Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Tree proximity dominates meter-scale microbial activity beneath mature Faidherbia albida trees on a sandy tropical soil in Zambia: evidence from spatial regression and blocked validation

Elijah Phiri[1]

ephiri@unza.zm

University of Zambia

Patience Chanda[1]

University of Zambia

Alice Mutiti Mweetwa[1]

University of Zambia

Research Article

Keywords: agroforestry, Faidherbia albida, fertility islands, microbial respiration, spatial cross-validation, Zambia

Posted Date: August 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10461765/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Fertility islands surrounding mature *Faidherbia albida* trees are often inferred from under-canopy and open-field comparisons, but these contrasts do not separate continuous distance gradients from inherited spatial structure. We re-analysed 169 georeferenced surface-soil observations collected on a 2-m grid around five mature trees in a 24 × 24 m organically managed field near Lusaka, Zambia. Microbial activity (MA) was modelled on the natural-log scale, and sand, silt and clay were represented by isometric log-ratio coordinates. Inference combined hierarchical regression with heteroskedasticity-consistent standard errors, a spatial-error model, Moran statistics, five-fold spatially blocked cross-validation, grouped permutation importance, block-bootstrap mediation and ordinary kriging. Mean MA was 4.19 within 0–2 m of a tree and 2.50 beyond 6 m, representing a 68% near-tree elevation. Log(MA) was spatially autocorrelated. Standardised tree distance remained negatively associated with log(MA) in both the additive model and the spatial-error model, whereas soil organic carbon (SOC) showed a positive but model-sensitive association. The block-bootstrap indirect association through SOC excluded zero, but adding SOC, texture, pH or a random forest did not improve prediction in held-out spatial blocks. Tree proximity therefore captured a robust but modestly predictive biological gradient. SOC was consistent with a partial pathway, while texture and pH provided no independent support. Because the study was observational and conducted in one field, the results identify spatial associations rather than causal tree effects.

1 Introduction

Faidherbia albida (Delile) A. Chev. is a characteristic fertilizer tree of African parklands and smallholder cropping systems. Its reverse phenology—loss of foliage during the rainy cropping season and canopy development during the dry season—can reduce competition for light during crop growth while sustaining litter, fodder and below-ground inputs during other parts of the year (Barnes and Fagg 2003; Boffa 1999). These traits underpin its role in “evergreen agriculture” and broader efforts to rebuild soil fertility with trees in maize-based farming systems (Akinnifesi et al. 2010; Garrity et al. 2010).

The effects of an individual tree are spatially heterogeneous. Litterfall, fine-root turnover, rhizodeposition, biological nitrogen cycling, stemflow, faunal deposition and reduced erosive transport may enrich carbon, nutrients and biological activity near a stem or canopy. Root competition, cultivation, microtopography and inherited soil variation can generate counter-gradients. Consequently, “fertility islands” are rarely perfect circular halos. A meta-analysis of *F. albida* systems found broad under-canopy enrichment but also strong variation in magnitude and spatial extent among tree sizes, soils and management contexts (Sileshi 2016). Zambian studies similarly reported higher organic carbon and nutrients closer to mature trees and changes in microbial communities and nitrogen mineralization under canopies and litter inputs (Umar et al. 2013; Yengwe et al. 2018a, b).

Most field comparisons classify locations as under canopy or in the open. That design is operationally simple but embeds an arbitrary boundary, ignores non-linear distance decay and treats closely spaced

samples as independent. A spatially explicit analysis should distinguish three related questions: whether MA changes with tree proximity; whether soil properties such as SOC and texture account for that association; and whether fitted relationships transfer to new locations rather than merely explaining spatially clustered observations. Global and local Moran statistics describe dependence and clustering (Anselin 1995; Moran 1950); spatial-error models account for residual dependence (Anselin 1988), and spatially blocked cross-validation provides a more realistic estimate of geographic predictive performance than random folds (Roberts et al. 2017; Valavi et al. 2019).

Soil texture requires additional care because sand, silt and clay are closed components that sum to approximately 100%. Entering all three fractions as ordinary predictors causes exact or near-exact collinearity and can produce misleading interpretations. Isometric log-ratio (ILR) coordinates map the composition into an unconstrained Euclidean space while preserving relative information (Aitchison 1986; Egozcue et al. 2003). This provides a coherent basis for asking whether texture explains MA beyond tree proximity and SOC.

This study used an intensively sampled Zambian field to test four hypotheses: (H1) MA declines with increasing distance from the nearest mature *F. albida* tree; (H2) SOC is positively associated with MA after adjustment for proximity and texture; (H3) texture composition contributes independent explanatory or predictive information; and (H4) models that combine proximity, texture and SOC outperform a proximity-only model under spatially blocked validation. We combined robust regression, spatial-error modelling, observational mediation and cross-validated prediction, treating ordinary kriging as a descriptive mapping tool rather than a substitute for mechanistic inference.

2 Materials and methods

2.1 Study site and sampling design

The study was conducted at Kasisi Agricultural Training Centre in Lusaka Province, Zambia (approximately 15°24' S, 28°48' E). The site has a tropical seasonal climate with rainfall concentrated in the main rainy season and was managed under organic farming. The sampled field contained five mature *F. albida* trees older than approximately 25 years, aligned broadly along the centre of the field. Because the analysis concerns one field and one tree alignment, it is a high-resolution spatial case study rather than a replicated tree-treatment experiment.

A 24 × 24 m plot was sampled at 2-m intervals in the x and y directions, producing a regular 13 × 13 grid (n = 169). Surface soil was collected from approximately 0–0.20 m depth. Euclidean distance from each sampling point to every tree was calculated and the minimum retained as distance to the nearest tree. For descriptive comparisons, points were grouped into 0–2 m (n = 15), > 2–4 m (n = 26), > 4–6 m (n = 28) and > 6 m (n = 100) bands. The supplied tree-coordinate file placed all five stems at x = 12 m, with y coordinates of 0.0, 6.0, 11.6, 15.2 and 23.6 m; these coordinates were used directly without transposition. The sampling configuration is shown in Fig. 1.

2.2 Soil measurements

Air-dried soil samples were gently crushed and passed through a 2-mm sieve. Sub-samples used for soil organic carbon analysis were further ground to < 0.25 mm. All analytical results were corrected to an oven-dry soil basis determined at 105°C for 24 h.

2.2.1 Soil pH

Soil pH was measured potentiometrically in a 1:2.5 soil-to-deionised-water suspension following standard procedures for soil pH determination (Thomas 1996). A 20-g soil sample was mixed with 50 mL of deionised water and allowed to equilibrate for 30 min before measurement with a calibrated glass-combination electrode. Samples were measured in duplicate, and discrepant measurements were repeated.

2.2.2 Soil organic carbon

Soil organic carbon was determined using a modified Walkley–Black wet-oxidation method with spectrophotometric quantification (Walkley and Black 1934; Heanes 1984). A 0.5-g finely ground soil sample was oxidised with potassium dichromate and concentrated sulphuric acid. Following dilution and the addition of phosphoric acid and sodium fluoride, the remaining dichromate was measured at 590 nm. Carbon concentrations were obtained from a calibration curve, corrected using a factor of 1.33 and expressed as g C kg^{-1} oven-dry soil.

2.2.3 Particle-size distribution

Particle-size distribution was determined using the Bouyoucos hydrometer method following dispersion with 5% sodium hexametaphosphate (Bouyoucos 1962). A 50-g soil sample was dispersed in a 1-L sedimentation cylinder, and hydrometer readings were taken after 40 s and 2 h. Sand was recovered by wet sieving through a 53-µm sieve, clay was estimated from the 2-h hydrometer reading, and silt was calculated by difference. The fractions were classified according to USDA particle-size limits and expressed as percentages of oven-dry mineral soil.

2.2.4 Microbial respiration and activity

Microbial activity was assessed by measuring basal respiration using a closed static-incubation method adapted from Anderson (1982). Thirty grams of soil were adjusted to 50% water-holding capacity, pre-incubated for seven days and placed in sealed jars containing 10 mL of 0.5 M NaOH to trap evolved CO_2. The jars were incubated in darkness at 22 ± 1°C for 14 days. The NaOH traps were replaced after 2, 4, 6, 10 and 14 days, and the absorbed CO_2 was determined by titration with 0.1 M HCl following carbonate precipitation with $BaCl_2$.

Cumulative respiration was expressed as µg $C–CO_2$ g^{-1} oven-dry soil, while mean daily respiration was calculated by dividing cumulative respiration by 14 days. The microbial activity index was calculated as:

$MA = R_{cum} / SOC$

where (Rcum) is cumulative respiration and SOC is soil organic carbon, after converting both variables to compatible mass units. The index represents microbial respiration per unit of soil organic carbon.

2.2.5 Quality assurance and quality control

Each analytical batch included procedural blanks, reference soils and duplicate samples. Instruments were routinely calibrated using certified standards, and at least one duplicate was analysed for every ten samples. Measurements outside the specified acceptance limits were repeated. A subset of samples was independently analysed using dry combustion for SOC, laser diffraction for particle-size distribution and infrared gas analysis for respiration to verify analytical performance.

2.3 Data preparation and compositional texture coordinates

MA was strictly positive and right-skewed; the primary response was therefore ln(MA). Raw MA was retained for descriptive summaries, plots, and sensitivity analysis. Because all texture fractions were positive, two orthonormal ILR coordinates were calculated without zero replacement:

$ILR_1 = \sqrt{(2/3)} \ln[Sand / \sqrt{(Silt \times Clay)}]$

$ILR_2 = \sqrt{(1/2)} \ln(Silt / Clay)$

ILR_1 is the sand-versus-fine-particle balance, and ILR_2 is the silt-versus-clay balance. Predictors used in regression were standardised to have a mean of zero and a unit sample standard deviation. The linear distance term and its squared standardised value were included to allow curvature while retaining interpretable nested models.

2.4 Exploratory and spatial autocorrelation analyses

Descriptive statistics included mean, standard deviation, coefficient of variation, range, median, skewness, and the Shapiro–Wilk statistic. Pearson and Spearman coefficients summarized bivariate relations. Distance-band summaries were treated as descriptive evidence; continuous distance models were primary because the bands were unequal in size and spatially dependent.

A row-standardised eight-nearest-neighbor weight matrix represented local spatial connectivity. Global Moran's I was evaluated with 9,999 random permutations. Local Moran statistics classified significant high–high, low–low, high–low, and low–high locations at $p < 0.05$ (Anselin, 1995). A bivariate Moran statistic tested whether high SOC locations were surrounded by high ln(MA) values.

2.5 Hierarchical, non-linear and spatial-error models

A prespecified hierarchy of ordinary least-squares models was fitted to ln(MA): M0, intercept only; M1, standardised linear and quadratic distance; M2, M1 plus the two ILR texture coordinates; M3, M2 plus

SOC; M4, M3 plus pH; and M5, M4 plus SOC × distance and SOC × ILR_1 interactions. Heteroskedasticity-consistent HC3 standard errors were used for coefficient inference. Models were compared using AICc, adjusted R^2, in-sample RMSE, and Moran's I of residuals. Variance-inflation factors assessed collinearity.

A cubic B-spline with five degrees of freedom described the adjusted continuous distance response while holding SOC, texture and pH at their standardised means. To address residual spatial dependence, the full additive specification was refitted as a maximum-likelihood spatial-error model with the same nearest-neighbour matrix. The spatial parameter λ represents dependence in the error process; a significant λ indicates omitted spatially structured influences.

2.6 Spatially blocked prediction and grouped permutation importance

The plot was divided into approximately 6 × 6 m blocks (25 occupied blocks), and five-fold GroupKFold cross-validation held out complete blocks. Mechanistic candidates used a cubic spline for distance, standardised linear covariates, and ridge regression. Candidate sets were proximity only; proximity plus texture; plus SOC; plus pH; and a 1,000-tree random forest using all predictors (Breiman 2001). Performance was summarized by out-of-fold RMSE, MAE, and R^2 on ln(MA). This design estimates transfer to spatially separated locations and reduces optimistic error estimates caused by neighbouring training and test observations (Roberts et al. 2017).

Grouped permutation importance was calculated from the full mechanistic model. Within each held-out fold, tree distance, both texture coordinates jointly, SOC, or pH was permuted 100 times. The increase in out-of-fold RMSE measured dependence on that predictor group. Values near or below zero indicate no unique predictive contribution after adjustment for the other groups.

2.7 Observational mediation and sensitivity analyses

An observational mediation analysis evaluated the statistical pathway distance → SOC → ln(MA). The "a" model regressed standardised SOC on distance, its quadratic term, texture, and pH; the "b" model regressed ln(MA) on SOC and the same covariates. The indirect association was the product a × b. Uncertainty was estimated by 2,000 bootstrap resamples of whole spatial blocks. Because distance was not randomized and unmeasured confounding cannot be excluded, the result is interpreted as pathway-consistent association, not a causal mediation effect (Imai et al. 2010).

Sensitivity models compared ln(MA) with raw MA and ILR texture coordinates with the total fine fraction. Conclusions were considered robust when the sign and inference for distance and SOC were stable across these specifications and the spatial-error model.

2.8 Ordinary kriging

Ordinary kriging of ln(MA) was used for descriptive mapping. Spherical, exponential, and Gaussian variogram forms were compared by leave-one-out cross-validation. The lowest RMSE form was fitted to all observations, and predictions were back-transformed to the MA scale. Prediction variance was

mapped separately. Kriging was not used to estimate the tree effect because interpolation can reproduce spatial pattern without identifying its mechanism (Goovaerts 1999; Kravchenko 2003).

All analyses were conducted in a reproducible Python/Google Colab workflow using pandas, SciPy, statsmodels, scikit-learn, libpysal/esda/spreg and PyKrige. The random seed was 20260721. Statistical significance was assessed at $\alpha = 0.05$, but interpretation emphasized effect sizes, confidence intervals, residual spatial dependence, and blocked predictive performance. Figures were prepared in Python using Matplotlib.

3 Results

3.1 Distribution and bivariate structure

The field was sand fraction-dominated (mean = 80.94%, CV = 2.77%) and acidic (mean pH = 5.21; range 4.52–6.05). SOC averaged 1.11% and was comparatively homogeneous (CV = 17.0%). MA averaged 2.88, ranged from 0.15 to 7.61, and showed the greatest relative variability (CV = 47.3%) and positive skewness (0.84). The MA distribution was therefore summarized on its original scale but modelled after logarithmic transformation. Complete descriptive statistics are provided in Online Resource 1, Table S1, and the distribution is shown in Fig. 2a.

MA was negatively related to tree distance by Pearson correlation ($r = -0.377$) and Spearman correlation ($\rho = -0.336$). SOC was positively correlated with MA ($r = 0.263$) and negatively correlated with distance ($r = -0.326$). In contrast, the raw texture fractions displayed the expected closure-dominated relations: sand correlated strongly with clay ($r = -0.726$) and almost perfectly with ILR_1 ($r = 0.967$). These patterns justify modelling texture through ILR coordinates rather than entering the three fractions simultaneously. The correlation structure is shown in Fig. 3.

3.2 Tree-proximity gradient and adjusted distance response

Mean MA decreased from 4.19 within 0–2 m to 2.50 beyond 6 m (Table 1), equivalent to a 68% increase in the nearest band. Mean SOC declined from 1.28% to 1.07% (20% near-tree elevation), and the fine fraction declined from 20.22% to 18.67% (8.3%). pH did not show a monotonic distance pattern. The > 6 m band contained 59% of observations, but its lower median and compressed upper range were evident in the raw distribution. The distance-band distribution is shown in Fig. 2b.

Table 1
Soil properties across distance-to-nearest-tree bands

Distance band	n	MA, mean ± SD	SOC (%), mean ± SD	pH	Fine fraction (%)
0–2 m	15	4.19 ± 1.45	1.28 ± 0.20	5.46	20.22
> 2–4 m	26	3.33 ± 1.32	1.14 ± 0.22	5.25	19.45
> 4–6 m	28	3.10 ± 1.65	1.14 ± 0.18	5.13	19.45
> 6 m	100	2.50 ± 1.09	1.07 ± 0.16	5.19	18.67

Distance classes are descriptive summaries. Continuous and spatially explicit models provide the primary inference.

After adjustment for SOC, texture and pH, the spline predicted MA of 5.09 at the tree stem, 3.09 at 2 m, 2.60 at 6 m and 2.11 near 12 m. The sharpest decline occurred in the first approximately 2 m, followed by a slower decrease. Confidence intervals were widest at the stem because relatively few observations occurred at very short distances. The adjusted response is shown in Fig. 2c.

3.3 Global and local spatial pattern

All analysed variables exhibited positive global spatial autocorrelation (Online Resource 1, Table S2). Moran's I was 0.167 for ln(MA), 0.267 for SOC and 0.532 for pH (all permutation $p < 0.001$). Distance itself was highly structured ($I = 0.840$), as expected from the tree geometry. The bivariate Moran statistic for SOC and the spatial lag of ln(MA) was 0.066 ($p = 0.065$), indicating only weak evidence that SOC-rich points were surrounded by high-MA neighbours.

Spatial weights were row-standardised and based on eight nearest neighbours; inference used 9,999 permutations.

Local Moran classification identified 16 high–high and 15 low–low MA clusters, together with 3 high–low and 5 low–high spatial outliers. The clusters formed discontinuous patches rather than equivalent halos around each stem. This heterogeneity is consistent with variable tree size or rooting, local management, and unmeasured microsite conditions. The local clusters and spatial outliers are shown in Fig. 4.

3.4 Hierarchical and spatial-error models

The proximity-only model had the lowest AICc (246.50) and explained 13.2% of ln(MA) variation (Table 2). Adding SOC increased R^2 to 15.7% but raised AICc by only 1.35, leaving M3 within the conventional $\Delta AICc \leq 2$ support set. Texture alone, pH, and interactions were not supported. Residual Moran's I remained significant in every ordinary model ($p = 0.007–0.018$ for M1–M5), indicating that robust standard errors did not remove spatial dependence.

Table 2
Hierarchical model comparison for ln(MA)

Model	R^2	Adj. R^2	AICc	ΔAICc	RMSE	Residual I	Residual p
M1 Proximity	0.132	0.121	246.50	0.00	0.493	0.081	0.015
M3 + SOC	0.157	0.131	247.85	1.35	0.485	0.092	0.007
M4 + pH	0.159	0.127	249.69	3.20	0.485	0.088	0.009
M2 + texture	0.136	0.115	249.85	3.35	0.491	0.080	0.018
M5 + interactions	0.160	0.118	253.90	7.41	0.485	0.090	0.009
M0 Null	-0.000	-0.000	266.21	19.71	0.529	0.167	< 0.001

M1, proximity; M2, + ILR texture; M3, + SOC; M4, + pH; M5, + prespecified interactions. Models are ordered by AICc.

In M4, standardised distance had a negative coefficient (β = -0.146, 95% CI -0.228 to -0.064, p < 0.001). The quadratic term was not significant. SOC was positive, but the 95% CI included zero (β = 0.083, p = 0.061). ILR_1, ILR_2, and pH were small and non-significant. All predictor VIFs were ≤ 1.34, excluding problematic collinearity after ILR transformation.

The spatial-error model estimated significant residual dependence (λ = 0.349, 95% CI 0.103 to 0.595, p = 0.005). After accounting for this dependence, distance remained negative (β = -0.146, p = 0.010) and SOC became clearly positive (β = 0.106, p = 0.016); texture and pH remained null. Thus, the proximity association was not an artifact of residual clustering, while the SOC signal was sensitive to how spatial dependence was modelled.

Table 3
Full additive ordinary least-squares model with HC3 standard errors and spatial-error coefficients for ln(MA)

Predictor	OLS β (95% CI)	p	Spatial-error β (95% CI)	p
Tree distance	-0.146 (-0.228, -0.064)	< 0.001	-0.146 (-0.257, -0.036)	0.010
Distance²	0.040 (-0.051, 0.131)	0.389	0.051 (-0.047, 0.150)	0.306
ILR_1: sand vs fines	-0.009 (-0.090, 0.072)	0.830	0.011 (-0.074, 0.096)	0.807
ILR_2: silt vs clay	0.023 (-0.063, 0.109)	0.597	0.024 (-0.060, 0.109)	0.572
SOC	0.083 (-0.004, 0.169)	0.061	0.106 (0.020, 0.192)	0.016
pH	0.023 (-0.053, 0.099)	0.552	0.006 (-0.085, 0.096)	0.904
Spatial error λ	—	—	0.349 (0.103, 0.595)	0.005

All continuous predictors were standardised. The squared term is the square of standardised distance. OLS confidence intervals use HC3 standard errors.

3.5 Spatially blocked prediction, permutation importance and mediation

Out-of-fold prediction was weak for all candidates (Table 4). The proximity-only model performed best (RMSE = 0.516, MAE = 0.386, R^2 = 0.048). Adding SOC, pH, or texture increased RMSE slightly, and the random forest had a negative R^2 (-0.005). The ranking rejected H4: additional covariates that modestly improved in-sample fit did not improve transfer to held-out spatial blocks. Model performance is compared visually in Fig. 5a.

Table 4
Five-fold spatially blocked cross-validation

Model	Blocks	Folds	RMSE	MAE	R^2
Proximity	25	5	0.516	0.386	0.048
Proximity + texture + SOC	25	5	0.523	0.393	0.021
Proximity + texture + SOC + pH	25	5	0.524	0.393	0.019
Proximity + texture	25	5	0.526	0.394	0.011
Random forest	25	5	0.530	0.404	-0.005

The RMSE and MAE are on the ln(MA) scale. Complete 6 × 6 m spatial blocks were held out.

Grouped permutation importance was directionally consistent with the model ranking. Permuting tree proximity increased RMSE by 0.022 on average, the largest positive value, although its 95% permutation interval crossed zero. SOC produced a smaller increase (0.006). The importance of pH was essentially zero, whereas permuting texture reduced RMSE by 0.008, indicating that the texture terms added noise under this validation design rather than unique predictive information. The grouped permutation results are shown in Fig. 5b.

Distance was negatively associated with standardised SOC in the mediator model (a = -0.318), while SOC was positively associated with ln(MA) after covariate adjustment (b = 0.083). Their product was −0.026; the spatial-block bootstrap interval (-0.063 to -0.005) excluded zero. The result is consistent with SOC accounting for part of the proximity gradient, but it does not identify causal mediation and should be interpreted alongside the limited blocked predictive gain from SOC.

3.6 Sensitivity analyses and kriging

The negative distance coefficient was robust to response scale and texture representation. It remained significant for ln(MA) and raw MA and when the two ILR coordinates were replaced by the total fine

fraction. SOC was significant in the ln(MA) fine-fraction model (β = 0.090, p = 0.039) but not in the corresponding raw-MA model, reinforcing that its inferential strength was modest. Texture remained nonsignificant in every sensitivity model.

The spherical variogram produced the lowest leave-one-out kriging RMSE (0.499) and negligible bias (-0.00001). The exponential model was nearly equivalent (RMSE = 0.499), whereas the Gaussian model performed poorly (RMSE = 1.477). The kriged surface depicted discontinuous zones of elevated MA near portions of the tree alignment rather than five uniform circles. Prediction variance was broadly similar across much of the gridded domain because observations were regularly spaced, but uncertainty increased between and beyond sampling points. Full variogram cross-validation results are provided in Online Resource 1, Table S3, and the prediction and variance surfaces are shown in Fig. 6.

4 Discussion

4.1 Tree proximity was the most robust correlate of microbial activity

The analysis strongly supported H1. MA was approximately 68% higher within 0–2 m than beyond 6 m, the adjusted spline showed the steepest decline in the near-stem zone, and distance retained a negative coefficient in robust OLS, raw-response sensitivity models, and the spatial-error model. This convergence is stronger than any single under-canopy contrast because it links the observed gradient to continuous distance while controlling for carbon, pH, compositionally coherent texture and residual spatial dependence.

The pattern is ecologically plausible. Mature *F. albida* trees redistribute organic inputs through litter, roots and rhizodeposition; alter near-surface moisture and temperature; trap fine organic and mineral material; and create habitat for livestock and soil fauna (Haile et al. 2014; Tsegu 2019). Zambian studies have documented higher organic carbon and nutrients near *F. albida* and canopy- or litter-associated changes in microbial communities and nitrogen mineralization (Umar et al. 2013). The present study extends that evidence by showing that tree proximity remains informative after spatial adjustment, although the modest R^2 indicates that distance is a coarse proxy for a bundle of unmeasured processes rather than a complete mechanistic explanation.

4.2 SOC was a plausible partial pathway, not a stand-alone predictor

H2 received qualified support. SOC correlated positively with MA, entered the ΔAICc ≤ 2 model, was significant in M3 and the spatial-error model, and produced a block-bootstrap indirect association whose confidence interval excluded zero. These results are consistent with greater carbon availability contributing to near-tree respiration. They also align with studies in which *F. albida* canopy and litter inputs altered soil carbon, microbial biomass and nitrogen cycling (Yengwe et al. 2018a).

However, SOC added little spatially blocked predictive skill and the bivariate SOC–MA Moran statistic was only marginal. Total SOC integrates material with widely different turnover rates, whereas short-term respiration responds to labile substrates, rhizosphere inputs, moisture, temperature, oxygen and microbial community composition (Henneron et al. 2022; Kuzyakov and Gavrichkova 2010). The divergence between inferential and predictive evidence is therefore informative rather than contradictory: SOC may participate in the pathway while remaining an incomplete proxy for the spatially variable resources that control instantaneous respiration. Future work should measure dissolved or particulate organic matter, microbial biomass, moisture, root density and canopy dimensions alongside total SOC.

4.3 Compositional texture did not explain the gradient

H3 was not supported. Neither ILR coordinate was associated with ln(MA); the fine-fraction sensitivity model produced the same conclusion, and texture permutation decreased held-out RMSE. The strong correlations among raw sand, silt and clay were largely mathematical consequences of closure and should not be interpreted as independent soil processes. ILR transformation removed this structural collinearity and showed that meter-scale texture contrasts did not account for the biological gradient after proximity and SOC were considered.

This result does not imply that texture is ecologically irrelevant. Texture regulates water retention, aeration, aggregation and organo-mineral protection (Hoffland et al. 2020), but its narrow range in this sand-dominated plot may have provided insufficient independent contrast. Some apparent near-tree fine-particle enrichment could also be spatially aligned with tree distance, making it difficult to separate contemporary trapping from inherited field structure in a single site. Replicated tree–open controls across contrasting textures are needed to test whether texture moderates the *F. albida* effect.

4.4 Spatial dependence changed inference and constrained prediction

Residual Moran statistics and the significant spatial-error parameter demonstrate that neighbouring observations shared information not captured by measured predictors. Ignoring that dependence would produce confidence intervals that are too narrow and would overstate geographic generalization (Gelfand et al. 2006). Accounting for spatial error did not erase the distance effect; instead, it strengthened evidence for SOC. This pattern suggests a broad tree-related gradient superimposed on smaller-scale spatial processes such as management patches, moisture redistribution or unmeasured soil properties.

Prediction remained weak. The best blocked R^2 was only 0.048, and the random forest failed to outperform a proximity spline. This is a central result for high-resolution soil mapping: a model can identify a statistically defensible average gradient while still having limited ability to predict individual held-out locations. Random folds would allow neighbouring points to enter both training and test sets and could therefore give a misleadingly optimistic assessment (Roberts et al. 2017). Spatial blocking revealed the practical limitation of the current covariate set.

The kriged surface was almost unbiased under leave-one-out interpolation, but near-zero bias is not equivalent to high explanatory skill. Ordinary kriging borrows information directly from nearby MA observations, whereas blocked regression asks whether environmental predictors transfer to separated locations. The map is useful for visualizing the observed spatial field and targeting future sampling, but it should not be presented as a precise management prescription without external validation and explicit uncertainty.

4.5 Agronomic and sampling implications

The strong mean difference between locations within 2 m and those beyond 6 m shows that a single composite soil sample can misrepresent an agroforestry field. Sampling plans should stratify by tree distance, canopy geometry and position along a tree row. Where soil biological indicators are used to evaluate soil health, open-field and near-tree samples should be reported separately or combined using area-weighted designs.

The findings also caution against assuming that every near-tree zone is equally fertile. Local Moran clusters and the kriged surface were discontinuous, and MA differed among nearest-tree groups. Site-specific management should therefore integrate tree size, root and canopy distribution, crop performance, moisture and nutrient availability. The current results support retaining mature *F. albida* as a potential soil-health resource, but they do not establish a universal radius for fertilizer reduction or crop placement.

4.6 Limitations and priorities

The primary limitation is treatment-level pseudoreplication: 169 coordinates provide detailed spatial resolution but not 169 independent tree replicates. The field lacked a matched tree-free control and may contain pre-existing gradients in parent material, amendment history, tillage or drainage. Exact tree dimensions and canopy radii were unavailable, and laboratory protocols and the mass basis of MA must still be made explicit before the work can be fully reproducible.

The mediation analysis relies on strong no-unmeasured-confounding assumptions and is explicitly observational. The 6-m blocking scheme was chosen a priori and should be supplemented by alternative block sizes and leave-one-tree-out validation where tree replication permits. The local Moran map used unadjusted pointwise $p < 0.05$ and should be treated as exploratory. Finally, one-time respiration measurements may be sensitive to seasonal moisture and incubation conditions; repeated sampling would determine whether near-tree hotspots are temporally persistent.

5 Conclusions

Dense spatial sampling revealed a robust decline in microbial activity with increasing distance from mature *Faidherbia albida* trees in a sandy Zambian field. The proximity signal persisted after compositional texture adjustment and spatial-error modelling. SOC provided evidence of a partial pathway but did not materially improve prediction in held-out spatial blocks. Texture composition and pH

contributed neither independent coefficients nor predictive gains. Thus, H1 was supported, H2 was partially supported, and H3 and H4 were not supported.

The study's principal contribution is methodological as well as ecological; inferential significance, spatial dependence and geographic predictive performance led to different but complementary conclusions. Tree proximity explained an average fertility-island gradient, while low blocked R^2 and discontinuous local clusters showed that individual locations remained difficult to predict. Replicated fields with complete laboratory metadata; canopy and root measurements; and soil moisture and labile-carbon indicators are required before the observed associations can be generalized or interpreted causally.

Declarations

Acknowledgements

The authors acknowledge the technicians of the Department of Soil Science, University of Zambia, and the management of Kasisi Agricultural Training Centre for field and laboratory support.

Funding

This work received complementary support through the Agricultural Productivity Programme for Southern Africa, implemented by the Zambia Agricultural Research Institute with World Bank support.

Competing interests

The authors declare no competing financial or non-financial interests.

Author contributions

Elijah Phiri: conceptualisation, methodology, supervision, formal analysis, visualisation, writing—original draft, and writing—review and editing. Patience Chanda: investigation, laboratory analysis, data curation, formal analysis, and writing—review and editing. Alice Mutiti Mweetwa: methodology, supervision, soil microbiology, interpretation, and writing—review and editing.

Data and code availability

The coordinate-level dataset, reproducible Google Colab workflow, out-of-fold predictions, spatial diagnostics and processed tables are available from the corresponding author on reasonable request.

References

1. Aitchison J (1986) The statistical analysis of compositional data. Chapman & Hall, London. https://doi.org/10.1007/978-94-009-4109-0
2. Akinnifesi FK, Ajayi OC, Sileshi GW, Chirwa PW, Chianu J (2010) Fertiliser trees for sustainable food security in the maize-based production systems of East and Southern Africa: a review. Agronomy for Sustainable Development 30:615–629. https://doi.org/10.1051/agro/2009058
3. Anderson JPE (1982) Soil respiration. In: Page AL (ed) Methods of soil analysis, part 2: chemical and microbiological properties, 2nd edn. American Society of Agronomy and Soil Science Society of America, Madison. https://doi.org/10.2134/agronmonogr9.2.2ed.c41
4. Anselin L (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht. https://doi.org/10.1007/978-94-015-7799-1
5. Anselin L (1995) Local indicators of spatial association—LISA. Geographical Analysis 27:93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x
6. Barnes RD, Fagg CW (2003) Faidherbia albida: monograph and annotated bibliography. Tropical Forestry Papers No. 41. Oxford Forestry Institute, Oxford
7. Boffa JM (1999) Agroforestry parklands in sub-Saharan Africa. FAO Conservation Guide No. 34. Food and Agriculture Organization of the United Nations, Rome
8. Bouyoucos GJ (1962) Hydrometer method improved for making particle-size analyses of soils. Agronomy Journal 54:464–465. https://doi.org/10.2134/agronj1962.00021962005400050028x
9. Breiman L (2001) Random forests. Machine Learning 45:5–32. https://doi.org/10.1023/A:1010933404324
10. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C (2003) Isometric logratio transformations for compositional data analysis. Mathematical Geology 35:279–300. https://doi.org/10.1023/A:1023818214614
11. Garrity DP, Akinnifesi FK, Ajayi OC, Weldesemayat SG, Mowo JG, Kalinganire A, Larwanou M, Bayala J (2010) Evergreen agriculture: a robust approach to sustainable food security in Africa. Food Security 2:197–214. https://doi.org/10.1007/s12571-010-0070-7
12. Gelfand AE, Holder MT, Latimer AM, Lewis PO, Rebelo AG, Silander JA, Wu S (2006) Explaining species distribution patterns through hierarchical modeling. Bayesian Analysis 1:41–92. https://doi.org/10.1214/06-BA102
13. Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89:1–45. https://doi.org/10.1016/S0016-7061(98)00078-0
14. Haile G, Itanna F, Lemenhi M (2014) Status of soil properties of scattered Faidherbia albida in agricultural landscapes in the central highlands of Ethiopia. African Journal of Agricultural Research 9:3379–3387. https://doi.org/10.5897/AJAR2014.9127
15. Heanes DL (1984) Determination of total organic C in soils by an improved chromic acid digestion and spectrophotometric procedure. Communications in Soil Science and Plant Analysis 15:1191–1213. https://doi.org/10.1080/00103628409367551

16. Henneron L, Balesdent J, Alvarez G et al (2022) Bioenergetic control of soil carbon dynamics across depth. Nature Communications 13:7676. https://doi.org/10.1038/s41467-022-34951-w
17. Hoffland E, Kuyper TW, Comans RNJ, Creamer R (2020) Eco-functionality of organic matter in soils. Plant and Soil 455:1–22. https://doi.org/10.1007/s11104-020-04651-9
18. Imai K, Keele L, Yamamoto T (2010) Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science 25:51–71. https://doi.org/10.1214/10-STS321
19. Kravchenko AN (2003) Influence of spatial structure on accuracy of interpolation methods. Soil Science Society of America Journal 67:1564–1571. https://doi.org/10.2136/sssaj2003.1564
20. Kuzyakov Y, Gavrichkova O (2010) Time lag between photosynthesis and carbon dioxide efflux from soil: a review of mechanisms and controls. Global Change Biology 16:3386–3406. https://doi.org/10.1111/j.1365-2486.2010.02179.x
21. Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–23. https://doi.org/10.1093/biomet/37.1-2.17
22. Roberts DR, Bahn V, Ciuti S et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical or phylogenetic structure. Ecography 40:913–929. https://doi.org/10.1111/ecog.02881
23. Sileshi GW (2016) The magnitude and spatial extent of influence of Faidherbia albida trees on soil properties and primary productivity in drylands. Journal of Arid Environments 132:1–14. https://doi.org/10.1016/j.jaridenv.2016.03.002
24. Thomas GW (1996) Soil pH and soil acidity. In: Sparks DL et al (eds) Methods of soil analysis, part 3: chemical methods. Soil Science Society of America, Madison. https://doi.org/10.2136/sssabookser5.3.c16
25. Tsegu ED (2019) The role of Faidherbia albida tree species in parkland agroforestry and its management in Ethiopia. Journal of Horticulture and Forestry 11:42–47. https://doi.org/10.5897/JHF2018.0570
26. Umar BB, Aune JB, Lungu OI (2013) Effects of Faidherbia albida on the fertility of soil in smallholder conservation agriculture systems in eastern and southern Zambia. African Journal of Agricultural Research 8:173–183. https://doi.org/10.5897/AJAR11.2464
27. Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2019) blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution 10:225–232. https://doi.org/10.1111/2041-210X.13107
28. Walkley A, Black IA (1934) An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. Soil Science 37:29–38. https://doi.org/10.1097/00010694-193401000-00003
29. Yengwe J, Gebremikael MT, Buchan D, Lungu OI, De Neve S (2018a) Effects of Faidherbia albida canopy and leaf litter on soil microbial communities and nitrogen mineralization in selected Zambian soils. Agroforestry Systems 92:349–363. https://doi.org/10.1007/s10457-016-0063-4
30. Yengwe J, Lungu OI, Buchan D, De Neve S (2018b) Quantifying nutrient deposition and yield levels of maize (Zea mays) under Faidherbia albida agroforestry systems in Zambia. European Journal of

Agronomy 99:148–155. https://doi.org/10.1016/j.eja.2018.07.004

Figures

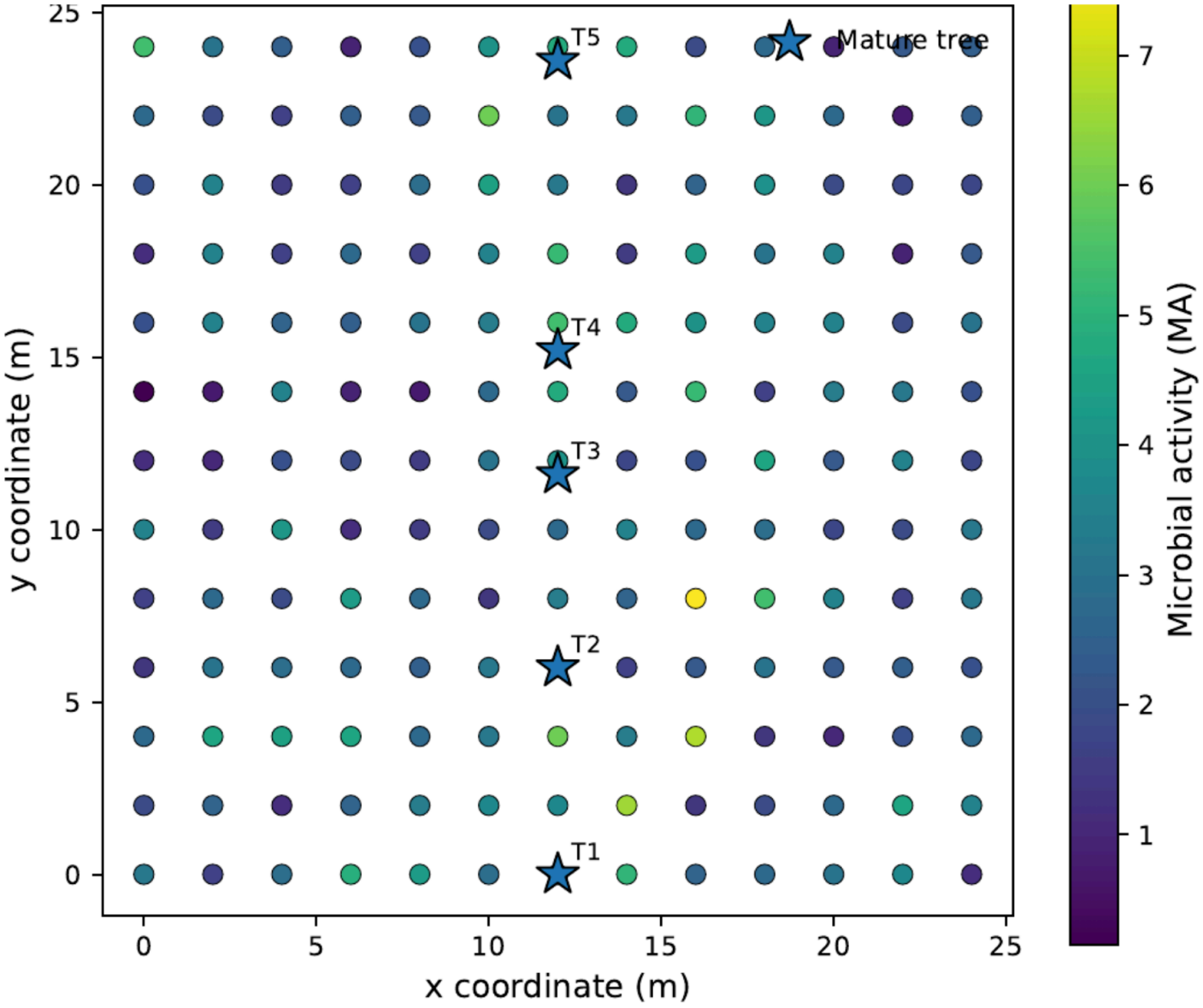


Figure 1

Sampling grid, mature tree positions and observed microbial activity (MA). The regular 2-m grid comprised 169 observations. Stars denote the five tree stems, and point colour represents measured MA

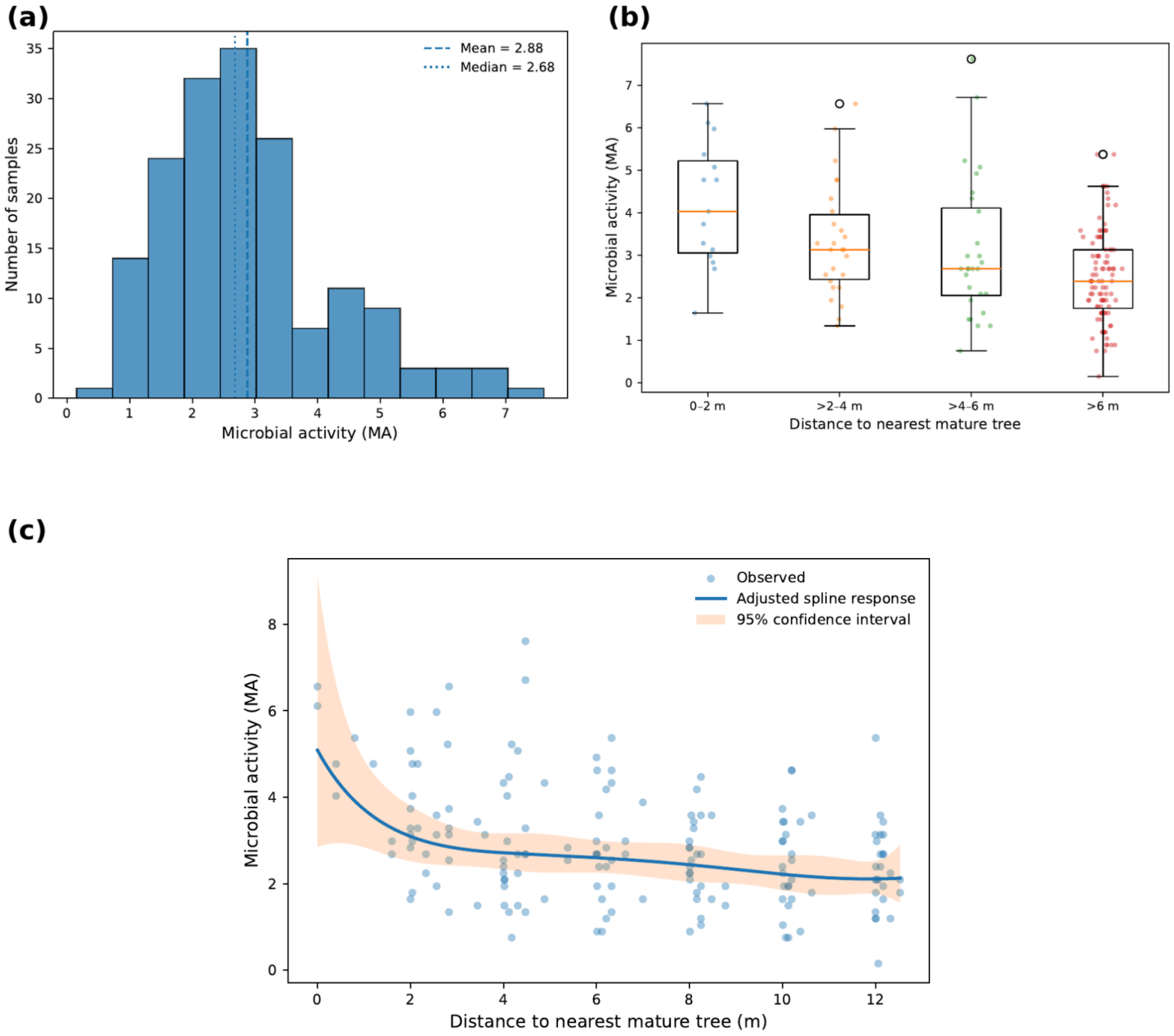


Figure 2

Distribution and tree-proximity response of microbial activity: (a) frequency distribution, with vertical lines marking the mean and median; (b) observations across distance-to-nearest-tree bands, with boxes showing interquartile ranges and horizontal lines showing medians; and (c) adjusted spline response, with the 95% confidence interval. Panel (c) holds SOC, pH and both texture coordinates at their means and presents predictions on the original MA scale

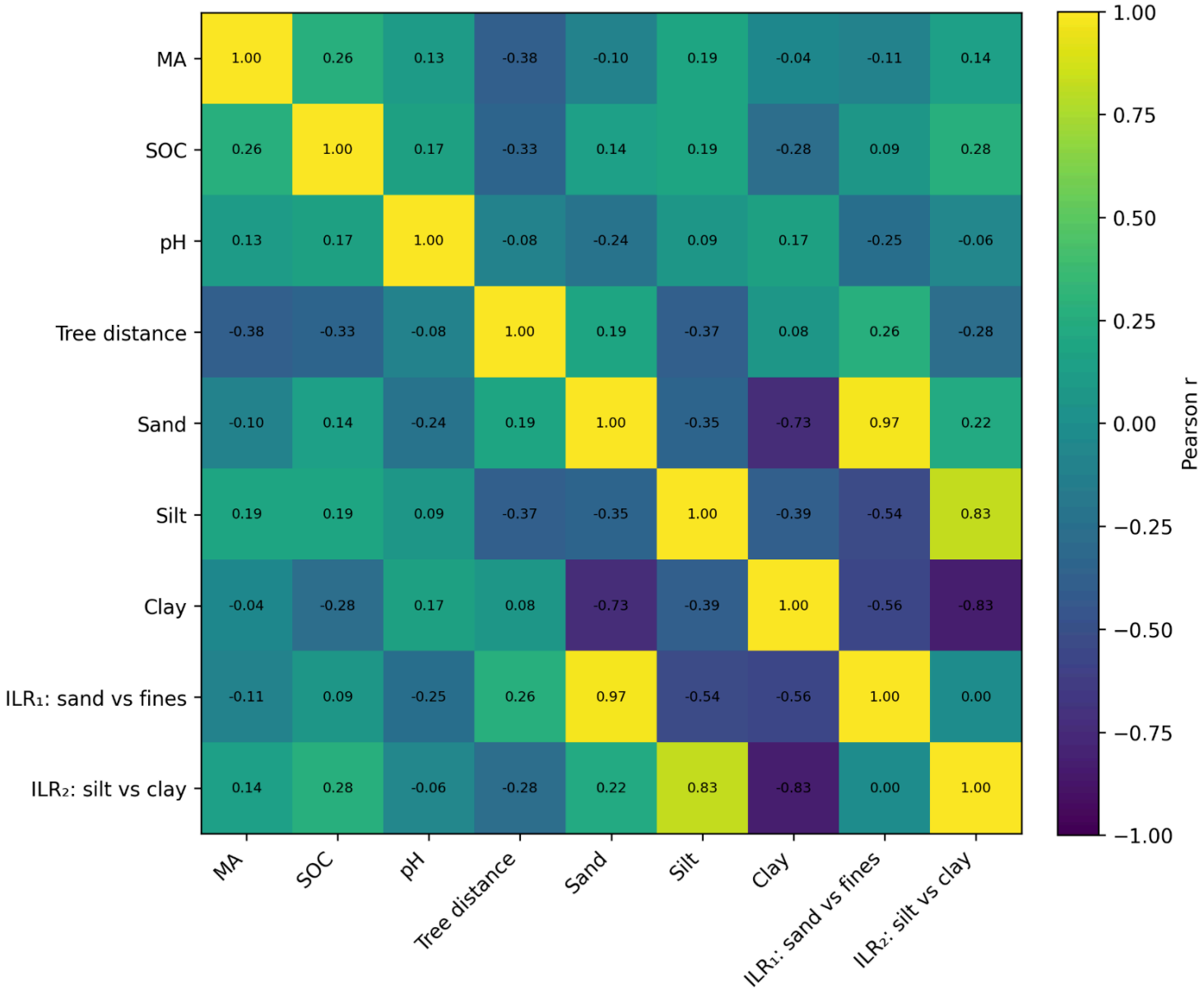


Figure 3

Pearson correlation matrix among MA, SOC, pH, tree distance, texture fractions and the two isometric log-ratio texture coordinates. Strong correlations between raw texture fractions and the log-ratio coordinates reflect the closed nature of particle-size data

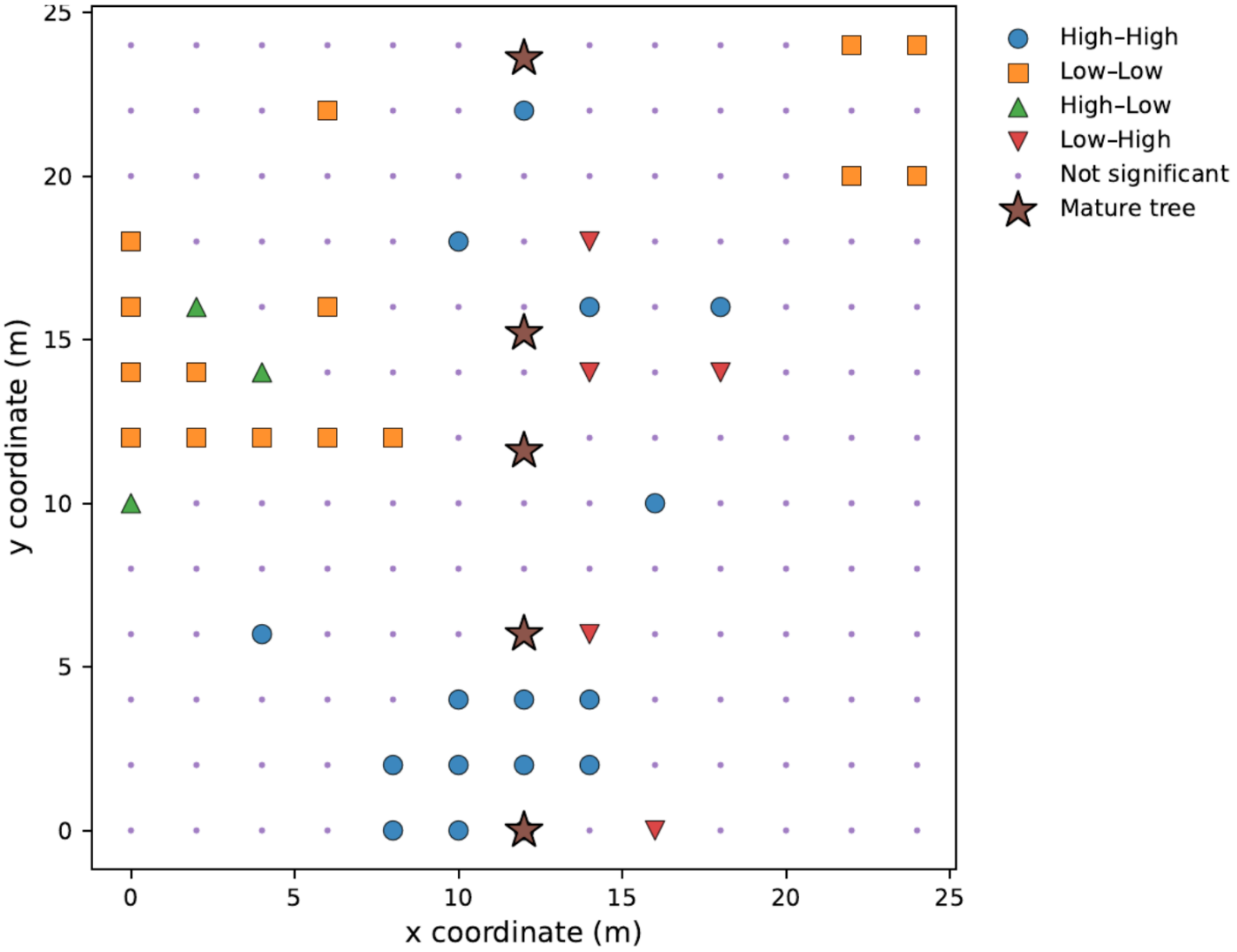


Figure 4

Local Moran clusters of ln(MA) at $p < 0.05$. High–high and low–low points identify local clusters; high–low and low–high points are spatial outliers. Stars denote mature tree positions

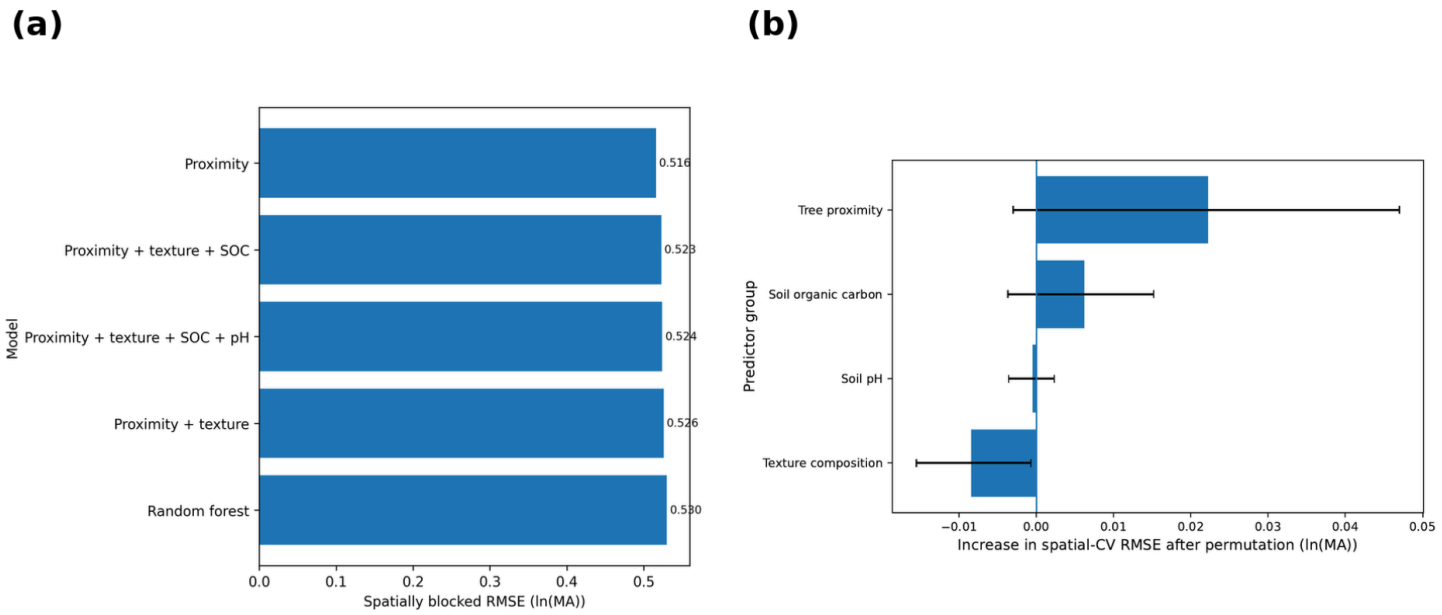


Figure 5

Spatial predictive evaluation: (a) spatially blocked root mean square error for the five candidate models, where lower values indicate better transfer to held-out blocks; and (b) grouped permutation importance under spatial cross-validation, where bars are mean increases in held-out root mean square error and whiskers are empirical 95% intervals across 100 permutations

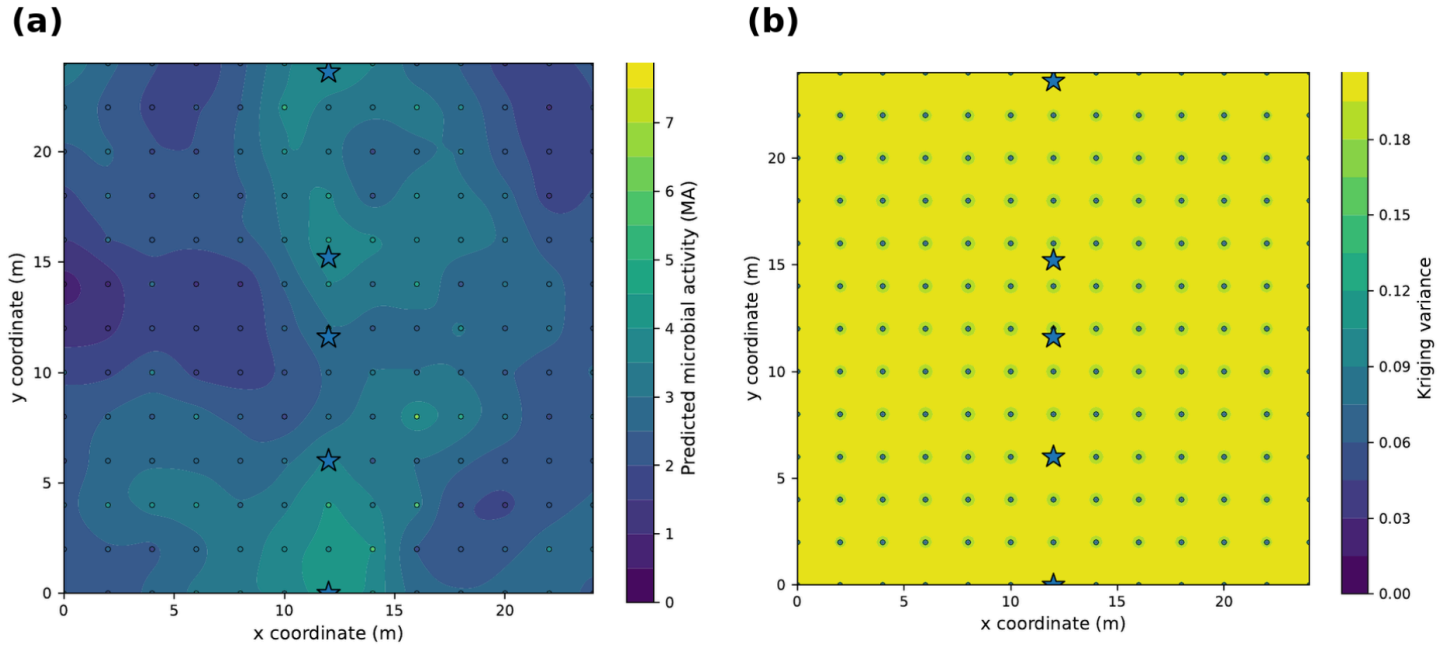


Figure 6

Ordinary kriging outputs: (a) predicted microbial activity using the selected spherical variogram, fitted on ln(MA) and back-transformed to the original scale; and (b) kriging prediction variance for ln(MA). Circles denote observations and stars denote mature trees

Supplementary Files

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

- Phirietal2026AgroforestrySystemsOnlineResource1.docx