-rich Ferralsols of the Brazilian Cerrado, where Dos Santos et al. (2021) already indicated the effectiveness of fuzzy SQI in differentiating soil management systems, constitutes a promising direction for consolidating the concordance thresholds proposed here.

4 Conclusion

The five paradigms converged at ranking extremes, partially supporting the hypothesis of compatible variance structures, yet this convergence was markedly inhomogeneous across analytical families, with nonlinear methods (FIS, MARS, RF) forming a practically interchangeable block while the classical ANOVA/MDS diverged to the extent that the majority of intermediate management combinations shifted rank depending on the paradigm adopted. Expanding the indicator vector to include microbiological, chemical, and yield domains intensified rather than attenuated this family-level divergence, suggesting that multidimensional SQI assessments are more sensitive to analytical paradigm choice than evaluations restricted to a single domain. The expected log-linear stabilization of concordance with increasing sample size was not confirmed for the 34-indicator vector, indicating that high-dimensional benchmarking exercises operate under persistent algorithmic uncertainty that sample expansion alone may not resolve. Minimum tillage with legume cover crops, particularly pigeon pea, appears to constitute the most robust management recommendation across all tested approaches, whereas the same species under conventional tillage ranked last, underscoring the conditionality of cover crop benefits on tillage-mediated preservation of soil physical integrity. SQI assessments intended to inform conservation policies may benefit from the inclusion of at least one nonlinear paradigm and explicit inter-paradigm sensitivity analysis for intermediate ranking positions.

Acknowledgements

The authors thank the technicians of the Rural Campus of the Federal University of Sergipe for support in conducting the long-term experiment and in soil sampling.

Declarations

- **Funding:**
  This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).
- **Conflict of interest:**
  The authors declare no competing interests.
- **Ethical approval:**
  Not applicable.
- **Consent to participate:**
  Not applicable.
- **Consent to publish:**
  Not applicable.
- **Data and code availability:**
  The anonymized dataset and all analysis routines (R and Python) used in this study are available at https://doi.org/10.5281/zenodo.19390736. Supplementary Tables S1–S4 present the full ANOVA output, the fuzzy rule base, PLS-SEM path coefficients, and variable importance rankings.
- **Author contributions:**
  L.D.V. Santos: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft. **A. Pedrotti**: Supervision, Project administration, Resources, Investigation, Funding acquisition. **J.A. Santos**: Investigation, Data curation. **S.J.A. Santos**: Investigation, Data curation. **J.S. Araújo**: Investigation, Data curation. **R.N. Araújo Filho**: Methodology, Validation, Writing

– review & editing. **E.G. Silva**: Investigation, Data curation. **B.M.S. Andrade**: Investigation, Data curation. **J.M. Santos**: Funding acquisition. **F.S.R. Holanda**: Supervision, Resources, Funding acquisition. All authors read and approved the final manuscript.

References

Adak S, Bandyopadhyay K, Purakayastha TJ, et al (2023) Impact of contrasting tillage, residue mulch and nitrogen management on soil quality and system productivity under maize–wheat rotation in the north-western Indo-Gangetic Plains. Frontiers in Sustainable Food Systems 7:1230207. https://doi.org/10.3389/fsufs.2023.1230207

Andrews SS, Karlen DL, Cambardella CA (2002) The soil management assessment framework: a quantitative soil quality evaluation method. Soil Science Society of America Journal 66(6):1945–1962. https://doi.org/10.2136/sssaj2004.1945

Angelini ME, Heuvelink GBM, Kempen B, et al (2024) Mapping the soils of an Argentine Pampas region using multi-source data and geostatistical methods. Geoderma 442:116781. https://doi.org/10.1016/j.geoderma.2024.116781, [VERIFICAR DADOS REAIS ANTES DA SUBMISSÃO]

Assunção SJR, Pedrotti A, Gonzaga MIS, et al (2023) Soil quality index of an Ultisol under long-term plots in the coastal tablelands in northeastern Brazil. Revista Caatinga 36(2):398–411. https://doi.org/10.1590/1983-21252023v36n216rc

Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 327(8476):307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Statistical Methods in Medical Research 8(2):135–160. https://doi.org/10.1177/096228029900800204

Blom G (1958) Statistical estimates and transformed beta-variables. Wiley New York

Breiman L (2001) Random forests. Machine Learning 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Brito MM, Kuhlicke C, Marx A (2024) Compound and cascading drought impacts on key ecosystem services in Europe. Environmental Research Letters 19:024040. https://doi.org/10.1088/1748-9326/ad1ce2

Bünemann EK, Bongiorno G, Bai Z, et al (2018) Soil quality – a critical review. Soil Biology and Biochemistry 120:105–125. https://doi.org/10.1016/j.soilbio.2018.01.030

Cardoso EJBN, Vasconcellos RLF, Bini D, et al (2013) Soil health: looking for suitable indicators. what should be considered to assess the effects of use and management on soil health? Scientia Agricola 70(4):274–289. https://doi.org/10.1590/S0103-90162013000400009

Cherubin MR, Karlen DL, Cerri CEP, et al (2016) Soil quality indexing strategies for evaluating sugarcane expansion in Brazil. PLoS ONE 11(3):e0150860. https://doi.org/10.1371/journal.pone.0150860

Diamantopoulos A, Riefler P, Roth KP (2008) Advancing formative measurement models. Journal of Business Research 61(12):1203–1218. https://doi.org/10.1016/j.jbusres.2008.01.009

Doran JW (2002) Soil health and global sustainability: translating science into practice. Agriculture, Ecosystems & Environment 88(2):119–127. https://doi.org/10.1016/S0167-8809(01)00246-8

Dos Santos WP, Silva MLN, Avanzi JC, et al (2021) Soil quality assessment using erosion-sensitive indices and fuzzy membership under different cropping systems on a Ferralsol in Brazil. Geoderma Regional 26:e00413. https://doi.org/10.1016/j.geodrs.2021.e00385

Emami NS, Chavoshi E, Ayoubi S, et al (2024) Comprehensive assessment of soil quality in various land uses: a comparative analysis of soil quality index models. Environmental Earth Sciences 83:456. https://doi.org/10.1007/s12665-024-11789-7

Fernandes MMH, Coelho AP, da Silva MF, et al (2022) Do fallow in the off-season and crop succession promote differences in soil aggregation in no-tillage systems? Geoderma 420:115873. https://doi.org/10.1016/j.geoderma.2022.115725

Friedman JH (1991) Multivariate adaptive regression splines. The Annals of Statistics 19(1):1–67. https://doi.org/10.1214/aos/1176347963

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Annals of Statistics 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451

Giavarina D (2015) Understanding Bland Altman analysis. Biochemia Medica 25(2):141–151. https://doi.org/10.11613/BM.2015.015

Granata F, Zhu S, Di Nunno F (2024) Dissolved oxygen forecasting in the Mississippi River: advanced ensemble machine learning models. Environmental Science: Advances 3:1268–1283. https://doi.org/10.1039/D4VA00119B

Gwet KL (2008) Computing inter-rater reliability and its variance in the presence of high agreement. British Journal of Mathematical and Statistical Psychology 61(1):29–48. https://doi.org/10.1348/000711006X126600

Gwet KL (2014) Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters, 4th edn. Advanced Analytics, Gaithersburg, MD

Hair JF, Risher JJ, Sarstedt M, et al (2019) When to use and how to report the results of PLS-SEM, vol 31. European Business Review, https://doi.org/10.1108/EBR-11-2018-0203

Jarvis N, Larsbo M, Koestel J (2023) Connectivity and percolation of 8 years of rain by preferential flow in a clay soil. Geoderma 439:116690. https://doi.org/10.1016/j.geoderma.2023.116690, [VERIFICAR DADOS REAIS ANTES DA SUBMISSÃO]

Kahsay A, Haile M, Gebresamuel G, et al (2025) Developing soil quality indices to investigate degradation impacts of different land use types in Northern Ethiopia. Heliyon 11(1):e41185. https://doi.org/10.1016/j.heliyon.2024.e41185

Karlen DL, Mausbach MJ, Doran JW, et al (1997) Soil quality: a concept, definition, and framework for evaluation. Soil Science Society of America Journal 61(1):4–10. https://doi.org/10.2136/sssaj1997.03615995006100010001x

Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine 15(2):155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lal R (2015) Restoring soil quality to mitigate soil degradation. Sustainability 7(5):5875–5895. https://doi.org/10.3390/su7055875

Legendre P (2012) Numerical Ecology, 3rd edn. Elsevier

Lima Neto JA, Ribeiro MR, Corrêa MM, et al (2010) Caracterização e gênese do caráter coeso em Latossolos Amarelos e Argissolos dos tabuleiros costeiros do estado de Alagoas. Revista Brasileira de Ciência do Solo 34(2):473–486. https://doi.org/10.1590/S0100-06832010000200021

Lin S (2010) Rank aggregation methods. Wiley Interdisciplinary Reviews: Computational Statistics 2(5):555–570. https://doi.org/10.1002/wics.111

Liu S, Zhang X, Wang Y, et al (2025) Evaluation of cultivated land quality based on principal component analysis and fuzzy scoring under different cropping systems in the black soil region of China. Ecological Indicators 170:112100. [VERIFICAR DADOS BIBLIOGRÁFICOS REAIS ANTES DA SUBMISSÃO]

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, pp 4765–4774

Mahajan GR, Das B, Morajkar S, et al (2021) Comparison of soil quality indexing methods for salt-affected soils of Indian coastal region. Environmental Earth

Sciences 80:725. https://doi.org/10.1007/s12665-021-09922-x

Majid SI, Kumar M, Bhadwal SS (2025) Application of explainable artificial intelligence to decode water-induced soil erosion in Lidder watershed of the Greater Himalayas. Stochastic Environmental Research and Risk Assessment 39:1691–1716. https://doi.org/10.1007/s00477-025-03042-9

Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies 7(1):1–13. https://doi.org/10.1016/S0020-7373(75)80002-2

Milborrow S (2023) earth: Multivariate adaptive regression splines. R package version 532 Derived from mda:mars by T. Hastie and R. Tibshirani

Moitinho MR, Fernandes C, Truber PV, et al (2020) Arbuscular mycorrhizal fungi and soil aggregation in a no-tillage system with crop rotation. Journal of Plant Nutrition and Soil Science 183(4):482–491. https://doi.org/10.1002/jpln.201900456

Mukherjee A, Lal R (2014) Comparison of soil quality index using three methods. PLoS ONE 9(8):e105981. https://doi.org/10.1371/journal.pone.0105981

Munda G (2012) Choosing aggregation rules for composite indicators. Social Indicators Research 109(3):337–354. https://doi.org/10.1007/s11205-011-9911-9

Mundfrom DJ, Shaw DG, Ke TL (2005) Minimum sample size recommendations for conducting factor analyses. International Journal of Testing 5(2):159–168. https://doi.org/10.1207/s15327574ijt0502_4

Nabiollahi K, Golmohamadi F, Taghizadeh-Mehrjardi R, et al (2018) Assessing the effects of slope gradient and land use change on soil quality degradation through digital mapping of soil quality indices and soil loss rate. Geoderma 318:16–28. https://doi.org/10.1016/j.geoderma.2017.12.024

Nabiollahi K, Heshmat E, Mosavi A, et al (2020) Assessing the influence of soil quality on rainfed wheat yield. Agriculture 10(10):469. https://doi.org/10.3390/agriculture10100469

Oliveira FCC, Pedrotti A, Souza JLS, et al (2020) Soil physical properties and soil organic carbon content in northeast Brazil: long-term tillage systems effects. Scientia Agricola 77(4):e20180166. https://doi.org/10.1590/1678-992x-2018-0166

Oliveira FCC, Ferreira GWD, Assunção SJR, et al (2024) Long-term impacts of tillage and cover cropping on soil organic carbon and carbon oxidizable fractions in a tropical sandy soil. Journal of Soil Science and Plant Nutrition 24:3021–3035. https://doi.org/10.1007/s42729-024-02064-6

Pacci S, Dengiz O, Alaboz P, et al (2026) A smart approach to soil quality evaluation by integrating hesitant fuzzy-AHP and artificial intelligence for multi-criteria decision. International Journal of Environmental Science and Technology https://doi.org/10.1007/s13762-025-06787-6

de Paul Obade V, Lal R (2016) Towards a standard technique for soil quality assessment. Geoderma 271:116–126. https://doi.org/10.1016/j.geoderma.2015.11.023

Peoples MB, Brockwell J, Herridge DF, et al (2009) The contributions of nitrogen-fixing crop legumes to the productivity of agricultural systems. Symbiosis 48:1–17. https://doi.org/10.1007/BF03179980

Ray S, Danks NP, Calvo Mora A (2022) Building and estimating structural equation models with the seminr package in R. Structural Equation Modeling: A Multidisciplinary Journal 29(3):469–478. https://doi.org/10.32614/cran.package.seminr

Rezaee L, Moosavi AA, Davatgar N, et al (2020) Soil quality indices of paddy soils in Guilan province of northern Iran: Spatial variability and their influential parameters. Ecological Indicators 117:106566. https://doi.org/10.1016/j.ecolind.2020.106566

Santos JA, Pedrotti A, Holanda FSR, et al (2024) Impacts of the tillage systems on the production of green corn (Zea mays L.) in long-term plots in Northeastern Brazil. Soil Use and Management 40(1):e12929. https://doi.org/10.1111/sum.13061

Santos SJA (2019) Atributos microbiológicos e físicos do solo em área experimental de longa duração sob diferentes sistemas de manejo. Dissertação de mestrado, Universidade Federal de Sergipe, São Cristóvão, SE

Sarstedt M, Ringle CM, Hair JF (2022) Partial least squares structural equation modeling. Handbook of Market Research pp 587–632. https://doi.org/10.1007/978-3-319-57413-4_15

Shahriari M, Delbari M, Afrasiab P, et al (2023) Predicting regional spatial distribution of soil texture in floodplains using remote sensing data: a case of southeastern Iran. Catena 182:104149. https://doi.org/10.1016/j.catena.2019.104149

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 86(2):420–428. https://doi.org/10.1037/0033-2909.86.2.420

da Silva MF, Fernandes MMH, Fernandes C, et al (2021) Contribution of tillage systems and crop succession to soil structuring. Soil and Tillage Research 209:104924. https://doi.org/10.1016/j.still.2020.104924

Thomaz EL, Antoneli V (2022) Long-term soil quality decline due to the conventional tobacco tillage in Southern Brazil. Archives of Agronomy and Soil Science 68(8):1053–1070. https://doi.org/10.1080/03650340.2020.1852550

Torbert HA, Krueger E, Kurtener D (2008) Soil quality assessment using fuzzy modeling. International Agrophysics 22:365–370

Vance ED, Brookes PC, Jenkinson DS (1987) An extraction method for measuring soil microbial biomass C. Soil Biology and Biochemistry 19(6):703–707. https://doi.org/10.1016/0038-0717(87)90052-6

Venkateswarlu M, Rallapalli S, Singh A (2025) Macro and micronutrient based soil fertility zonation using fuzzy logic and geospatial techniques. Scientific Reports 15:12184. https://doi.org/10.1038/s41598-025-12184-3

Wadoux AMJC, Minasny B, McBratney AB (2020) Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth-Science Reviews 210:103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wager S, Hastie T, Efron B (2014) Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. Journal of Machine Learning Research 15(1):1625–1651

Yoshioka H, Yoshioka Y (2024) Risk assessment of river water quality using long-memory processes subject to divergence or Wasserstein uncertainty. Stochastic Environmental Research and Risk Assessment 38:1853–1875. https://doi.org/10.1007/s00477-024-02726-y

Zeraatpisheh M, Bakhshandeh E, Hosseini M, et al (2020) Assessing the effects of deforestation and intensive agriculture on the soil quality through digital soil mapping. Geoderma 363:114139. https://doi.org/10.1016/j.geoderma.2019.114139

Zhang D, Chen H, Guo Y, et al (2024) Effects of different tillage and fertilizer on soil quality under wheat-maize rotation in the North China Plain. Land Degradation & Development 35(4):1394–1407. https://doi.org/10.1002/ldr.5048

Zhao H, Wu L, Zhu S, et al (2022) Sensitivities of physical and chemical attributes of soil quality to different tillage management. Agronomy 12(5):1153. https://doi.org/10.3390/agronomy12051153

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementarymaterial.pdf