Smart Agricultural Technology 12 (2025) 101357

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

How to select reference sites for long-term agricultural experiments: A data-driven approach

Nishita Thakur [a,b], Marco Donat [a], Sonoko Bellingrath-Kimura [a,c], Wiebke Niether [b], Andreas Gattinger [b], Deise Aline Knob [b], Eva-Maria L. Minarsch [b], Philipp Weckenbrock [b], Franz Schulz [b], Lutz Breuer [d,e], Suzanne Jacobs [d,e], John Clifton-Brown [f], Karolina Golicz [g,*]

[a] Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany
[b] Department of Agronomy and Plant Breeding II, Institute for Organic Farming (Focus on Sustainable Soil Use), Justus Liebig University Giessen, Germany
[c] Institute of Agriculture and Horticulture, Faculty of Life Sciences, Humboldt University of Berlin, Germany
[d] Institute for Landscape Ecology and Resources Management (ILR), Research Centre for Biosystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, Germany
[e] Centre for International Development and Environmental Research (ZEU), Justus Liebig University Giessen, Germany
[f] Department of Agronomy and Plant Breeding I, Research Centre for Biosystems, Land Use and Nutrition (iFZ), Justus Liebig University Giessen, Germany
[g] Research Division, EU Research Office, Justus Liebig University Giessen, Germany

ARTICLE INFO

Keywords:
Experimental design
Land use change
Spatial data analysis
K-means clustering

ABSTRACT

Measuring the effects of land use changes such as agroforestry, photovoltaics or afforestation, on environmental variables poses an ongoing challenge for scientists. These difficulties stem from the long time frame for changes to manifest and the difficulty in isolating pre-existing site conditions from land use change effects.. This study addresses the key issue of reference site selection, which is often oversimplified to relying on spatially proximate but potentially heterogeneous neighbouring fields.. Advances in high-resolution spatial and temporal environmental data provide opportunities to enhance the robustness of land use change studies by enabling the identification of statistically comparable reference sites across the landscape, rather than relying on fixed, proximal references. The aim of this study was to develop a generalizable and easy-to-implement approach for the identification of potential reference sites, based on matching site conditions. Using 12 variables across three key categories—agronomic, topographic, and edaphic (soil)—we evaluated six fields (totalling 42.7 ha) to identify locations most comparable to a set of ad-hoc experimental plots located in a case-study agroforestry field (3.5 ha). High-resolution maps (3 × 3 m) were generated using state-of-the-art satellite imagery and data from a proximal multi-sensor platform. Geographically weighted principal component analysis (GWPCA) combined with K-means clustering stratified the fields and identified areas most similar to the experimental plots, demonstrating the methodology. Within individual fields, we identified specific areas that closely matched the target conditions of the investigated plots, optimizing the selection of reference sites for future sampling. Results showed that the fields nearest to the case-study field exhibited the highest similarity, as expected based on the Tobler's proximity law and that the soil data generated from a proximal multi-sensor platform had limited impact on the selection. This approach, supported by open-access, well-documented Python code, is designed to be flexible and easily adapted to various research needs.

1. Introduction

A growing interest in agricultural land use systems involving perennial crops such as trees [1] and rhizomatous grasses [2] as well as modern technologies, e.g., agri-photovoltaics [3], calls for careful consideration of their impacts on a range of agronomic and environmental parameters. Robust impact assessments require reference systems that are characterised by similar biotic and abiotic conditions, reducing the potential for misattribution of the effects of pre-existing site conditions to the effects of the change of the land use system.

* Corresponding author.
E-mail address: karolina.golicz@admin.uni-giessen.de (K. Golicz).

https://doi.org/10.1016/j.atech.2025.101357
Received 5 December 2024; Received in revised form 21 August 2025; Accepted 22 August 2025
Available online 23 August 2025
2772-3755/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Sampling campaigns which collect baseline (time zero) data prior to the land use change, are common to studies of long-term agricultural experiments, for example, from treeless cropland (or grassland) into agroforestry i.e., a land use system where trees and crops are cultivated simultaneously on the same unit of land [40]. The baseline data enables robust quantification of temporal changes in monitored variables. but can be resource-intensive and limited by (1) a lack of data for systems which were not previously under scientific investigation and (2) a limited applicability to crop yield assessments, which are dependent on weather and crop variety. Instead, a space-for-time substitution principle can be employed by investigating multiple fields (both neighbouring and non-neighbouring) in different stages of transition into a new land use system, i.e., a land use chronosequence [4], to better understand the trajectory of change in monitored parameters [5]. However, a land use chronosequence by itself cannot account for changes in the environment during the observation period – therefore the fields to be investigated as potential reference sites must be carefully selected.

On a farm-scale, paired-plot experimental designs are frequently employed. Golicz et al. [6] summarised recent literature involving temperate European agroforestry systems and showed that 70 % of the studies used an adjacent treeless field to act as a reference site. However, a neighbouring field may differ in land use history, crop rotation, and management practices such as tillage frequency, fertilization, and pest control, yield performance, topography, soil, or microclimate conditions, which poses a challenge for comparisons between paired-plots at the landscape scale (discussed in, e.g., [7,8]). Furthermore, the introduction of linear features such as tree rows into cropland induces edge effects [9] between perennial rows and crop alleys in formerly unaffected (no-edge) parts of the field [10–12].

In the absence of neighbouring sites with similar site conditions, it is possible to establish an in-field reference (Fig. 1), assuming that these in-field reference plots are not affected by the new land use. Existing literature (e.g., [10]) and ground investigations can be used to delineate zones free from the effects of land use change, such as tree shading or competition for water and nutrients, by establishing the maximum distance beyond which the effects are minor and/or absent. For example, in agroforestry systems, microclimatic effects were shown to extend over distances of up to 30 times the tree height [13,14]. This approach, however, carries a risk of underestimating the changes to the measured parameters because the effects might extend beyond the assumed distance.

Studies of long-term agricultural experiments such as agroforestry illustrate the complexity of reference system selection (Fig. 1). Overall, to ensure robust assessment of the effects of land use change on a range of environmental parameters, site stratification [15] should be carried out to ensure that the in-field (plots within the experimental field) and the out-field plots (plots within conventional cropland) are constrained to the same stratum. On the landscape scale, this might involve identifying suitable sections of individual fields, which might not directly neighbour the investigated experimental field. The in-field and out-field reference plots can be established simultaneously, with the out-field reference plots mirroring the biotic (e.g., vegetation cover) and abiotic conditions (e.g., soil texture, nutrient levels) of the in-field plots. Once the out-field reference plots are identified, additional reference plots can be set up for individual in-field plot designated as intervention (treatment). These reference plots should be arranged spatially to allow for meaningful comparisons between the in-field experimental plots and their corresponding out-field controls, while also accounting for small-scale field heterogeneity. For example, in many agroforestry studies, the experimental plots are arranged in a transect i.e., plots are sampled every few meters in a straight line moving from the tree trunk towards the centre of the field with the centre acting as the in-field reference [6]. This spatial arrangement can be replicated in an out-field setting.

In this study, we aimed to develop a data-driven method for a farm-scale reference site selection. This was achieved by means of collating

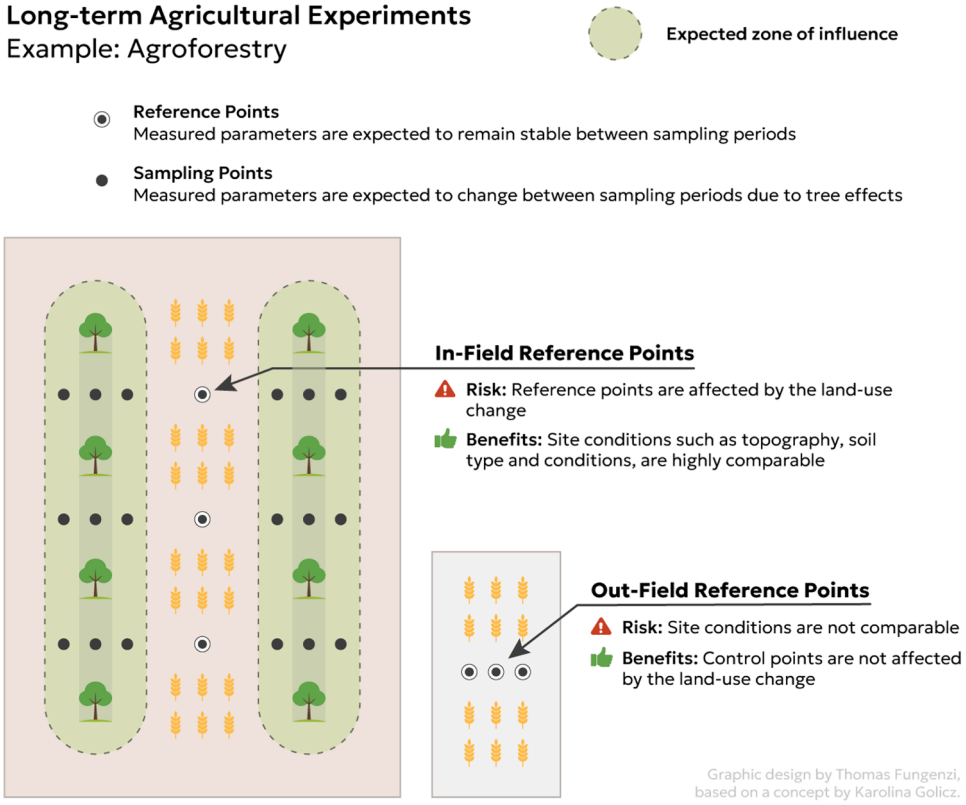


Fig. 1. Two types of reference plots, in-field and out-field, which can be used to investigate the impacts of long-term agricultural experiments involving land use change such as agroforestry on crop and environmental (biotic and abiotic) parameters.

publicly accessible geodata on agronomic and topographic parameters ($n = 8$) alongside proximal sensor-derived soil parameters ($n = 4$) and applying a K-means clustering algorithm to stratify the sites to demonstrate the necessary operational steps. We identified areas across multiple crop fields with site conditions similar to six in-field reference plots located in our experimental (agroforestry) field. We expected that, as per Tobler's Law [16], there is greater similarity among fields closest to each other because of their geographic proximity. However, we hypothesised that proximity alone is insufficient to guide the selection of reference sites. Furthermore, we hypothesised that the inclusion of proximal soil sensor data would significantly affect the stratification and thus, have an impact on reference site selection.

2. Site description

The case-study field was located at the Gladbacherhof experimental farm in Villmar, Aumenau (coordinates: 50° 24′ N, 8° 15′ E), managed by Justus Liebig University Giessen (Germany). The arable field (3.5 ha in size) was converted into an agroforestry system (F1) in February 2020. The F1 (Fig. 2A-D) is an alley-cropping field with six tree rows (3 m-wide) alternating with crop alleys (18 m-wide). Within F1, a total of six target points ($n = 6$), i.e., plots located between two tree rows and assumed to be outside of the zone of tree influence, were selected as target points. The target points acted as in-field reference plots with out-field reference plots to be placed in the same stratum, i.e., areas with matching biotic and abiotic conditions. We selected five fields from the same farm (F2-F6; total size: 42.7 ha) and characterised by organic farm management close to F1 to identify out-field reference plots. The target in-field reference points and the subsequent out-field reference points were generated ad-hoc to demonstrate the method and its operational procedure.

3. Methods

3.1. Variable selection and data pre-processing

We considered three major categories of variables, i.e., agronomy, topography, and soil conditions, which were relevant to the farm under study, with the criteria for the final selection involving the availability of geodata and the proximal sensor-derived soil data (Fig. 2A-D). In practical applications, variable selection should be guided by the specific research objectives and data availability..

To demonstrate the method (Fig. 3), we considered a subset of these variables - a full dataset (with 12 variables) and a reduced dataset (with 8 variables that excluded soil properties obtained by proximal sensing) (Table 1). This is not to be confused with dimensionality reduction, which refers to the method of excluding specific variables from the dataset to avoid multi-collinearity. The data was processed separately for each investigated field. The soil properties (electrical resistivity, soil moisture index and topsoil gamma activity) were interpolated across the target fields by ordinary kriging. Subsequently, we extracted the values for the six target points from each raster.

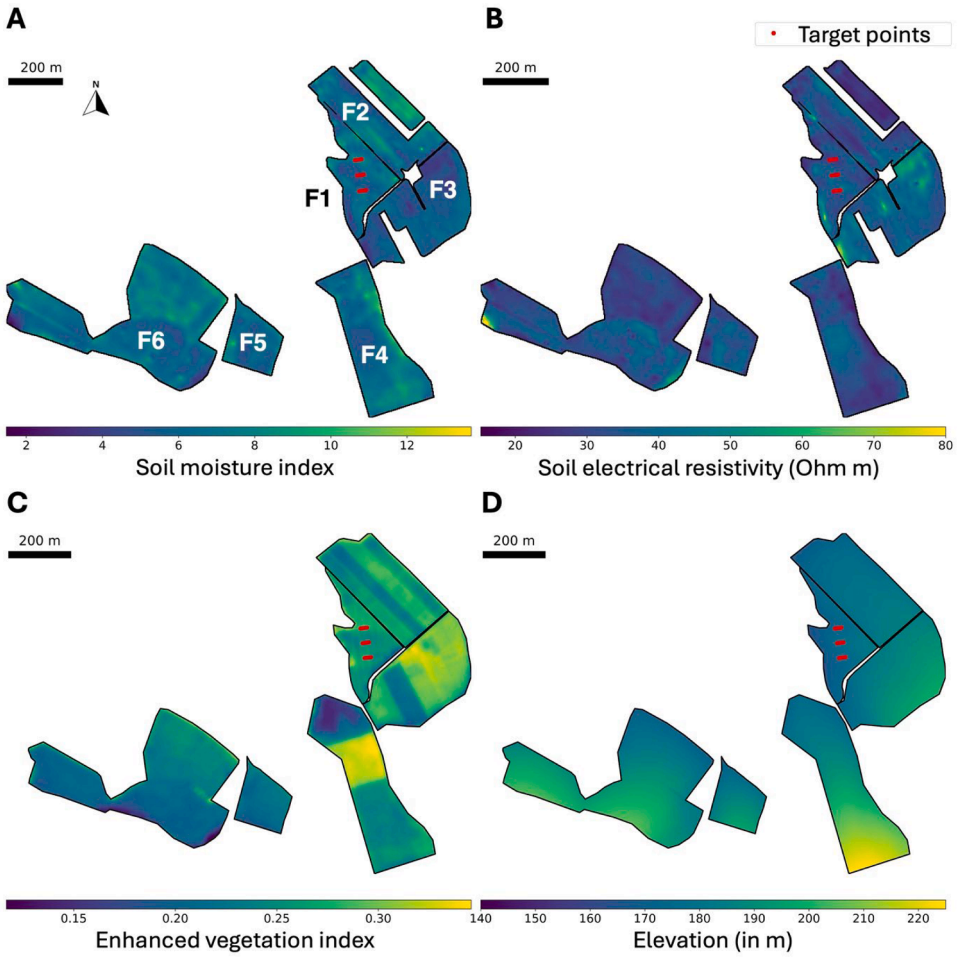


Fig. 2. A-D. The investigated agroforestry field (F1) with target points highlighted in red alongside candidate fields for investigation (F2–6) for the identification of reference sites. There is substantial spatial variability in soil conditions (A-B), agronomic indicators (C) and topography (D). Letters and numbers refer to field IDs.

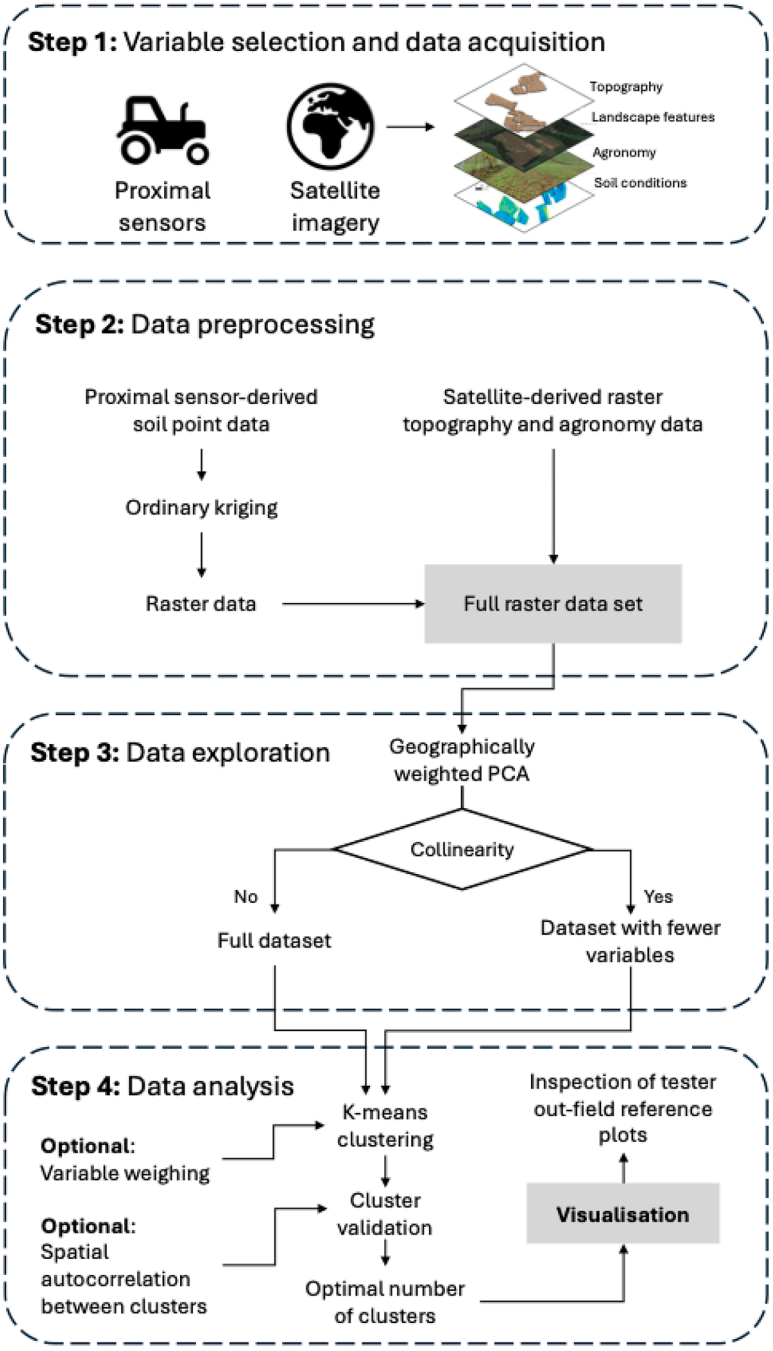


Fig. 3. A data-driven methodology for identification of reference sites that integrates different environmental data, including agronomic, topographical and soil parameters, to ensure comparability with experimental sites.

3.1.1. Agronomy

We acquired satellite data (3 m resolution) from Planet Scope (includes material © Planet (2020–2023), provided by BKG and BMI, all rights reserved) for the vegetation growing period of 4 years (2020–2023) for which data at the high resolution was available. We used cloud-free images to compute the mean and median normalised difference vegetation index (NDVI) and enhanced vegetation index (EVI) for the 4 years. The variables we generated from this were used to indicate crop biomass and productivity [17].

3.1.2. Topography

We acquired the digital elevation model (DEM) as a tile grid with a resolution of 1 m, and combined with the 'Mosaic to New Raster' tool in ArcGIS Pro (ver. 3.0) to cover the entire region of interest. Subsequently, we used the raster file generated to derive elevation and calculate topographic indices of slope, aspect, topographic wetness index and terrain curvature (Gao, 2021).

3.1.3. Soil conditions

We used a proximal soil sensing platform called the Geophilus Soil Mapping System [18] in 2022 to obtain electrical resistivity values. Geophilus maps complex electrical volume resistivity, which is indicative of soil structure, up to a depth of 1.5 m [42]. To limit errors associated with the signal penetration through the soil layers, we included only the topsoil (0–30 cm; Rh1) and the subsoil (30–60 cm; Rh2 and 60–90 cm; Rh3) in the analysis. Other proximal soil sensor-derived

Table 1
Parameters selected for site selection. Projection for the covariates: CRS WGS 1984 UTM 32 N. The common geometry for the raster data format was the resolution of 3 m x 3 m.

Category	Variables	Unit	Resolution	Source
Agronomy	Normalised difference vegetation index (NDVI)*	(-)	3 × 3 m	Planet Scope, © Planet (2020–2023), provided to ZALF by BKG and BMI, all rights reserved
	Enhanced vegetation index (EVI)*	(-)		
Topography	Elevation (ELE)	Meters	1 × 1 m	Hessische Verwaltung für Bodenmanagement und Geoinformation (HVBG)
	Aspect	(-)		
	Slope (SL)	(°)		
	Terrain curvature (CUR	(-)		
	Topographic wetness index (TWI)	(-)		
Soil characteristics	Proximal sensor-derived electrical resistivity at a depth of 0–30 cm (Rh1)	(Ohm m^{-1})	Point data	Geophilus
	Proximal sensor-derived electrical resistivity at a depth of 30–60 cm (Rh2)	(Ohm m^{-1})		
	Proximal sensor-derived electrical resistivity at a depth of 60–90 cm (Rh3)	(Ohm m^{-1})		
	Topsoil Gamma activity (GM)	(counts s^{-1})		
	Soil moisture index (SMI)	(-)		

Explanations and abbreviations: *Mean and median; averaged growing periods of 4 years.

parameters included the gamma activity of the topsoil and the soil moisture index, i.e., the measure of relative moisture distribution in the area calculated as the quotient of gamma and Rh1. Soil variables were expressed as point data (measurements dispersed in a regular pattern along a straight line every 1.8 m with measurement lines equivalent to the width of the machinery, [18]) which we interpolated across the fields using ordinary kriging (following [19]). We then determined the spatial covariance structure of the sampled points, after which we fitted circular variograms to the gamma and soil moisture index values and an exponential variogram to the electrical resistivity values. Due to the missing data, we extrapolated soil values across larger distances for sections of fields F3 and F4.

Once the rasters were derived, we then extracted values for individual variables of the selected target points (for the values of individual target points, see Table S2 in the Supplementary Material).

3.2. Data exploration and statistical analyses

3.2.1. Geographically weighted PCA

We used the Pearson's correlation coefficients and a geographically weighted PCA (GWPCA) to investigate the collinearity of the investigated dataset (for the correlation matrix, see Figure S1 in the Supplementary Material). The GWPCA was used to account for the spatial structure of the dataset, including spatial autocorrelation [43]. The datasets were processed with the GWPCA twice – once with and once without the electrical resistivity-derived soil variables. We applied the latter to test the applicability of this method for sites with limited area-specific soil information, as the selection of experimental sites is usually carried out before the collection of baseline soil data. The second computed dataset involved fewer variables ($n = 8$). The output of the GWPCA was used to examine spatial autocorrelation, which we verified by the Moran's Index. The index measures spatial autocorrelation, indicating how similar or dissimilar values are across nearby geographic locations, with values ranging from +1 (strong clustering) to −1 (strong dispersion), and 0 (randomness) (see Table S1 in the Supplementary Material), as well as to inspect whether a reduction of dimensions (variables) is necessary prior to clustering. Dimensionality reduction was unnecessary because of limited collinearity (Figure S1); and all variables were included in the subsequent analysis. In case of a high number of variables, it is recommended to carry out this data exploration step.

3.2.2. K-means clustering with Moran's index

We resampled rasters to a spatial resolution of 3 m for data uniformity, prior to clustering and applied a mask to ensure that the field boundaries are uniform across all rasters. We subsequently stacked the rasters to form a multiband raster, with each band representing a single variable and converted to multidimensional arrays using the *numpy* package (ver. 1.20.1) where each row represented a pixel, and each column represented a variable. All invalid values were removed, and the rasters were scaled with *StandardScaler* (ver. 1.5.1) for the joint distribution of all bands to be considered equally. We applied the K-means clustering algorithm (in *scikit-learn* package, ver. 1.5.1) and examined it using the elbow method with the Within-Cluster Sum of Squares WCSS (see Figure S2 in the Supplementary Material). We determined the number of clusters to be optimal at $n = 4$, using various indices in addition to the WCSS method (see Table S4 in the Supplementary Material). We assigned the labels to each pixel in the raster. We remapped the array to a raster and plotted the raster with *matplotlib* (ver. 3.9.1), with and without soil variables. The clusters were visualized as a choropleth map and a cluster difference map was composed to highlight regions where clusters differed.

3.2.3. Variable weighing

We implemented variable weighting as an additional step, i.e., the method allows for assigning different levels of importance, or weights, to variables when grouping data into clusters. The end-user can decide whether they want to have all variables treated equally or assign a higher degree of relevance to specific ones. By weighting variables, more influential features can have a greater impact on the formation of clusters, leading to more meaningful groupings. This approach is expected to streamline the clustering depending on the variables that are given the maximum importance or are otherwise prioritized.

3.2.4. Spatial autocorrelation between clusters

We used spatial variograms to assess the spatial autocorrelation between clusters in both reduced and full datasets, afterwards computing the pairwise distances between points, and the semi-variance for each distance [20]. We divided the distance into bins (lags) and plotted the average semi-variance for each lag. Given the large dataset size, computing the full pairwise distance matrix for a large number of points would have required excessive amounts of memory, so we used subsampling of a random subset of points to capture the overall spatial structure of the data. Variograms for reduced and full datasets are available in the Supplementary Material (Figure S3 and Figure S4). Additionally, we validated the clusters with respect to proximity and spatial coherence (see Table S5 in Supplementary Material) using a Gaussian distance-decay function [21,22]. We validated the method by

means of 5 testing distances (between 50–800 m on a k-nearest neighbor graph). We recorded a cohesion ratio of approximately 97 %, which indicated highly contiguous and spatially coherent clusters across multiple scales [38–41].

3.2.5. Inspection of tester out-field reference plots

We calculated three similarity metrices for the clusters (in *Scipy* ver 1.6.2 and *Matplotlib*, ver.3.9.1), i.e., Euclidean distance (ED), Cosine similarity (CS) and Jensen-Shannon divergence (JSD), to better understand the distribution of areas where reference plots should not be established. Subsequently, we generated two pairs of tester out-field reference points ($n = 12$) d ad-hoc. The points were placed within the same cluster (stratum) and retained the same distance between the points, as the in-field reference points to mimic common approaches to sampling of agroforestry systems. We generated the values for each reference point in relation to each predictor for further inspection. To aid final decision-making, we applied a Student's *t*-test to compare in-field and out-field reference plots.

3.3. Code

The annotated code detailing each step of the method for reference site selection can be found on GitHub: https://github.com/nishita2695/ControlSiteAnalysis.git

4. Results

4.1. Summary data

The investigated environmental variables displayed significant spatial variability across the studied fields (Table 2). F3 had the highest average topsoil electrical resistivity values (40.18±7.02) and F5 the lowest (33.06±4.86) compared to the case-study (agroforestry) field (34.73±6.57). The electrical resistivity decreased along the soil profile depth. F2 had the highest mean soil moisture index (7.05±1.49) and F3 the lowest (5.47±0.87) relative to the case-study field (6.49±1.08). The gamma values, indicative of soil mineral composition, and a useful variable used in soil texture prediction [18], were broadly similar across all fields. We recorded the highest elevation difference at F4, with a range of 45 m, in comparison to the range for the target points equivalent to 18 m. Mean EVI and NDVI values were highest at F3 (0.28 and 0.76, respectively) with the mean of 0.26 and 0.72 for the target points. We recorded lower vegetation indices for the southern fields (5 and 6).

4.2. Cluster distribution based on agronomy and topography

The target points (in-field reference points) were situated in Cluster 1 (Fig. 4). The results we derived from the available topographic and agronomic data showed that substantial sections of F2, lowermost sections of F3 and uppermost section of F5 and F6 were characterised by similar site conditions to the target points, i.e., these areas constituted a stratum to which out-field reference plots should be constrained.

4.3. Change in cluster distribution in response to proximal sensor-derived soil data

We found that the reference site selection, in terms of constraining selection to a single stratum (Cluster 1) was largely transferable across the reduced (Fig. 4) and full (Fig. 5A) datasets. The difference map (Fig. 5B) indicated the differences in cluster distribution, which reflected underlying environmental variability in respect to soil conditions.

4.4. Identification of areas deemed unsuitable to act as reference sites

We calculated similarity indices calculated for different clusters to identify locations deemed unsuitable for setting up out-field reference plots. In case of the reduced dataset, we found Cluster 3 to be dissimilar to Clusters 1 (large ED of 28.90 and moderate JSD of 0.0039), to Cluster 2 (moderate ED of 16.43 and moderate JSD of 0.0016) and to Cluster 4 (large ED of 20.12 and high JSD of 0.0060). We therefore excluded it from reference site selection. In case of the full dataset, Clusters 1 and 2 (moderate ED of 40.26 and high CS of 0.9857) and Clusters 1 and 3 (large ED of 67.47 and CS of 0.9517) had high to moderate similarity indices. Cluster 4 was dissimilar to all clusters (large EDs, negative CS, and infinite JSD values) and was an outlier. For detailed breakdown of values, see Table S3 in the Supplementary Material.

Table 3

Table 2
Range and mean ± SD (in brackets) of soil, topographic, and agronomic properties calculated for six fields (F1–6, see Fig. 4) at Gladbacherhof Farm, Germany. Abbreviations: Rh1 – Geophilus-derived proxy for electrical resistivity at a depth of 0–30 cm, Rh2 - Electrical resistivity (30–60 cm), Rh3 - Electrical resistivity (60–90 cm), GM – Topsoil gamma activity, SMI – Soil moisture index, ELE - Elevation, SLP - Slope, TWI – Topographical wetness index, TC - Terrain curvature, EVI - Enhanced vegetation index, NDVI – Normalised difference vegetation index.

Properties	F1	F2	F3	F4	F5	F6
Rh1 (Ohm m^{-1})	16.58 – 83.83 (34.73 ±6.57)	8.51 – 148.69 (32.24 ±6.88)	18.06 – 71.85 (40.18 ±7.02)	20.39 – 44.05 (31.68 ±4.97)	18.73 – 44.79 (33.06 ±4.86)	15.02–87.62 (33.39 ±7.64)
Rh2 (Ohm m^{-1})	23.02 – 83.87 (43.69 ±8.57)	11.55 – 152.59 (45.66 ±15.50)	8.04 – 271.57 (47.14 ±31.12)	15.82 – 227.46 (37.56 ±27.90)	16.67–98.01 (31.56 ±8.85)	15.53 – 95.28 (36.54 ±63.23)
Rh3 (Ohm m^{-1})	22.22 – 79.81 (41.62 ±8.19)	12.29 – 91.13 (35.28 ±7.30)	7.90 – 772.47 (134.34 ±100.02)	8.05 – 489.60 (48.21 ±62.94)	10.97 – 174.40 (33.36 ±17.79)	12.79 – 98.25 (44.31 ±137.97)
GM (counts s^{-1})	1.96 – 2.19 (2.19±0.08)	1.98 – 2.30 (2.19 ±0.06)	2.02 – 2.49 (2.14 ±0.05)	1.97 – 2.45 (2.16 ±0.09)	1.93 – 2.26 (2.14±0.06)	1.90 – 2.31 (2.18±0.06)
SMI (-)	2.67 – 13.54 (6.49 ±1.08)	1.45 – 26.07 (7.05 ±1.49)	2.99 – 12.17 (5.47 ±0.87)	4.77 – 11.16 (7.00 ±1.29)	4.48 – 11.08 (6.61 ±1.05)	1.69 – 14.40 (6.80 ±1.27)
ELE (m)	165.00–183.00 (173.98 ±3.87)	167.00 – 185.00 (178.93±4.23)	172.00 – 199.00 (188.95±5.53)	180.00 – 225.00 (199.34±12.90)	172.00 – 197.00 (184.79 ±6.33)	171.00 – 206.00 (190.53 ±8.71)
SLP (°)	0.00 – 29.15 (5.75 ±8.80)	0.00 – 28.85 (3.45 ±7.40)	0.00 – 29.05 (4.13 ±7.95)	0.00 – 28.68 (5.83 ±9.22)	0.00 – 29.06 (8.09 ±9.83)	0.00 – 29.02 (7.03 ±9.75)
TWI (-)	0.68 – 15.38 (7.25 ±2.75)	0.00 – 15.64 (7.82 ±2.69)	0.00 – 16.13 (7.74 ±3.18)	0.00 – 13.85 (7.38 ±2.85)	0.00 – 29.34 (7.87 ±9.85)	0.00 – 15.36 (7.30 ±2.95)
TC (-)	0.00 – 29.27 (5.82 ±8.79)	0.00 – 29.36 (3.43 ±7.45)	0.00 – 29.49 (4.24 ±9.00)	0.00 – 29.49 (10.39 ±5.81)	0.00 – 29.24 (7.91 ±9.71)	0.00 – 29.21 (7.10 ±9.80)
EVI (-)	0.19 – 0.33 (0.26±0.02)	0.21 – 0.33 (0.26 ±0.02)	0.17 – 0.34 (0.28 ±0.03)	0.14 – 0.35 (0.25 ±0.05)	0.18 – 0.28 (0.22±0.01)	0.12 – 0.33 (0.22 ± 0.03)
NDVI (-)	0.66 – 0.87 (0.74±0.03)	0.65 – 0.84 (0.73 ±0.03)	0.63 – 0.85 (0.76 ±0.04)	0.47 – 0.86 (0.72 ±0.10)	0.65 – 0.75 (0.68±0.01)	0.57 – 0.83 (0.68 ± 0.05)

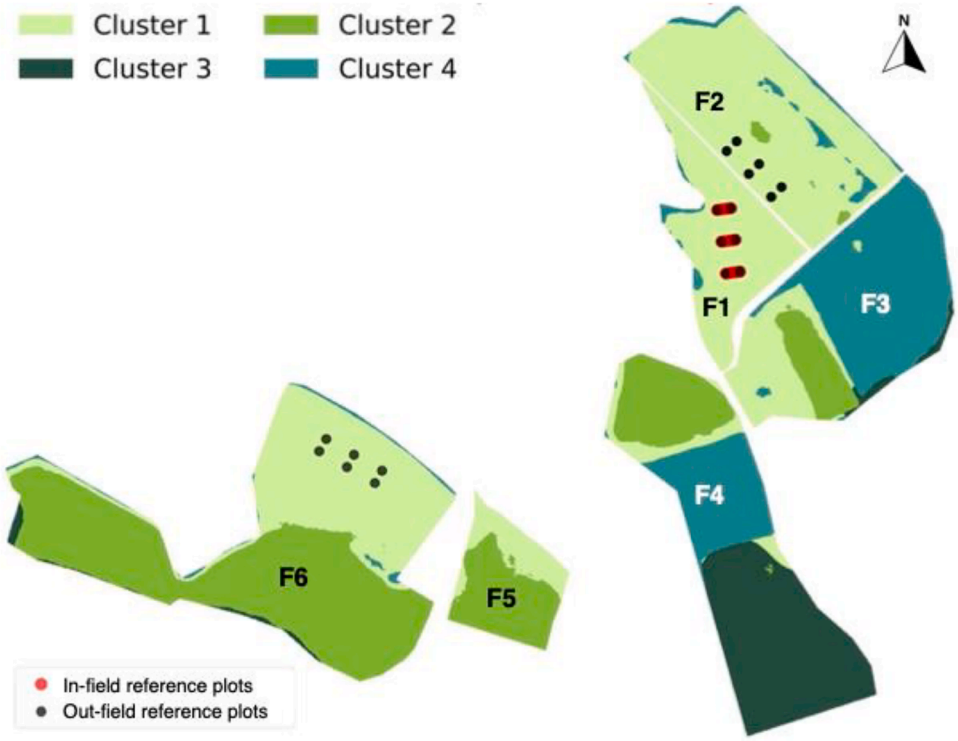


Fig. 4. Cluster distribution ($n = 4$) based on agronomic and topographic predictors ($n = 8$), without proximal sensor-derived soil variables ($n = 4$). Exemplary in-field reference plots in red and out-field in black.

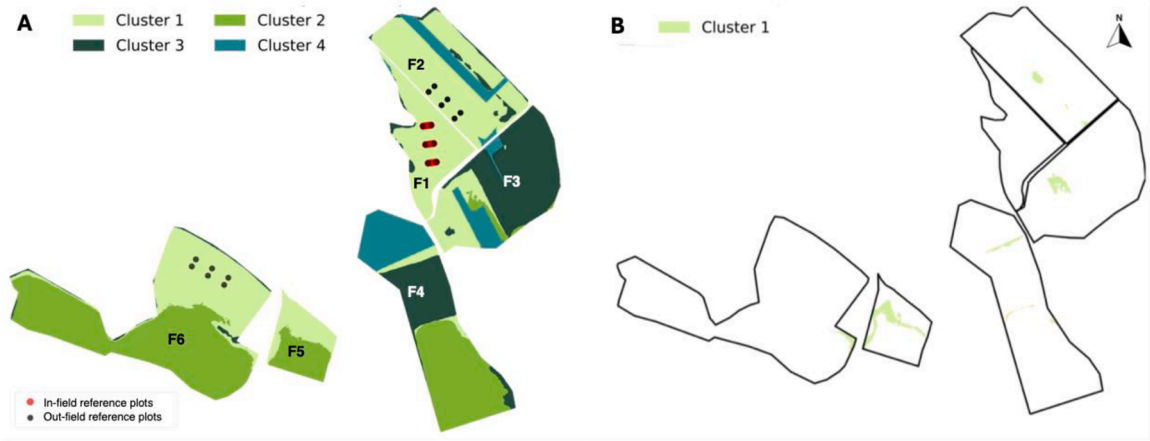


Fig. 5. A-B. Cluster distribution ($n = 4$) involving proximal sensor-derived soil variables ($n = 4$) showing similarities between in-field and out-field plots (A) and the difference map (indicating Cluster 1 differences) between reduced and full dataset-derived clusters (B).

4.5. Further inspection of tester in-field and out-field reference points

We found that the in-field and out-field reference points were comparable across most investigated predictors with the exception of elevation and electrical resistivity (Table 4). Recorded elevation at in-field reference points was significantly lower ($M = 174.06$, SD = 1.93) than that recorded for the out-field reference in F2 ($M = 179.33$, SD = 1.03), $t_{(5)} = -5.9$, $p < 0.05$). Similarly, elevation recorded at in-field reference points was significantly lower than for the out-field reference points in F6 ($M = 180$, SD = 1.67), $t_{(5)} = -5.7$, $p < 0.05$). The soil moisture index was found to be lower for the in-field reference plots ($M = 6.30$, SD = 1.17) in comparison to the out-field reference points in F2 ($M = 7.79$, SD = 0.82), $t_{(5)} = -2.6$, $p = 0.03$). In-field reference plots showed a significantly lower soil electrical resistivity for soil depth of 0.30–0.60 cm and 0.60–0.90 cm ($M = 44.36$, SD = 7.36 and $M = 42.41$, SD = 6.42, respectively) in comparison to the out-field reference points in F6 ($M = 30.99$, SD = 2.06 and $M = 32.77$, SD = 1.85, respectively), $t_{(5)} = -2.6$, $p = 0.01$ and $t_{(5)} = 3.5$, $p = 0.01$, respectively. For the full breakdown of the Student *t*-test values, see Table S3 in the Supplementary Material.

5. Discussion

Our primary objective in this study was to develop a data-driven approach that selected areas best suited to establishing reference plots for long-term agricultural experiments, to reduce the risk of mistakenly attributing effects of farm-scale heterogeneity to that of land use change. We used agroforestry as a relevant case study owing to the widely

Table 3
Similarity among clusters to identify locations unsuitable to act as reference sites.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Outliers
Reduced dataset					
Cluster 1		+	−	+	
Cluster 2	+		−	+	
Cluster 3	−	−		−	
Cluster 4	+	+	−		
Full dataset					
Cluster 1		+	+	−	
Cluster 2	+		+	−	
Cluster 3	+	+		−	
Cluster 4	−	−	−		+

published and fresh doubt in land use research over whether the methods used in prior studies were sufficiently controlled and the evidence produced still valid [23,24]. Our study involved an in-depth examination of the outputs e.g., in case when proximal sensor data was included in the analysis, which was carried out as part of the approach.

5.1. Proximity as a guide for reference site selection

Tobler's Law states that 'Everything is related to everything, but near things are more related than distant things` [16]. We hypothesised that distance will not be the sole determinant of suitable reference sites. Our hypothesis was supported by the results shown in this study. When different environmental factors were accounted for, we saw that only sections of neighbouring fields had similar site conditions to the target in-field reference points. Similarly, we identified that fields that were not neighbouring could be suitable for establishment of out-field reference plots.

Our method demonstrated the potential for application of geodata as well as proximal sensor data. However, this approach was developed to be adapted to specific research needs. For instance, Mandal et al. [25] show that in experiments involving soil properties such as soil organic carbon, other data types should be additionally considered, e.g., geological formations and soil types [26,27] or land use history [15]. In studies of crops, the distance to small woody landscapes such as hedges or agroforestry treelines might result in reduction of photosynthetically active radiation due to shading [28], which negatively impacts (particularily, C4) species such as maize and sorghum [29,44] or changes in microclimate [14]. Distance to the field edge was also found to be an important factor in crop studies [9]. These variables should constitute additional filters to identify locations within clusters e.g., a pixelwise buffer calculation of distance could be used for identification of areas nearby to small woody features, forests or water bodies, which could act as robust out-field reference points at smaller-scales, i.e., plot as opposed to the field-scale.

Investigators should curate geodata in terms of spatial and temporal scales in accordance with the research question. For example, while freely available Sentinel images with a resolution of 10 m might be suitable for the presented method, they were considered inadequate for the selection of targeted in-field reference points in this study. Specifically, with a spacing of 21 m between reference points (in the east-west direction), satellite images with a 10-meter resolution might not capture spatial variability at an adequate scale. Investigators should consider the potential impact of Sentinel imagery resolution limitations on small-scale variability detection [30] at the planning stage. For some applications, e.g., plot-scale experiments, fine-scale geodata obtained through Unmanned Aerial Vehicles [31], might be preferable. Finally, the temporal stability of parameters, e.g., when dynamic variables which change from year to year such as NDVI or EVI are selected [17], as well as measurements, e.g., when they involve soil moisture assessment with proximal soil sensors [32], is a factor that should be considered.

5.2. Proximal sensor data and assessment of soil conditions

Our inclusion of proximal sensor-derived soil data had limited effect on our ability to select suitable areas for establishment of out-field reference plots. This might be due to (1) extrapolation of data, i.e., from point-to-raster which reduces the resolution and introduces errors with implications for accuracy in the assessment of heterogeneity of soil conditions, and (2) overall homogeneity of soil conditions, i.e., Rh values (as points) were shown to be relatively uniform across the investigated fields. These results are site-specific and are not generalisable. Overall, we recommend investigators to obtain soil information prior to sampling campaigns to account for different conditions. For example, in agroforestry research, organic carbon constitutes a frequently measured parameter. Organic carbon retention and sequestration potential are strongly affected by soil texture [33], which should be similar between the experimental and reference areas. Through application of proximal sensing technology, these differences can be quantified at very high-resolution. However, whereas recent studies have highlighted the stability of proximal sensing measurements across different time points [32], we found the quality of proximal soil sensor data important. As a result of missing values due to higher-than-average soil moisture content at the time of measurement, we interpolated across fragments of F2, F3 and F4 at larger distances. This resulted in the field fragments being assigned to Cluster 4, which was an extreme outlier. In such scenarios, we recommend another field campaign involving proximal soil sensing during more appropriate time of the year. In addition, other platforms and methods such as EMI (electromagnetic induction; [34]), LiDAR-based soil moisture mapping [35] or simple soil moisture maps might lead to greater differentiation between fields.

5.3. Cluster similarity and tester out-field reference plots

Cluster similarity provides information, which might be useful in situations when fields corresponding to the preferred option for

Table 4
Parameter values for individual in-field and out-field reference points. **Abbreviations: M – Mean, SD – Standard deviation, GM – Topsoil gamma activity, Rh1 – Geophilus-derived proxy for electrical resistivity at a depth of 0–30 cm, Rh2 - Electrical resistivity (30–60 cm), Rh3 - Electrical resistivity (60–90 cm), SMI – Soil moisture index, ELE - Elevation, SLP - Slope, TC - Terrain curvature, TWI – Topographical wetness index, EVI - Enhanced vegetation index, NDVI – Normalised difference vegetation index.**

ID	GM	Rh1	Rh2	Rh3	SMI	ELE	SLP	TC	TWI	EVI	NDVI
In-field reference points											
M	2.15	35.02	44.36	42.41	6.30	174.06	10.86	20.50	6.60	0.26	0.72
SD	0.09	5.54	7.36	6.42	1.17	1.93	12.70	42.82	2.94	0.01	0.01
Out-field reference points (F2)											
M	2.24	29.04	50.60	36.10	7.79*	179.33*	10.75	14.81	5.92	0.26	0.72
SD	0.01	3.05	7.15	3.29	0.82	1.03	11.48	40.77	3.56	0.00	0.01
Out-field reference points (F6)											
M	2.22	31.90	30.99*	32.74*	6.99	180.00*	4.43	5.56	8.45	0.25	0.72
SD	0.02	2.69	2.06	1.85	0.57	1.67	10.84	13.61	3.17	0.00	0.00

establishment of out-field reference plots (Cluster 1 in this example) are not available due to, e.g., different stages of crop rotation, land rights (suitable areas belonging to another farm) or farm management (organic vs conventional). When it is not possible to establish out-field reference plots in the same strata, avoiding extremely dissimilar strata might prove to be the best strategy. However, in these circumstances, the variation between strata should be quantified via dedicated field surveys prior to sampling campaigns.

As a final step of the analysis, we recommend investigators perform a visual inspection and statistical comparisons to ensure that the selected out-field reference plots match the conditions of in-field reference plots across key parameters. For example, our analysis showed that elevation was significantly different between in-field and out-field reference plots. We found that elevation affected crop yield in agroforestry systems across small (single field) scales [36]. Thus, under real-life conditions, elevation might be allocated a higher weighting to optimise the field-to-field comparisons. Similarly, in studies targeting soil chemistry, e.g., soil organic carbon, detectable differences in soil resistivity might disqualify certain areas from establishing out-field reference plots.

6. Conclusions and future directions

Research into suitability of reference systems remains limited – so far, only one other study described a similar approach of using spatial and weather data to identify high quality controls [37]. In contrast to the research by Das and colleagues (2024), who focus on benchmarking for soil health assessment at the eco-region scale, our approach targets small-scale agricultural applications and provides novel insights into the best-practice experimental design. Our method constitutes a demonstration of how geodata and proximal soil sensing data can be used for reference site selection. We however, developed our method, including the available code, to be modified in response to specific research questions. A follow-up study, which could be used to validate the presented methodology, should involve (1) a research question with targeted variables which are responsive to site heterogeneity and are likely to impact the outcomes of the study, and (2) dedicated field surveys, following the desktop assessment, to confirm whether the selected candidate reference sites are suitable. The costs associated with performing field surveys of experimental and reference sites are justified with a view of implementing long-term agricultural experiments.

CRediT authorship contribution statement

Nishita Thakur: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation. **Marco Donat:** Writing – review & editing, Validation, Investigation. **Sonoko Bellingrath-Kimura:** Writing – review & editing, Project administration, Funding acquisition. **Wiebke Niether:** Writing – review & editing. **Andreas Gattinger:** Writing – review & editing, Project administration, Funding acquisition. **Deise Aline Knob:** Data curation. **Eva-Maria L. Minarsch:** Writing – review & editing. **Philipp Weckenbrock:** Writing – review & editing. **Franz Schulz:** Data curation. **Lutz Breuer:** Writing – review & editing, Supervision, Funding acquisition. **Suzanne Jacobs:** Writing – review & editing. **John Clifton-Brown:** Writing – review & editing. **Karolina Golicz:** Writing – original draft, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Funding

This work was supported by the project Agroforstsysteme Hessen II funded by the Hessische Ministerium für Landwirtschaft und Umwelt, Weinbau, Forsten, Jagd und Heimat, Grant No 895-0736-23-511-8-028, the DFG Research Unit 5664 'Agroforestry for sustainable multifunctional agriculture (FORMULA)', Grant No BR2238/44-1 and the project GreenDairy (Integrated animal-crop agricultural systems) funded by the Hessian Ministry of Higher Education, Research, and the Arts, Grant No LOEWE/2/14/519/03/07.001-(0007)/80.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.atech.2025.101357.

Data availability

Data will be made available on request.

References

[1] E.-M.L. Minarsch, P. Schierning, F. Wichern, et al., Transect sampling for soil organic carbon monitoring in temperate alley cropping systems - A review and standardized guideline, Geoderma Reg. 36 (2024) e00757, https://doi.org/10.1016/j.geodrs.2024.e00757.
[2] J.P. McCalmont, A. Hastings, N.P. McNamara, et al., Environmental costs and benefits of growing *Miscanthus* for bioenergy in the UK, GCB Bioenergy 9 (2017) 489–507, https://doi.org/10.1111/gcbb.12294.
[3] M. Laub, L. Pataczek, A. Feuerbacher, et al., Contrasting yield responses at varying levels of shade suggest different suitability of crops for dual land-use systems: a meta-analysis, Agron. Sustain. Dev. 42 (2022) 51, https://doi.org/10.1007/s13593-022-00783-7.
[4] L.R. Walker, D.A. Wardle, R.D. Bardgett, B.D. Clarkson, The use of chronosequences in studies of ecological succession and soil development, J. Ecol. 98 (2010) 725–736, https://doi.org/10.1111/j.1365-2745.2010.01664.x.
[5] S.M. Bell, S.J. Raymond, H. Yin, et al., Quantifying the recarbonization of post-agricultural landscapes, Nat. Commun. 14 (2023) 2139, https://doi.org/10.1038/s41467-023-37907-w.
[6] K. Golicz, H.-P. Piepho, E.-M.L. Minarsch, et al., Highlighting the potential of multilevel statistical models for analysis of individual agroforestry systems, Agrofor. Syst. (2023), https://doi.org/10.1007/s10457-023-00871-x.
[7] P. Balandier, C. Dupraz, Growth of widely spaced trees. A case study from young agroforestry plantations in France, Agrofor. Syst. 43 (1998) 151–167, https://doi.org/10.1023/A:1026480028915.
[8] P.T. Birteeb, C. Varghese, S. Jaggi, et al., An efficient class of tree network balanced designs for agroforestry experimentation, Commun. Stat. B: Simul. Computat. 51 (2022) 7169–7179, https://doi.org/10.1080/03610918.2020.1825739.
[9] W.N.W. Fincham, J.W. Redhead, B.A. Woodcock, R.F. Pywell, Exploring drivers of within-field crop yield variation using a national precision yield network, J. App. Ecol. 60 (2023) 319–329, https://doi.org/10.1111/1365-2664.14323.
[10] P. Pardon, B. Reubens, J. Mertens, et al., Effects of temperate agroforestry on yield and quality of different arable intercrops, Agric. Syst. 166 (2018) 135–151, https://doi.org/10.1016/j.agsy.2018.08.008.
[11] A. Swieter, M. Langhof, J. Lamerre, Competition, stress and benefits: trees and crops in the transition zone of a temperate short rotation alley cropping agroforestry system, J. Agronomy Crop Sci. 208 (2022) 209–224, https://doi.org/10.1111/jac.12553.
[12] A. Swieter, M. Langhof, J. Lamerre, J.M. Greef, Long-term yields of oilseed rape and winter wheat in a short rotation alley cropping agroforestry system, Agroforest Syst. 93 (2019) 1853–1864, https://doi.org/10.1007/s10457-018-0288-5.
[13] C. Böhm, M. Kanzler, D. Freese, Wind speed reductions as influenced by woody hedgerows grown for biomass in short rotation alley cropping systems in Germany, Agroforest Syst 88 (2014) 579–591, https://doi.org/10.1007/s10457-014-9700-y.
[14] S.R. Jacobs, H. Webber, W. Niether, et al., Modification of the microclimate and water balance through the integration of trees into temperate cropping systems, Agric. For. Meteorol. 323 (2022) 109065, https://doi.org/10.1016/j.agrformet.2022.109065.
[15] D. Emde, C. Poeplau, A. Don, et al., The centennial legacy of land-use change on organic carbon stocks of German agricultural soils, Glob. Chang. Biol. 30 (2024) e17444, https://doi.org/10.1111/gcb.17444.
[16] W.R. Tobler, A computer movie simulating urban growth in the Detroit region, Econ. Geogr. 46 (1970) 234–240, https://doi.org/10.2307/143141.
[17] S.A. Shammi, Q. Meng, Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling, Ecol. Indic. 121 (2021) 107124, https://doi.org/10.1016/j.ecolind.2020.107124.
[18] E. Lueck, J. Ruehlmann, Resistivity mapping with GEOPHILUS ELECTRICUS — Information about lateral and vertical soil heterogeneity, Geoderma 199 (2013) 2–11, https://doi.org/10.1016/j.geoderma.2012.11.009.
[19] T. Liu, K. Juang, D. Lee, Interpolating soil properties using kriging combined with categorical information of soil maps, Soil Sci. Soc. Am. J. 70 (2006) 1200–1209, https://doi.org/10.2136/sssaj2005.0126.
[20] X. Guo, S. Kurtek, K. Bharath, Variograms for kriging and clustering of spatial functional data with phase variation, Spat. Stat. 51 (2022) 100687, https://doi.org/10.1016/j.spasta.2022.100687.

[21] H. Morlon, G. Chuyong, R. Condit, S. Hubbell, D. Kenfack, D. Thomas, R. Valencia, J.L. Green, A general framework for the distance-decay of similarity in ecological communities, Ecol. Lett. 11 (9) (2008) 904–917, https://doi.org/10.1111/j.1461-0248.2008.01202.x.
[22] A. Natarajan, M. De Iorio, A. Heinecke, E. Mayer, S. Glenn, Cohesion and repulsion in bayesian distance clustering, J. Am. Stat. Assoc. 119 (546) (2023) 1374–1384, https://doi.org/10.1080/01621459.2023.2191821.
[23] S. Boinot, K. Barkaoui, D. Mézière, et al., Research on agroforestry systems and biodiversity conservation: what can we conclude so far and what should we improve? BMC Ecol Evo 22 (2022) 24, https://doi.org/10.1186/s12862-022-01977-z.
[24] A.-C. Mupepele, C.F. Dormann, Reply to: "research on agroforestry systems and biodiversity conservation: what can we conclude so far and what should we improve?" by Boinot et al. 2022, BMC Ecol Evo 22 (2022) 65, https://doi.org/10.1186/s12862-022-02016-7.
[25] A. Mandal, A. Majumder, S.S. Dhaliwal, et al., Impact of agricultural management practices on soil carbon sequestration and its monitoring through simulation models and remote sensing techniques: a review, Crit. Rev. Environ. Sci. Technol. 52 (2022) 1–49, https://doi.org/10.1080/10643389.2020.1811590.
[26] P.-T. Guo, M.-F. Li, W. Luo, et al., Digital mapping of soil organic matter for rubber plantation at regional scale: an application of random forest plus residuals kriging approach, Geoderma 237–238 (2015) 49–59, https://doi.org/10.1016/j.geoderma.2014.08.009.
[27] M. Wiesmeier, L. Urbanski, E. Hobley, et al., Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales, Geoderma 333 (2019) 149–162, https://doi.org/10.1016/j.geoderma.2018.07.026.
[28] M. Donat, J. Geistert, K. Grahmann, Bellingrath-Kimura SD, Orientation of tree rows in alley cropping systems matters – The "ShadOT" modelling tool for tree growth and shading effects, MethodsX. 11 (2023) 102282, https://doi.org/10.1016/j.mex.2023.102282.
[29] A. Dilla, P.J. Smethurst, K. Barry, et al., Potential of the APSIM model to simulate impacts of shading on maize productivity, Agroforest Sys 92 (2018) 1699–1709, https://doi.org/10.1007/s10457-017-0119-0.
[30] J. Segarra, M.L. Buchaillot, J.L. Araus, S.C. Kefauver, Remote sensing for precision agriculture: sentinel-2 improved features and applications, Agronomy 10 (5) (2020) 641, https://doi.org/10.3390/agronomy10050641.
[31] M. Wengert, H.-P. Piepho, T. Astor, et al., Assessing spatial variability of barley whole crop biomass yield and leaf area index in silvoarable agroforestry systems using UAV-borne remote sensing, Remote Sens. (Basel) 13 (2021) 2751, https://doi.org/10.3390/rs13142751.
[32] L.S. Conway, K.A. Sudduth, N.R. Kitchen, S.H. Anderson, Repeatability of commercially available visible and near infrared proximal soil sensors, Precision Agric. 24 (2023) 1014–1029, https://doi.org/10.1007/s11119-022-09985-1.
[33] A. Tariq, A. Gunina, N. Lamersdorf, Initial changes in soil properties and carbon sequestration potential under monocultures and short-rotation alley coppices with poplar and willow after three years of plantation, Sci. Total Environ. 634 (2018) 963–973, https://doi.org/10.1016/j.scitotenv.2018.03.391.
[34] H. Shaukat, K.C. Flowe, M. Leopold, et al., Comparing quasi-3D soil moisture derived from electromagnetic induction with 1D moisture sensors and correlation to barley yield in variable duplex soil, Soil and Tillage Res. 236 (2024), https://doi.org/10.1016/j.still.2023.105953. ISSN 0167-1987.
[35] A.M. Ågren, J. Larson, S.S. Paul, H. Laudon, W. Lidberg, et al., Use of multiple LIDAR-derived digital terrain indices and machine learning for high-resolution national-scale soil moisture mapping of the Swedish forest landscape, Geoderma 404 (2021) 0016–7061, https://doi.org/10.1016/j.geoderma.2021.115280.
[36] M. Habibur-Rahman, A. Raza, H.E. Ahrends, et al., Impact of in-field soil heterogeneity on biomass and yield of winter triticale in an intensively cropped hummocky landscape under temperate climate conditions, Precis. Agric. 23 (2022) 912–938, https://doi.org/10.1007/s11119-021-09868-x.
[37] S. Das, A. Hird, B. Maharjan, et al., Reference site selection based on state-and-transition models for soil health gap evaluation within cropland reference ecological units, Soil Security 16 (2024) 100142, https://doi.org/10.1016/j.soisec.2024.100142.
[38] F.S. Dias, M. Betancourt, P.M. Rodríguez-González, et al., Analysing the distance decay of community similarity in river networks using Bayesian methods, Sci. Rep. 11 (2021) 21660, https://doi.org/10.1038/s41598-021-01149-x.
[39] S. Watson, Efficient design of geographically-defined clusters with spatial autocorrelation, J. Appl. Stat. 49 (13) (2021) 3300–3318, https://doi.org/10.1080/02664763.2021.1941807.
[40] MH ur Rahman, Ahrends HE, A. Raza, T. Gaiser, Current approaches for modeling ecosystem services and biodiversity in agroforestry systems: challenges and ways forward, Front. For. Glob. Change 5 (2023), https://doi.org/10.3389/ffgc.2022.1032442.
[41] A.K. Singh, S. Mittal, P. Malhotra, Y.V. Srivastava, Clustering evaluation by Davies-Bouldin Index(DBI) in cereal data using K-means, in: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 306–310.
[42] E. Boenecke, E. Lueck, J. Ruehlmann, et al., Determining the within-field yield variability from seasonally changing soil conditions, Precision Agric. 19 (2018) 750–769, https://doi.org/10.1007/s11119-017-9556-z.
[43] P. Harris, C. Brunsdon, M. Charlton, Geographically weighted principal components analysis, Int. J. Geogr. Inf. Sci. 25 (10) (2011) 1717–1736, https://doi.org/10.1080/13658816.2011.554838.
[44] P. Pardon, P. Quataert, J. Bracke, et al., Crop yield in young temperate alley cropping systems is affected by tree height, distance to trees and crop type, while impacts on crop quality remain limited, Agroforest Syst. 99 (2025) 162, https://doi.org/10.1007/s10457-025-01260-2.