Soil spatial variation to guide the development of fertilizer use recommendations for smallholder farms in western Kenya

Article *in* Geoderma Regional · May 2020

DOI: 10.1016/j.geodrs.2020.e00300

Journal Pre-proof

Soil spatial variation to guide the development of fertilizer use recommendations for smallholder farms in western Kenya

Stephen M. Ichami, Keith D. Shepherd, Ellis Hoffland, George N. Karuku, Jetse J. Stoorvogel

PII: S2352-0094(20)30049-3

DOI: https://doi.org/10.1016/j.geodrs.2020.e00300

Reference: GEODRS 300

To appear in: *Geoderma Regional*

Received date: 10 December 2018

Revised date: 17 April 2020

Accepted date: 19 April 2020

Please cite this article as: S.M. Ichami, K.D. Shepherd, E. Hoffland, et al., Soil spatial variation to guide the development of fertilizer use recommendations for smallholder farms in western Kenya, *Geoderma Regional* (2019), https://doi.org/10.1016/j.geodrs.2020.e00300

Soil spatial variation to guide the development of fertilizer use recommendations for smallholder farms in western Kenya

Stephen M. Ichami[1 a, b, c, d, e], Keith D. Shepherd[b], Ellis Hoffland[d], George N. Karuku[a], Jetse J. Stoorvogel[e].

[a] Department of Land Resource Management and Agricultural Technology
College of Agriculture and Veterinary Medicine
Faculty of Agriculture
University of Nairobi,
P.O Box 29053 - 00625, Kabete
Nairobi, Kenya
stephen.ichami@gmail.com; gmoe@unobi.ac.ke

[b] World Agroforestry Centre (ICRAF)
United Nations Avenue, Gigiri
P.O. Box 30677 - 00100
Nairobi, Kenya
k.shepherd@cgiar.org

[c] International Centre for Tropical Agriculture (CIAT-Kenya)
DuduVille Lane off Kasarani Road
P.O. Box 823 - 00621
Nairobi, Kenya
sichami@cgiar.org

[d] Soil Biology Group

[1] Corresponding author (email: *stephen.ichami@gmail.com*; Tel: +254727610349/+254735568410)

Wageningen University
P.O. Box 47
6700AA Wageningen
The Netherlands
ellis.hoffland@wur.nl

[e] Soil Geography and Landscape Group
Wageningen University
P.O. Box 47
6700AA Wageningen
The Netherlands
jetse.stoorvogel@wur.nl

Abstract

A farm survey was conducted within a 100 km^2 sampling block to collect data on the spatial variation in unfertilized maize biovolume and grain yields in relation to soil organic carbon (SOC), total nitrogen, phosphorus and extractable cations. Key soil factors associated with crop performance were identified using stepwise multiple linear regression modelling. The spatial variation of key soil factors and crop performance indicators (CPIs) was described in terms of spatial dependency. An analysis of variance indicated the variation explained by soil types, sampling units, and administrative units. Soil properties displayed high variability with coefficients of variation of in the range of 50% to 89% for extractable nutrients. Grain yield ranged widely from 0.1 to 11.3 t ha^{-1}, with 31% of the variation being accounted for by measured soil properties. SOC was identified as key soil factor associated with variation in crop performance. SOC displayed moderate spatial dependency with a range of 523 m. Analysis of variation indicate that variation in SOC was sufficiently described by small spatial units (fields). These insights were used to provide a framework for determining an appropriate scale for developing digital soil maps or distance for soil sampling in heterogenous smallholder farming systems. Strategies aimed at refining fertilizer use recommendation can therefore use this guideline.

Key words: farm survey; scale of variation; maize; fertilizer recommendations; digital soil mapping.

1. Introduction

Agricultural production in sub-Saharan Africa (SSA) can be characterized by smallholder farming and low productivity. The latter is caused by low inherent soil fertility (Bekunda *et al.*, 2010), soil nutrient depletion (Stoorvogel *et al.*, 1993), limited nutrient inputs, and poor germplasm (Vanlauwe *et al.*, 2011). Fertilizers are required to replenish soil nutrient stocks and provide nutrients to the crop to increase productivity. There is increasing interest from governments and fertilizer companies in developing specific fertilizer products and blends targeted to specific geographies. Smallholder farmers also face the basic question of what type and how much fertilizer to apply given the local conditions on their farm and available resources. They rely on their own experience from previous years and fertilizer use recommendations. Typically, the recommendations are spatially coarse and developed on the basis of soil surveys and agronomic experiments. The recommendations are valid for the area or a spatial unit, for which the experiment is considered representative, which could be an administrative unit such as country or an agro-ecological zone (AEZ). These, so-called, blanket fertilizer use recommendations are a single fertilizer use recommendation for a given area that do not account for the variation in conditions within that area.

Currently, several countries still use blanket fertilizer use recommendations to guide decisions on nutrient management. In the past decades, many studies have focused on refining or improving fertilizer recommendations, with the aim of attaining higher crop yields (*e.g.,* Mowo and Mlingano, 1993). In the 1980s, for example, countries such as Kenya provided blanket fertilizer use recommendations for the entire country to guide decisions by smallholder farmers on their fertilizer management options (AIC, 1981). Later, in the 1990s, fertilizer use recommendations were refined on the basis of agro-ecological zones in an attempt to deal with within-country variation through the Fertilizer Use Recommendation Project (FURP) in Kenya (FURP, 1994). Agro-ecological zones were defined based on climate, soil, and topography (Geurts and Van den Berg, 1998). However, the variability in growing conditions within AEZs can limit the efficiency_of fertilizer use recommendations developed at the AEZ-level (Giller *et al.*, 2006). For example, Diarisso *et al.* (2015) reviewed soil spatial heterogeneity in smallholder landscapes and soil responsiveness to interventions; they concluded that a form of precision agriculture is required that recognizes fine scale spatial heterogeneity.

Many recent studies focus on improving fertilizer management and include the use of decision support tools such as the Quantitative Evaluation of Fertility in Tropical Soils (QUEFTS) model (Janssen *et al.*, 1990), the derived Nutrient Expert (Pampolino *et al.*, 2012), and crop growth simulation models like APSIM (Kisaka *et al.*, 2016). These tools have to be re-calibrated for every region of interest (Xu *et al.*, 2013). Furthermore, Molefe *et al.*, (2012) observed that such decision support tools fail to capture the complexity within smallholder farms. Matthews *et al.*, (2002) reported poor quality of data limits the application of such nutrient management tools in smallholder farms.

A further refinement of fertilizer recommendations is hampered by the lack of detailed soil data (Sanchez *et al.*, 2009). New, more detailed soil surveys are rare, particularly in SSA. The available national soil survey maps are spatially course (*e.g.*, 1:250,000 to 1:1 million), and are produced using different methods, resulting to varying levels of accuracy (regional or national) and data incompleteness (Baruck *et al.*, 2016). Two new developments in the collection of soil data may create new options to refine fertilizer recommendations even further:

i). Digital soil mapping (McBratney *et al.*, 2003) has evolved into an operational tool that can provide detailed insight in soil variability in an efficient way. Examples include the

100 m resolution digital soil map of Machakos and Makueni district in Kenya (Mora-Vallejo *et al.*, 2008) but also various continental to global initiatives (*e.g*. Stoorvogel *et al.*, 2017).

ii). Fertilizer recommendations for a farm can be based on soil test values for that farm. Where traditional soil analysis is often too expensive and, therefore, out of reach for smallholder farmers, new proximal sensing techniques like infrared spectroscopy (Shepherd *et al.*, 2015) can be used to provide soil analysis at a low cost.

A better understanding of soil spatial variability may provide guidelines for refining fertilizer use recommendations to optimize crop productivity. Those guidelines should include at which scale fertilizer use recommendations need to be developed.

This study seeks to develop an approach to assess, ex-ante, the optimal level of scale, that reflects the variability in local growing conditions in smallholder farms. The objectives of this study were: (i) to describe the spatial variability of soil properties and crop performance, (ii) to identify key soil factors associated with crop performance, and (iii) identify a scale in which soil spatial variability can be sufficiently described. Our hypothesis is that the scale of variability of key soil factors corresponds to that of fertilizer response across smallholder farming landscapes. We focused on smallholder farming systems and western Kenya region has been taken as a case study. Previous research conducted in western Kenya mainly focused on within farm variability (Tittonell *et al.*, 2013) and, spatial and temporal variability in maize response (S. Njoroge *et al.*, 2017). In this region, blanket fertilizer recommendations are still being used.

2. Materials and methods

2.1. Site description

The study area is a heterogeneous smallholder landscape in western Kenya (0°26′ - 0°18′ northern latitude; 33°58′ - 34°33′ eastern longitude) delimited by the administrative boundaries of Siaya and Kakamega counties (Fig. 1). The area is characterized by its (sub-) humid conditions and classified as the Lower Midland (LM_1) (Jaetzold *et al.*, 2007). FURP provided a blanket fertilizer recommendation for the LM_1 AEZ of 60 kg N ha^{-1} yr^{-1} and 30 kg P ha^{-1} yr^{-1} for monocrop maize (FURP, 1994). The FURP experiment site is located 2 km away outside the current study area but within the LM_1 AEZ (Fig. 1c).

The smallholder landscape is characterized by a distinct bimodal rainfall pattern and a mean annual temperature of 20°C. Long rains (March–June) have a mean precipitation of 1350 mm whereas short rains (September-December) have a mean precipitation of 850 mm (Jaetzold *et al.*, 2007). The mean potential evapotranspiration is estimated at 1287 mm per maize growing season (Ademba *et al.*, 2015). The altitudes vary in the gently undulating landscape (slopes <3 %) between 1400 and 1500 m above sea level. The main soil types (IUSS Working Group WRB, 2014) are presented in Figure 1b and include Rhodic Ferralsols (well drained, moderately to very deep, clay soils) and Ferralic Cambisols (well drained, moderately deep, loamy clay soils) on the hills, and Dystric Gleysol (poorly drained, shallow, sandy loam soils) in the plains (Waswa *et al.*, 2013). The fertility of the soils is limiting in N, P and K (Lijzenga, 1998).

Farming systems are subsistent with dominantly rain-fed crops and low fertilizer input. The mixed crop-livestock system includes maize (*Zea mays L.*) as the dominant staple crop, usually intercropped with common bean (*Phaseolus vulgaris L.*). Average maize yield levels achieved with the current local conditions using conventional farming practices range from 400 to 2000 kg ha^{-1} (Vanlauwe *et al.*, 2014) for the long rainy season. Other crops cultivated include bananas (*Musa paradisiaca L*.), sweet potatoes (*Ipomoea batata L*.) and groundnuts (*Arachis hypogaea L*). The smallholder farmers are supported by governmental agricultural extension services.

Fig. 1

2.2. Steps used to determine a relevant scale

This study included five consecutive steps to: (i) conduct a farm survey, (ii) determine variability in soil properties and crop performance indicators (CPIs), (iii) relate soil properties to CPIs and identify main soil factors associated with the variation in CPIs, (iv) characterize the scale of spatial variability of key soil factors in order to describe a relevant scale of variability in inherent soil fertility, and (v) evaluate if there is any correspondence in the scale of variability of key soil factors with the scale of variability in fertilizer response.

2.3 Step 1: Farm survey

A farm survey was carried out within the Land Degradation Sampling Framework (LDSF) (Vågen *et al*., 2010). The LDSF is a stratified hierarchical sampling design that captures variability at different scale levels: block, tiles, sub-tiles fields and plots across a given landscape (Fig. 2). First, a 100 km^2 block, chosen to typify a smallholder landscape, was allocated within the study area. The block was sub-divided into 16 (2.5 km^2) tiles and each tile further sub-divided into 10 (0.25 km^2) sub-tiles. A total of 8 tiles and 32 sub-tiles (four from each tile) were randomly selected, within which three points were randomly located. Near each point maize fields for sampling were sought during the short and long rains maize cropping seasons of 2012 and 2013. Different points were drawn for each of the two seasons. Unfertilized, well-managed (*e.g.*, free of weeds, pest, not affected with drought and diseases), mono-crop maize fields were selected for sampling that had a maize crop at the ear-leaf growth stage (silking stage *i.e.*, 70 - 75 days after plant emergence) at the time of sampling. The ear-leaf growth stage is considered optimum for diagnosis of nutrient constraints in maize. Per sub-tile two different locations were selected for sampling in every two of the consecutive seasons. At the locations, a maize field according to the above criteria was searched. A third point was selected for cases where no appropriate maize field was found near one of the two selected points. If the two selected points met the criteria, the third point was not sampled. During this farm survey, only unfertilized maize fields were sampled where the yield is taken to reflect the inherent soil fertility.

The locations were found with a GPS and maize fields were identified nearby. In the selected maize fields, a Y frame layout was placed to locate four plots measuring 2.5 m^2 (Fig 2 d). The central plot was located first, by measuring 20 m from the main boundary, towards the centre of the maize field. The main boundary is defined as the boundary located from the direction of smallholder farmers' homestead, towards the maize field. Subsequently, three plots were located 12.2 m from the central plot and distributed uniformly around it. In a first visit, the exact coordinates were recorded, soils were sampled and plant density (count number of maize plants), and plant biovolume, as a proxy of plant biomass determined. In a second visit, just before harvest by the farmer, the grain yield (yield) was measured. The biovolume (BV) was estimated using the basal diameter (BD) and height (H) of the maize plant following Chomba *et al.*, (2013), Equation 1:

$$BV(cm^3) = H(cm) \times \left(\frac{BD(cm)}{2}\right)^2 \pi \qquad 1$$

The BD was measured in duplicate, 2 cm above the soil surface for all maize plants in the plot. Mean biovolume was estimated using the BD and H measurements of all plants in the plot. Yield was measured from dry maize that was hand-harvested and the kernels removed and weighed (kg) at plant maturity (between 50-60 days after the silking stage). Yield and biovolume are the CPIs and were used as proxies of crop response, which reflect variability in the inherent soil fertility across the maize fields. In the rest of this article, the term CPIs is used to refer to both yield or biovolume.

To characterize soil properties, composite soil samples were taken per plot. Using a zig zag pattern, six topsoil (0-20 cm) samples were taken with an Edelman soil auger (600 cm^3) within the 2.5 m^2 plots to have a representative composite soil sample. Subsamples from the composite soil samples, obtained using coning and quartering, were analysed at Crop Nutrition Laboratory Services and World Agroforestry Centre laboratories. These samples were air-dried, thoroughly mixed and ground to pass a 2 mm sieve prior to the analysis. Soil pH was measured with a pH meter with a 1:2 soil/water suspension (Okalebo *et al.*, 2002). SOC and total N were analysed by dry combustion using a C/N analyser (Wright and Bailey, 2001), with an acidification pre-treatment to remove carbonates for SOC detemination. Extractable Ca, Mg, K, Na, and P were determined using the Mehlich-3 extract (Mehlich, 1984) and an inductive coupled plasma optical emission spectrometer (Sikora *et al.*, 2005).

Fig. 2

2.4 Step 2: Variability in soil properties and crop performance

To assess the variability (distribution) in soil properties and CPIs, descriptive statistics (means, median, coefficient of variation (CV), minimum and maximum) were calculated. Density plots were used to check whether soil properties and CPIs followed a normal distribution as a requirement for the subsequent steps of multivariate statistical analysis. Variables with a left-skewed distribution were log-transformed for the frequency to achieve near normality as a requirement for regressions and geostatistical analysis. Relationships among soil properties, and between soil properties and CPIs were examined by pair-wise correlation analysis to derive the Pearson correlation coefficients (*r*). Normally, problems with multi-collinearity occur when highly correlated variables are included in a regression model (Wold *et al.*, 1984). Therefore, to assess the degree of variable interactions amongst soil properties and CPIs, highly correlated ($r >$ 0.80) variables were identified.

2.5 Step 3: Key soil drivers of crop performance indicators

Key soil properties that can be attributed to the variation in CPIs were identified by Stepwise Multiple Linear Regression (SMLR) modelling (Geladi *et al.*, 1999). The regression method was used to analyse the linear relationship between single dependent variables (CPIs) with the independent variables (soil properties) based on Equation 2.

$$y = a + \sum_{i=1}^{n} b_i \times x_i \pm \varepsilon \qquad 2$$

where "y" is the CPIs, "x_i" are the soil properties, "n" is the number of soil properties, "a" is the intercept, "b_i" are the regression coefficients and "ε" is the standard error of the estimate. The SMLR analysis was used because of; (i) the few numbers of independent variables included in the analysis, - eight soil properties (pH, SOC, TN, P, K, Ca, Mg, Na) and, (ii) to avoid the problem of overfitting by adding or deleting variable in SMLR analysis (Guan *et al.*, 2013). The data were also evaluated for outliers which were expunged from the analysis to minimize this problem of overfitting. The relative importance of the predictors in the regression model were calculated and used to discern the key soil factors. The significance probability and coefficient of determination r^2 statistics were used as basis of evaluation. Hereafter, in the rest of the text, the r^2 refer to the conventional coefficient of determination - the proportion of variance explained by soil factors. We used a bootstrap re-sampling strategy to assess the strength of evidence, that the identified soil predictors were truly independent and reproducible. Hence, the mean confidence intervals for each soil predictors were estimated using 1000 iterations. The relative importance was calculated using the "*relaimpo*" R package (Grömping, 2006)

To further discern which soil properties were key, the soil predictors were gradually eliminated using a stepwise, backward elimination and forward selection criterion, to build the best regression equations describing CPIs as a function of the soil properties. The predictors with the lowest significant contribution to the regression model were eliminated and then tested until the remaining ones had a significant contribution. The key predictors of CPIs were then identified in the final regression equation. We calculated the Akaike's Information Criteria (AIC) value, and used it to evaluate whether the identified soil predictors were similar to those obtained using the aforementioned relative importance (r^2) statistic. The magnitude of AIC values formed the basis for interpretation, where the regression model with the lowest AIC was considered as the best, while the soil predictors with the highest AIC value were taken to be the key soil property. All the statistical regression modelling was done using the "*lme4*" R- package (Bates *et al.*, 2015).

2.6. Step 4: Scale of variability

Discrete map units, such as AEZ, soil map units, and administrative boundaries, form a logical basis for the development of blanket fertilizer recommendations (Smaling and Van De Weg, 1990). The study area fell within a single AEZ. However, different soil map units and administrative units were identified. These units are only useful for the refinement of the fertilizer use recommendations if they describe the variation in the key soil properties that influence the CPIs. Soil types are expected to describe the variation in soil properties and could form a logical basis for fertilizer recommendations. Administrative units are instrumental to agricultural extension officers for logistical purposes of disseminating fertilizer recommendations. However, the administrative units may include considerable variation in soil properties, making them less useful for fertilizer recommendations. The hierarchy of scales across the LDSF sampling framework (Fig. 2) was also considered since nutrient variability occurs at different scale levels across the smallholder landscape (Tittonell *et al.*, 2013).

An ANOVA mixed-effect linear models were conducted to analyse the variation across the soil map units, administrative units, and different scales of the LDSF sampling frame using the identified key soil predictors and CPIs. The "*nlme*" *R*-package was used to conduct the unbalanced ANOVA, where the plots (2.5 m^2) were fitted as random effects (Pinheiro *et al.*, 2019). The administrative units, soil types, tiles, sub-tiles and fields were considered as the fixed effects and provided a measure of explained variability (EV) by each of the mapping units. One mapping unit was modelled at a time with the response variable being the identified key soil properties or the CPIs.

To evaluate the proportion of EV between the mapping units, two pseudo R^2 summary statistics for mixed-effects models were estimated as described by Nakagawa and Schielzeth, (2013). The "marginal" (R_m^2), which considers the variance of the fixed effects, and "conditional" (R_c^2) that takes the variance of both the fixed and random effects into account (Nakagawa *et al.*, 2017). These statistics were computed following Equations 3 and 4, respectively.

$$R_m^2 = \frac{var_f}{var_f + var_r + var_e} \qquad 3$$

$$R_c^2 = \frac{var_f + var_r}{var_f + var_r + var_e} \qquad 4$$

where var_f is the variance of the fixed effects, var_r is the variance of the random effect and var_e is the variance of the model residuals. High values of R^2 indicates the mapping units (*e.g.*, soil types/maize fields) may be appropriate for a blanket fertilizer recommendation.

Often discrete mapping units do not properly describe the relevant spatial variability, specifically the local variation on smallholder farms (*e.g.* where the administrative boundaries are used). An alternative approach to describe the spatial variation in the key soil properties is to carry out a geostatistical analysis to determine the spatial dependency of the key soil factors and CPIs. We derived the semi-variograms for key soil factors and CPIs following Kerry *et al*., (2010) using the "*gstat*" *R*-package (Pebesma, 2004). The range was interpreted as the basis to define the relevant scale of variability following Kerry and Oliver, (2004). This interpretation depends on the strength of spatial dependencies that was determined by the nugget/sill ratio following Cambardella *et al*., (1994) with ratios < 0.25 depict strong, 0.25 – 0.75 moderate and > 0.75 weak spatial dependencies. Weak spatial dependence implied that there was considerable short distance variation and no logical patterns (Costa *et al*., 2015).

2.7 Step 5: Validation

In steps 1-4, we made use of a relatively quick farm survey to analyse the spatial variability in soils and crop performance. However, to properly derive fertilizer use recommendations, it is necessary to look at fertilizer response. However, under normal conditions it would be very resource intensive to carry out a large number of fertilizer response trials. Therefore, we hypothesized that the scale of variability of key soil factors associated with CPIs corresponds to the scale of variability in fertilizer response across smallholder farms. In the study area trials were carried out n forty two, 100 m^2 plots by the African Soil Information Service (http://afsis-dt.ciat.cgiar.org) and the International Institute of Plant Nutrition (IPNI) (Huising e*t al*., 2013; Zingore *et al*., 2014). This allowed us to test the above hypothesis. Data of fertilized maize trials were obtained in the short and long rain season of 2010 and 2013. These trials consisted of N, P and K fertilizer treatments. Maize yields of the fertilized and control plots were used to calculate the Fertilizer Response (FR), computed as a response ratio following Hedges *et al*., (1999):

$$ln\,\mathrm{FR} = ln\left(\frac{y_c}{y_t}\right),$$

where y_c is yield from the control plot and y_t is the yield from the treatment plots all reported in Mg ha^{-1}. The FR was transformed into the *ln* FR to achieve near normality. Geographical coordinates, corresponding to the centroid of each fertilized plot were used to determine the spatial dependency of *ln*FR as described in section 2.6. The spatial dependency in FR was compared to the spatial dependency in soil properties and CPI's.

3 Results and discussion

3.1 Step 1: Farm survey

A total of 64 maize fields on smallholder farms were sampled within 32 sub-tiles that were randomly distributed within 8 tiles across the 100 km^2 block. An average of 7 maize fields within each tile was sampled. Out of 256 plots sampled, 203 had complete observations on soil properties, yield and biovolume. Yield was not observed in 20% of the fields as farmers harvested the field prior to the planned sampling date. Seven observation were identified as outliers. Incomplete records and extreme values were excluded from the statistical analysis. The total number of maize plants at silking stage per plot ranged from 7 to 25. At harvest, a 10% reduction in total number of maize plants was observed for all fields sampled during biovolume estimates.

3.1 Step 2: Variability in soil properties and crop performance indicators

Table 1 presents the descriptive statistics for soil properties and CPIs. Soil properties in the topsoil indicate considerable variation in soil fertility. SOC concentrations ranged from 0.56 to 5.23%. The highest SOC concentrations were found in fields that had been recently converted to maize cultivation and those that displayed intensive soil management (14% of the observations). Low SOC concentrations were observed in maize fields that were intensively cultivated. Soil pH varied from slight acidity to near neutrality (4.8 to 7.4) and within the optimum range for maize growth. Mehlich-3 extractable P was below the critical concentration of 15 mg kg^{-1} in 55% of the sampled plots. Yield ranged widely from 0.8 Mg ha^{-1} to 11.8 Mg ha^{-1}.

Coefficients of variation indicated different degrees of variation of the soil properties and CPIs (Table 1). Mehlich-3 extractable P and K were highly variable with CVs of 74% and, 89% respectively; Extractable Ca and Mg had CVs of 60% and 52%, respectively, while SOC and total N has lower CVs of 32% and, 26% respectively. Yield and biovolume exhibited a moderate variation as indicated by their CVs of 57% and, 43% respectively. We used the coefficient of variation to assess variability since it allowed comparison among variables with different units of measurement (soil properties and CPIs). However, the CV statistics could not allow us to explicitly evaluate the spatial variation in soil properties (Haileslassie *et al.*, 2005).

Table 1

Density plots for Mehlich-3 extractable P, K, Ca, Mg and Na displayed a negatively skewed distribution, indicating a high prevalence of low values in the dataset, which signifies low nutrient levels in the study area. Results of the natural log-transformed soil properties, prior to correlation analysis are presented in Supplementary Material 1 (S1, Fig. 1). Soil pH, total N, yield and biovolume displayed a near normal distribution.

Significant correlations between soil properties and CPIs was evident, as shown by *r* values (Table 2). SOC was positively correlated with yield (r = 0.55, p < 0.001) and biovolume (r=0.88, p < 0.0001). Relationships between P and yield (r = 0.01, p = 0.02), Na and yield (r = 0.01, p = 0.101) were weak and insignificant, and so was the relation between P and biovolume (Table 2; see Supplementary Material 1, S1 Fig. 2).

Results of pairwise correlation also revealed a high correlation among soil properties (r > 0.8). SOC and total N were highly correlated (r= 0.95, p < 0.001), and so were Ca and Mg (r= 0.89, p < 0.001). Correlation coefficients between pH, SOC, P, Ca, K and Na were relatively low. Having known the association between soil properties, all the eight soil properties were included in the regression analysis. The next step of analysis was to identify key factors which influence underlying variation in CPIs for the study area.

Table 2

3.4 Step 3: Key soil drivers of crop performance.

Regression results indicate soil predictors explained 32% and 79% of the variability in maize yield and biovolume, respectively (Table 3). SOC was the main factor that significantly ($p < 0.001$) contributed to the variability in CPIs. The explained variance by the soil predictors indicate that each individual soil property played a role in influencing the underlying variation. But the contribution of pH, Total N, P, Ca, Mg, K and Na was not statistically significant ($p > 0.001$). The problem of multicollinearity could have had an influence on the performance of the model, since SOC and Total N, as well as Mg and Ca were highly correlated (see Supplementary Material 1, S1 Fig. 2). To test the aforementioned influence, total N and Mg were removed from the model. This reduced the explained variance to 31% and 78% for yield and biovolume, respectively. Although the intercept become significant ($p < 0.1$), meaning the model accuracy was not affected. Hence, the result indicates no significant influence of multicollinearity when all eight soil properties were included in the regression models. Multicollinearity creates high coefficient estimators that inflates variances and may lead to selecting the wrong soil predictors (Kroll and Song, 2013). The problem is magnified when the samples size is small contrary to ours, which had 196 observations (Kroll and Song, 2013). Thus, the multiple linear regression models predicted variability of CPIs fairly well as shown by the explained variance for the study area.

Table 3

The relative importance results for the eight predictor of yield and biovolume are shown Fig. 3. The predictor with highest r^2 was SOC with values of 0.41 for yield and 0.43 for biovolume. SOC was identified as the key factor that influence variation in CPIs for the study area. Bootstrapping stimulations, employing different statistical method, confirmed SOC as key factor (See Supplementary Material 1, S1 Fig. 3). The lowest observed r^2 values were 0.12 (pH) for yield and 0.12 (*ln*Na) for biovolume. The negative influence of low acidity and high sodicity explain why pH and sodium were the least important soil predictors (Mbakaya *et al.*, 2008). The regression models represent part of variation in CPIs as explained by the soil properties, and was useful for differentiating the contribution of each soil factor. We used SMLR modelling as a strategy of reducing the number of candidates to be considered for evaluation of spatial structure. This also simplifies the proposed approach. Our results suggest that SOC was the key soil factor that influenced spatial variation in CPIs.

Fig. 3

To further discern which of these soil properties are key soil factors, AIC values were evaluated from the stepwise regression models (Fig. 4). The best regression model included SOC and Na as the main soil predictors for CPIs. SOC was the only significant ($p < 0.0001$) predictor in the models with the highest AIC value of 257 for yield and 2113 for biovolume (Fig. 4). Even though Na was included in this regression equation, its contribution was not significant ($p = 0.7635$) and the AIC values for CPIs were the lowest (187). Thus, the results confirmed that indeed SOC was key soil factors that influence variation in crop performance of the study area.

Fig. 4

3.5 Step 4: Spatial variability of key soil drivers and crop performance indicators

The ANOVA models showed that the large soil units (soil types and administrative boundaries) describe less than 10% of the variation in the SOC (Fig. 5). The mixed-effect models resulted in a low marginal R^2 (Fig 5. a c) but high conditional R^2 (Fig 5 b d). This meant that the fixed effects (administrative boundaries, soil types, LDSF scales (tile, sub-tile and field)) explain low variability (< 5 %) in SOC. Most of the spatial variability in SOC was attributed to differences in fields, and between sampling the plots, indicated by high marginal R^2 values (Fig 5a c). This implied that most of the local variability was capture at field level (within variation), which make them good basis for the development of a fertilizer use recommendation. A similar trend was observed for the CPIs (Fig. 6). The high conditional R^2 is attributed to the inclusion of variance for both the larger mapping units (fixed effects) and plots (the random effects). Although the mean SOC, yield and biovolume were significantly different between the three soil types, there is considerable variation within these soil units. Here, it also becomes apparent that the smaller mapping units (fields) are describing considerable variation in SOC, yield and biovolume compared to all the other stratifications of the landscape that were applied, even despite the fact that they are just randomly located squared in the landscape.

Fig. 5

Fig. 6

Table 4 shows the semi-variogram parameters with the best-fitted model for the SOC, FR and CPIs. SOC and CPIs showed moderate to strong spatial dependencies based on the definition of Cambardella *et al.* (1994). However, semi-variograms showed considerable short distance variability, confirming the results in the literature that these systems present considerable short distance variability. Although SOC showed considerable spatial dependency, the patterns were found to occur at relatively short distances with a range of 523 m. Despite the relations between the soil properties and the CPIs, the short distance spatial dependency shown for the soil properties is not found for the CPIs. The CPIs show a stronger spatial dependency and also a longer range. The value of the range can be considered as the scale of distance beyond which SOC do not show any spatial correlation. This implied that the range could be interpreted further to provide guidance on optimum scale for developing digital nutrient maps for this study area.

The results of the geostatistical analysis are in line with the analysis of variance. The short distance variability found for SOC explain that the very general soil units and large tiles that both cover areas of > 5 km^2 do not describe the variation. Smaller areas like the sub-tiles and the administrative units are roughly the size of the range of the semi-variograms and describe more variation in SOC as shown by the analysis of variance. The results imply that the sampling distance for representative soil test results should be smaller for SOC (523m) in this area.

The results confirm that blanket fertilizer use recommendations on the basis of coarse digital soil map are not likely to be efficient in increasing of food production. The development of fertilizer use recommendations will require intensive sampling to describe the variation in soil conditions. Following Kerry and Oliver (2004), the results can be re-interpreted towards optimal sampling densities. The optimum sampling distance should be less than half the range of a fitted semi-variogram model. Given the relatively short range of 523 m for SOC, it is necessary to sample at distances of less than 250 m. Although this would describe the large trends, it should be recognized the results would still not be very effective due to short distance variability as indicated by the nugget. Thus, local variability within smallholder farms may be captured at 250 m resolution. Given the 250 m resolution, two options for improving fertilizer recommendations can be explored for this area; (i) predict soil test values through interpolation, using suitable environmental covariates *e.g.* detailed satellite imagery (digital elevation models) through regression kriging, or (ii) farmers to rely on soil testing on their fields.

The two CPIs showed different levels of spatial variation with different spatial dependencies. This is not surprising given the relatively low correlation coefficient between them (r^2 = 0.56).

Table 4

3.6 Step 5: Validation

The results showed that there is considerable soil variation at local scale (< 523 m). The best semi-variogram model fitted for FR was spherical, which corresponded to the least root mean square error. The nugget/sill ratio suggest FR exhibited moderate of spatial dependencies across fertilized plots for the study area (Table 4). We observed a high nugget effect (Table 4) suggesting that in fertilizer response there was small-scale variation among fertilized maize plots. Moderate spatial dependencies implied that we could interpret the range. The range of FR was 426 m and of a similar order of magnitude as that of SOC (523 m) (Table 5). Hence our hypothesis, that the scale of fertilizer response corresponds to the scale of variability of key soil properties is not rejected.

4. General discussion

This study aimed at establishing an ex-ante approach to determine a rough scale that reflects local spatial variability in smallholder farming landscapes. We envisaged that the study will inform decisions for refining fertilizer use recommendations. This would be accomplished by providing guidance on a rough scale for using digital soil maps or help inform sampling distance for soil testing in heterogenous smallholder fields. The knowledge will aid farmers and policy makers resolve problems related to in-situ soil testing, by making better nutrient management decisions. A generalized workflow is provided in Fig. 7. Empirical rules were derived, to describe the observed variability of key soil factors (SOC) and CPIs, which are then used to determine a directional flow of decisions that would help refine the recommendation. Although, we consider this study site as a good representation of smallholder farming landscapes, our approach is only applicable to this site and other regions (rain-fed) with similar topography, climate and soils conditions.

Fig. 7

The relevant scale at which we can properly describe inherent soil spatial variation has rarely been explicitly considered when developing digital soil maps for heterogeneous smallholder farming systems. Variability in soil properties and CPIs provided a framework for estimating a rough scale for developing digital nutrient maps. Our results provide evidence of existence of variability as indicated by soil properties, that displayed high variation, as shown by high CV values (Table 1). Anthropic influence affects spatial structure of soil properties in maize fields. Variability in soil properties can be attributed to natural intrinsic variation, parent material (Deckers, 2002) and difference in management across maize fields (Zingore *et al.*, 2007). The high variability in yield were within ranges of maize yield reported by Kihara *et al.*, (2016). Variability of maize yield has also been reported in other studies conducted in western Kenya at landscape level (Burke and Lobell, 2017; Tittonell *et al.*, 2013). Studies have shown the impact of high soil variability on nutrient requirement for maize crop in smallholder farms of Nigeria

(Shehu *et al.*, 2018).

High variation in soil properties, has consequently led to variable fertilizer use efficiency and fertilizer response, within these smallholder farms (R. Njoroge *et al.*, 2017; Tittonell *et al.*, 2007). Maps displaying spatial patterns of soil properties may capture the variability at specific locations and provide information for the local inherent soil variability (Antwi *et al.*, 2016). The use of auxiliary information in digital soil maps would further allow integration of covariate information such as digital elevation (as raster maps) in creation of the digital soil maps. As a result, improve on the accuracy of predicting soil test values for nutrient management. This study provides a sequential framework that can aid in capturing local variation in soil properties (Fig. 7)

Our study showed robust relationships between CPIs and soil properties, which were evaluated further (Table 2). Pair-wise correlations results between soil properties and CPIs were statistically significant and provide evidence of the existence of a relation between inherent soil fertility and CPIs (Table 2). The most striking correlations were between SOC and biovolume ($r = 0.88$, $p < 0.0001$) and between SOC and yield ($r = 0.55$, $p < 0.001$). This confirms that SOC is an important soil factor, and proper management of SOC would result to high CPIs in the region. These findings also agree with those reported by Chomba *et al.* (2013). It is difficult to give a reason why there were good correlation between SOC with biovolume and not yield (Table 2). A similar correlation between SOC and maize yield ($r = 0.59$, $p < 0.001$) has been observed in other studies conducted in Uganda on Ferrosols (Musinguzi *et al.*, 2016). Correlations between soil properties and CPIs, agree with those reported by (Mtangadura *et al.*, 2017) who found positive relations between maize yield and Ca in Zimbabwe.

The SMLR revealed SOC as key factors associated with the variation of crop performance for the study area (Fig. 5 and 6). This can be explained by the fact that SOC has a dominant influence on N supply, nutrient retention and sulphur supply, soil structure and soil responsiveness to fertilizer application (Lal, 2016; Six *et al.*, 2002). These results are consistent with other studies that reported SOC as key soil factor that influence maize yield as well as fertilizer response (Musinguzi *et al.*, 2016). The knowledge of the spatial variability of SOC may be important for understanding the variance structure of CPIs and how they can be related. The SMLR statistical methodology implemented in analysing soil data has wider applicability and can be applied to other similar sites and crops. Climatic and management factors that influence crop performance (Tittonell *et al.*, 2008; Waithaka *et al.*, 2007) and could also be included, but would require a larger sample size to increase confidence level of the results (Maas and Hox, 2005). In this case when we ran a regression model, and reduced or including additional soil factors did not explain more of the variation in CPIs than SOC and Na alone.

We evaluated the variation of key soil factors and CPIs using ANOVA and their spatial structure. ANOVA model results displayed significant (p < 0.001) variation between administration boundaries, tile and sub-tiles. Low EV for SOC suggests that not much variation (< 25%) was captured by the large discrete mapping units (administration boundaries, tile and sub-tiles) for this region (Fig. 5a). Broad discrete mapping unit such as administrative boundaries are likely to be inappropriate for delineating fertilizer recommendations, as indicated by the fixed -effects, and marginal R^2 values (Fig 5b & 6b d). The EV was considerable (> 25%) for the small spatial units (*e.g.*, field). This was also evident based on the high conditional R^2 values (> 50 %), which accounted for both the random and fixed effect in the ANOVA analysis. High EV for maize fields implied that fertilizer recommendations should be provided at field level for the study area. This may require each smallholder farmer to conduct soil testing of their fields.

Variation of key soil factors was described by very small spatial units and confirmed by low nugget values obtained from the semi-variogram models, indicating short distance variability (Table 4). This further confirmed that fertilizer use recommendations will need to rely on digital soil maps and/or local sampling at a defined distance. Short distance variability has been reported by Diarisso *et al.*, (2015) in the small villages of west Africa. However, short distance variability may limit application of digital soil mapping. For such a scenario, soil testing would be an alternative. However, researchers have argued that, the use of soil testing can be more effective, when it is combined with plant tissue analysis (Webb *et al.*, 2011) or when cheaper and rapid soil characterisation methods such as infrared spectroscopy (IR) are employed, especially as IR predicts SOC well (Shepherd *et al.*, 2015).

Our ex-ante approach demonstrates local spatial variability can be captured for the heterogenous smallholder landscape. While estimating spatial variability of soil properties and crop response, the unit of measurement is an important consideration, as it influences the range (Bhati 2005). Findings reported in other studies indicate plots size did not affect the crop yields (Bhati 2005). However, it recommended to use smaller sampling units, as they could increase the precision of estimating variance. Therefore, the small plots ($2.5m^2$) captured spatial variation on the maize fields, since small units of measurement lead to spatial variance close to the true value, while large units may introduce biases (Western and Blöschl, 1999).

Optimization of a relevant scale has been a major bottleneck for nutrient management in smallholder farms (Vasu *et al.*, 2017). The spatial structure of SOC with a strong relation with the CPIs and moderate dependencies was a proper basis. Moderate spatial dependencies for SOC have previously been reported elsewhere in western Kenya (Okeyo *et al.*, 2009). Occurrence of moderated spatial dependencies can be explained by influence of extrinsic management factors such as ploughing and other local management practices that weaken spatial dependencies after long history of cultivation (Mzuku *et al.*, 2005). For this study area, we propose 250 m as optimum resolution for digital soil maps, given the effective range of 543 m for SOC. This distance can serve as threshold scale below which maps would capture the local growing conditions of the study area. Other studies have proposed a similar distance of 323 m for rain-fed conditions (Vasu *et al.*, 2017). The reliability of these digital nutrient maps will also depend on the sampling protocol and accuracy of the semi-variogram model (Liu *et al.*, 2014). However, at 250 m sampling distance may have impact on the cost of soil analysis.

We assumed that crop response and fertilizer response will display similar spatial patterns. Based on this assumption, a comparison was made between ranges of SOC and FR. The range of SOC (523 m) was similar to that of FR (423 m) (Table 4). Even though the ranges were not exact in their magnitude, our hypothesis was accepted. The minor discrepancies in terms of lack of exact correspondence of range between SOC and fertilizer response for trial data can be attributed the difference in sampling density between the two approaches. The sampling density for farm survey across the landscape was higher (2.3 samples per 100 km^2) compared to that of fertilizer trials was (0.42 samples per 100 km^2). Many studies report that high density sampling is required for better results where soil pattern is complex due to the topography (Cobo *et al.*, 2010; Tesfahunegn *et al.*, 2011). Difference in factors such as weather, and germplasm can also explain the discrepancy, since data collection was conducted in the same study area, but during different seasons and, sampled different maize varieties among farms.

The findings of a correspondence between the range of FR to that of SOC provides useful insights for strategies aimed at refining fertilizer recommendation in western Kenya. This justifies the need for adjusting the fertilizer application rate based on the observed local spatial variation in these maize fields. In this region, fertilizer use for maize varies with the available fertilizer type in the local agro-dealers, soil properties and rainfall (Ichami *et al.*, 2018; N.Sanginga *et al.*, 2009). The current fertilizer application rates of 31 kg per ha have been projected to increase to 50 kg per ha (Bezu and Holden, 2014), and could be realized if appropriate fertilizer management strategies are put in place. As a result, fertilizer use efficiency could be improved from the small amounts applied by farmers. Therefore, DSM, provided at a relevant scale (fine resolution) would play a critical role as a nutrient management tool for this study.

5. Conclusions

This study demonstrates an ex-ante, approach for establishing a relevant scale for making fertiliser recommendations that captures spatial variation of soil properties and CPIs based on local conditions (Fig. 7). We conclude the following:

- SOC was the key soil factors that determined variability in unfertilized maize grain yield and plant biovolume of this region.
- Discrete mapping units based on soil classification units and administrative boundaries, may not be suitable for delineating fertilizer recommendations for smallholder farms in the study area.
- Only SOC show moderate spatial dependencies and was used for interpretation of a suitable scale that could provide the relevant spatial detail of maps for nutrient

management for this study area.

- Based on the spatial correlation distance of SOC, which displayed an effective range of 523 m, it implies that, within this distance, local variability within smallholder farms may be captured. However, based on previous research (Kerry and Oliver, 2004), we propose a resolution/distance of 273 m as a threshold scale for developing digital nutrient maps or sampling intervals for soil testing. This can be complemented by the already existing map of 250 m resolution digital soil properties available for SSA through Hengl *et al.*, (2015). This finding provide approximation of scale as a basis for guiding fertiliser recommendations and future efforts should be directed at improving its accuracy.

The results of this study provide a rough estimate of scale that can be used for digital soil mapping, that would capture local variation in smallholder farms. Alternatively, the soil sampling distances can be based on 250 m, which would be the appropriate distance for capturing local variation. This can be implemented in a refined soil-based fertilizer management strategy for rainfed smallholder systems where digital soil mapping and soil testing would be important techniques. Additionally, 250 m sampling distance (interval) would certainly result to high number of soil samples. Thus, proximal sensing techniques such as infrared spectroscopy for can be utilized alongside DSM to reduce the costs for soil analysis.

Acknowledgement

This study was funded in collaboration of the University of Nairobi, Wageningen University sandwich PhD and CIAT-Kenya and World Agroforestry Centre (ICRAF) fellowship programs. The unlimited cooperation and support extended by International Plant Nutrition Institute (IPNI) by Dr. Shamie Zingore in carrying out the field research work in western Kenya are gratefully acknowledged. We also acknowledge the support of the CGIAR research program on 'Water, Land and Ecosystems'.

References

Ademba, J.S., Kwach, J.K., Esilaba, A.O., Ngari, S.M., 2015. The Effects of Phosphate Fertilizers and Manure on Maize Yields in South Western Kenya. East African Agric. For. J. 81, 1–11. https://doi.org/10.1080/00128325.2015.1040640

AIC, 1981. Agricultural Information Centre (AIC) Major Crops Technical Handbook. .

Antwi, M., Duker, A.A., Fosu, M., Abaidoo, R.C., 2016. Geospatial approach to study the spatial distribution of major soil nutrients in the Northern region of Ghana. Cogent Geosci. 2. https://doi.org/10.1080/23312041.2016.1201906

Baruck, J., Nestroy, O., Sartori, G., Baize, D., Traidl, R., Vrščaj, B., Bräm, E., Gruber, F.E., Heinrich, K., Geitner, C., 2016. Soil classification and mapping in the Alps: The current state and future challenges. Geoderma. https://doi.org/10.1016/j.geoderma.2015.08.005

Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. https://doi.org/10.18637/jss.v067.i01

Bekunda, M., Sanginga, N., Woomer, P.L., 2010. Restoring soil fertility in sub-sahara Africa.

Bezu, S., Holden, S., 2014. Are rural youth in ethiopia abandoning agriculture? World Dev. https://doi.org/10.1016/j.worlddev.2014.06.013

Burke, M., Lobell, D.B., 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1616919114

Cambardella, C.A., Moorman, T.B., Parkin, T.B., Karlen, D.L., Novak, J.M., Turco, R.F., Konopka, A.E., 1994. Field-Scale Variability of Soil Properties in Central Iowa Soils. Soil Sci. Soc. Am. J. 58, 1501–1511. https://doi.org/10.2136/sssaj1994.03615995005800050033x

Chomba, S.K., Okalebo, J.R., Imo, M., Mutuo, P.K., Walsh, M.G., 2013. Predicting maize response to fertilizer application using growth curves in western Kenya . Glob. J. Arab. Crop Prod. 1, 78–85.

Cobo, J.G., Dercon, G., Yekeye, T., Chapungu, L., Kadzere, C., Murwira, A., Delve, R., Cadisch, G., 2010. Integration of mid-infrared spectroscopy and geostatistics in the assessment of soil spatial variability at landscape level. Geoderma 158, 398–411.

Costa, C., Papatheodorou, E.M., Monokrousos, N., Stamou, G.P., 2015. Spatial Variability of Soil Organic C, Inorganic N and Extractable P in a Mediterranean Grazed Area. L. Degrad. Dev. https://doi.org/10.1002/ldr.2188

Deckers, J., 2002. A systems approach to target balanced nutrient management in soilscapes of sub-Saharan Africa. Integr. plant Nutr. Manag. Sub-Saharan África. From concepts to Pract. CAB Int. Wallingford, Oxon, UK 47–62.

Diarisso, T., Corbeels, M., Andrieu, N., Djamen, P., Douzet, J.-M., Tittonell, P., 2016. Soil variability and crop yield gaps in two village landscapes of Burkina Faso. Nutr. Cycl. Agroecosystems 105, 199–216. https://doi.org/10.1007/s10705-015-9705-6

Diarisso, T., Corbeels, M., Andrieu, N., Djamen, P., Douzet, J.-M., Tittonell, P., 2015. Soil variability and crop yield gaps in two village landscapes of Burkina Faso. Nutr. Cycl. Agroecosystems 1–18. https://doi.org/10.1007/s10705-015-9705-6

FURP, 1994. The Fertilizer Use Recommendation Project , Kenya Agricultural Research Institute, National Agricultural Research Laboratories, Nairobi. Final Report. Annex I Fertil. trial Doc. (FERDOC). Minist. Agric. Nairobi.

Geladi, P., Hadjiiski, L., Hopke, P., 1999. Multiple regression for environmental data: Nonlinearities and prediction bias, in: Chemometrics and Intelligent Laboratory Systems. https://doi.org/10.1016/S0169-7439(98)00204-4

Geurts, P.M.H., Van den Berg, M., 1998. A simple agro-ecological zonation for fertilizer recommendations in Mozambique. Soil Use Manag. 14, 136–141. https://doi.org/10.1111/j.1475-2743.1998.tb00138.x

Giller, K.E., Rowe, E.C., de Ridder, N., van Keulen, H., 2006. Resource use dynamics and interactions in the tropics: Scaling up in space and time. Agric. Syst. 88, 8–27. https://doi.org/http://dx.doi.org/10.1016/j.agsy.2005.06.016

Grömping, U., 2006. Relative importance for linear regression in R: The package relaimpo. J. Stat. Softw. https://doi.org/10.18637/jss.v017.i01

Guan, Y., Wei, J., Zhang, D., Zu, M., Zhang, L., 2013. To identify the important soil properties affecting dinoseb adsorption with statistical analysis. Sci. World J. https://doi.org/10.1155/2013/362854

Haileslassie, A., Priess, J., Veldkamp, E., Teketay, D., Lesschen, J.P., 2005. Assessment of soil nutrient depletion and its spatial variability on smallholders' mixed farming systems in Ethiopia using partial versus full nutrient balances. Agric. Ecosyst. Environ. 108, 1–16. https://doi.org/10.1016/j.agee.2004.12.010

Hedges, L. V., Gurevitch, J., Curtis, P.S., 1999. The meta-analysis of response ratios in experimental ecology. Ecology. https://doi.org/10.1890/0012-9658(1999)080[1150:TMAORR]2.0.CO;2

Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. PLoS One 10, e0125814. https://doi.org/10.1371/journal.pone.0125814

Huising, E.J., Zingore, S., Kihara, J., Nziguheba, G., Palm, C.A., 2011. Diagnostic trials: a field guide.

Ichami, S.M., Shepherd, K.D., Sila, A.M., Stoorvogel, J.J., Hoffland, E., 2018. Fertilizer response and nitrogen use efficiency in African smallholder maize farms. Nutr. Cycl. Agroecosystems. https://doi.org/10.1007/s10705-018-9958-y

Jaetzold, Kutsch, 1982. Agro-ecological zones of the tropics with a sample from Kenya. Tropenlandwirtschaft 83: 15–34. Tropenlandwirtschaft 83, 15–34.

Janssen, B.H.H., Guiking, F.C.T.C.T., van der Eijk, D., Smaling, E.M.A.M.A., Wolf, J., van Reuler, H., 1990. A system for quantitative evaluation of the fertility of tropical soils (QUEFTS). Geoderma 46, 299–318. https://doi.org/10.1016/0016-7061(90)90021-Z

Kerry, R., Oliver, M.A., 2004. Average variograms to guide soil sampling. Int. J. Appl. Earth Obs. Geoinf. 5, 307–325.

Kerry, R., Oliver, M.A., Frogbrook, Z.L., 2010. Geostatistical Applications for Precision Agriculture. Precis. Agric. 305–312. https://doi.org/10.1007/978-90-481-9133-8

Kibblewhite, M.G., Ritz, K., Swift, M.J., 2008. Soil health in agricultural systems. Philos. Trans. R. Soc. B Biol. Sci. 363, 685 LP – 701.

Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., Njoroge, S., Palm, C., Huising, J., 2016. Understanding variability in crop response to fertilizer and amendments in sub-Saharan Africa. Agric. Ecosyst. Environ. 229, 1–12. https://doi.org/10.1016/j.agee.2016.05.012

Kisaka, M.O., Mucheru-Muna, M., Ngetich, F.K., Mugwe, J.N., Mugendi, D.N., Mairura, F., Muriuki, J., 2016. Using apsim-model as a decision-support-tool for long-term integrated-nitrogen-management and maize productivity under semi-arid conditions in Kenya. Exp. Agric. https://doi.org/10.1017/S0014479715000095

Kroll, C.N., Song, P., 2013. Impact of multicollinearity on small sample hydrologic regression models. Water Resour. Res. https://doi.org/10.1002/wrcr.20315

Lal, R., 2016. Soil health and carbon management. Food Energy Secur. https://doi.org/10.1002/fes3.96

Liu, Z., Zhou, W., Shen, J., He, P., Lei, Q., Liang, G., 2014. A simple assessment on spatial variability of rice yield and selected soil chemical properties of paddy fields in South China. Geoderma 235–236, 39–47. https://doi.org/10.1016/j.geoderma.2014.06.027

Maas, C.J.M., Hox, J.J., 2005. Sufficient sample sizes for multilevel modeling. Methodology 1, 86–92.

Matthews, R., Stephens, W., Hess, T., Middleton, T., 2002. Applications of Crop / Soil Simulation Models in Tropical Agricultural Systems. Adv. Agron. https://doi.org/http://dx.doi.org/10.1016/S0065-2113(02)76003-3

Mbakaya, D.S., Okalebo, J.R., Muricho, M., Lumasayi, S., 2008. EFFECTS OF LIMING AND INORGANIC FERTILIZERS ON MAIZE YIELD IN KAKAMEGA NORTH AND UGUNJA DISTRICTS , WESTERN KENYA. Sustain. Dev. 123–129.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/http://dx.doi.org/10.1016/S0016-7061(03)00223-4

Mehlich, A., 1984. Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. Commun. Soil Sci. Plant Anal. 15, 1409–1416.

Molefe, T., Thamaga-Chitja, J., Yobo, K.S., 2012. Qualitative evaluation of smallholder farmer decisions, support systems, knowledge and disease management tools. Indilinga African J. Indig. Knowl. Syst. 11, 160–178.

Mora-Vallejo, A., Claessens, L., Stoorvogel, J., Heuvelink, G.B.M., 2008. Small scale digital soil mapping in Southeastern Kenya. Catena 76, 44–53. https://doi.org/10.1016/j.catena.2008.09.008

Mowo, J.G., Mlingano, A.R.I., 1993. Review of fertilizer recommendations in Tanzania. National Soil Service, ARI Mlingano, Ministry of Agriculture.

Mtangadura, T.J., Mtambanengwe, F., Nezomba, H., Rurinda, J., Mapfumo, P., 2017. Why organic resources and current fertilizer formulations in Southern Africa cannot sustain maize productivity: Evidence from a long-term experiment in Zimbabwe. PLoS One. https://doi.org/10.1371/journal.pone.0182840

Musinguzi, P., Ebanyat, P., Tenywa, J.S., Basamba, T.A., Tenywa, M.M., Mubiru, D.N., 2016. Critical Soil Organic Carbon Range for Optimal Crop Response To Mineral Fertiliser Nitrogen on a Ferralsol. Exp. Agric. 52, 635–653. https://doi.org/10.1017/S0014479715000307

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., MacDonald, L., 2005. Spatial Variability of Measured Soil Properties across Site-Specific Management Zones. Soil Sci. Soc. Am. J. 69, 1572. https://doi.org/10.2136/sssaj2005.0062

N.Sanginga, P.W., Sanginga, N., Woomer, P.L., 2009. Integrated soil fertility management in Africa: principles, practices and developmental process. CIAT.

Nakagawa, S., Johnson, P.C.D., Schielzeth, H., 2017. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. J. R. Soc. Interface. https://doi.org/10.1098/rsif.2017.0213

Nakagawa, S., Schielzeth, H., 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol. Evol. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Njoroge, R., Otinga, A.N., Okalebo, J.R., Pepela, M., Merckx, R., 2017. Occurrence of poorly responsive soils in western Kenya and associated nutrient imbalances in maize (Zea mays L.). F. Crop. Res. 210, 162–174. https://doi.org/http://dx.doi.org/10.1016/j.fcr.2017.05.015

Njoroge, S., Schut, A.G.T., Giller, K.E., Zingore, S., 2017. Strong spatial-temporal patterns in maize yield response to nutrient additions in African smallholder farms. F. Crop. Res. 214, 321–330. https://doi.org/10.1016/j.fcr.2017.09.026

Okalebo, R.J., Gathua, W.K., Woomer, P.L., 2002. Laboratory Methods of Soil and Plant Analysis: A Working Manual. TSBF-CIAT and Sacred Africa, Nairobi, Kenya.

Okeyo, J.M., Shepherd, K.D., Wamicha, W., Shisanya, C., 2009. Spatial Variation in Soil Organic Carbon within Smallholder Farms in Western Kenya: A geospatial Approach.

Pampolino, M.F., Witt, C., Pasuquin, J.M., Johnston, A., Fisher, M.J., 2012. Development approach and evaluation of the Nutrient Expert software for nutrient management in cereal crops. Comput. Electron. Agric. 88, 103–110. https://doi.org/http://dx.doi.org/10.1016/j.compag.2012.07.007

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Comput. Geosci. 30, 683–691. https://doi.org/http://dx.doi.org/10.1016/j.cageo.2004.03.012

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2019. {nlme}: Linear and Nonlinear Mixed Effects Models.

Sanchez, P., Denning, G., Nziguheba, G., 2009. The African Green Revolution moves forward. Food Secur. 1, 37–44. https://doi.org/10.1007/s12571-009-0011-5

Shehu, B., Merckx, R., Jibrin, J., Kamara, A., Rurinda, J., 2018. Quantifying Variability in Maize Yield Response to Nutrient Applications in the Northern Nigerian Savanna. Agronomy. https://doi.org/10.3390/agronomy8020018

Shepherd, K.D., Shepherd, G., Walsh, M.G., 2015. Land health surveillance and response: A framework for evidence-informed land management. Agric. Syst. 132, 93–106. https://doi.org/10.1016/j.agsy.2014.09.002

Sikora, F.J., Howe, P.S., Hill, L.E., Reid, D.C., Harover, D.E., 2005. Comparison of colorimetric and ICP determination of phosphorus in Mehlich3 soil extracts. Commun. Soil Sci. Plant Anal. 36, 875–887.

Six, J., Callewaert, P., Lenders, S., De Gryze, S., Morris, S.J., Gregorich, E.G., Paul, E.A., Paustian, K., 2002. Measuring and Understanding Carbon Storage in Afforested Soils by Physical Fractionation. Soil Sci. Soc. Am. J. 66, 1981–1987. https://doi.org/10.2136/sssaj2002.1981

Smaling, E.M.A., Van De Weg, R.F., 1990. Using soil and climate maps and associated data sets to select sites for fertilizer trials in Kenya. Agric. Ecosyst. Environ. 31, 263–274. https://doi.org/http://dx.doi.org/10.1016/0167-8809(90)90225-3

Stoorvogel, J.J., Bakkenes, M., ten Brink, B.J.E., Temme, A.J.A.M., 2017. To What Extent Did We Change Our Soils? A Global Comparison of Natural and Current Conditions. L. Degrad. Dev. https://doi.org/10.1002/ldr.2721

Stoorvogel, J.J., Smaling, E.M.A., Janssen, B.H., 1993. Calculating soil nutrient balances in africa at different scales .1. supra-national scale. Fert Res 35, 227–235. https://doi.org/10.1007/BF00750641

Tesfahunegn, G.B., Tamene, L., Vlek, P.L.G., 2011. Catchment-scale spatial variability of soil properties and implications on site-specific soil management in northern Ethiopia. Soil Tillage Res. 117, 124–139. https://doi.org/10.1016/j.still.2011.09.005

Tittonell, P., Muriuki, A., Klapwijk, C.J., Shepherd, K.D., Coe, R., Vanlauwe, B., 2013. Soil Heterogeneity and Soil Fertility Gradients in Smallholder Farms of the East African Highlands. Soil Sci. Soc. Am. J. 77, 525–538. https://doi.org/10.2136/sssaj2012.0250

Tittonell, P., Vanlauwe, B., Corbeels, M., Giller, K.E., 2008. Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya. Plant Soil 313, 19–37. https://doi.org/10.1007/s11104-008-9676-3

Tittonell, P., Vanlauwe, B., de Ridder, N., Giller, K.E., 2007. Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: Soil fertility gradients or management intensity gradients? Agric. Syst. 94, 376–390. https://doi.org/10.1016/j.agsy.2006.10.012

Vågen, T.-G., Davey, F.A., Shepherd, K.D., 2012. Land Health Surveillance: Mapping Soil Carbon in Kenyan Rangelands, in: Nair, P.K.R., Garrity, D. (Eds.), Agroforestry - The Future of Global Land Use. Springer Netherlands, Dordrecht, pp. 455–462. https://doi.org/10.1007/978-94-007-4676-3_22

Vågen, T.-G., Winowiecki, L., Tamene, T., Tondoh, J.E., 2010. The Land Degradation Surveillance Framework (LDSF): Field Guide. World Agroforestry Centre (ICRAF).

Vanlauwe, B., Kihara, J., Chivenge, P., Pypers, P., Coe, R., Six, J., 2011. Agronomic use efficiency of N fertilizer in maize-based systems in sub-Saharan Africa within the context of integrated soil fertility management. Plant Soil 339, 35–50. https://doi.org/10.1007/s11104-010-0462-7

Vanlauwe, B., Wendt, J., Giller, K.E., Corbeels, M., Gerard, B., Nolte, C., 2014. A fourth principle is required to define Conservation Agriculture in sub-Saharan Africa: The appropriate use of fertilizer to enhance crop productivity. F. Crop. Res. 155, 10–13. https://doi.org/10.1016/j.fcr.2013.10.002

Vasu, D., Singh, S.K., Sahu, N., Tiwary, P., Chandran, P., Duraisami, V.P., Ramamurthy, V., Lalitha, M., Kalaiselvi, B., 2017. Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management. Soil Tillage Res. 169, 25–34. https://doi.org/10.1016/j.still.2017.01.006

Waithaka, M., Thornton, P., Shepherd, K., Ndiwa, N., 2007. Factors affecting the use of fertilizers and manure by smallholders: the case of Vihiga, western Kenya. Nutr. Cycl. Agroecosystems 78, 211–224. https://doi.org/10.1007/s10705-006-9087-x

Waswa, B.S., Vlek, P.L.G., Tamene, L.D., Okoth, P., Mbakaya, D., Zingore, S., 2013. Evaluating indicators of land degradation in smallholder farming systems of western Kenya. Geoderma 195–196, 192–200. https://doi.org/10.1016/j.geoderma.2012.11.007

Webb, M.J., Nelson, P.N., Rogers, L.G., Curry, G.N., 2011. Site-specific fertilizer recommendations for oil palm smallholders using information from large plantations. J. Plant Nutr. Soil Sci. 174, 311–320. https://doi.org/10.1002/jpln.201000178

Western, A.W., Blöschl, G., 1999. On the spatial scaling of soil moisture. J. Hydrol. https://doi.org/10.1016/S0022-1694(98)00232-7

Wold, S., Ruhe, A., Wold, H., Dunn W., I.I.I., 1984. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. SIAM J. Sci. Stat. Comput. 5, 735–743. https://doi.org/10.1137/0905052

Wright, A.F., Bailey, J.S., 2001. Organic carbon, total carbon, and total nitrogen determinations in soils of variable calcium carbonate contents using a Leco CN-2000 dry combustion analyzer. Commun. Soil Sci. Plant Anal. 32, 3243–3258. https://doi.org/10.1081/CSS-120001118

Xu, X., He, P., Pampolino, M.F., Chuan, L., Johnston, A.M., Qiu, S., Zhao, S., Zhou, W., 2013. Nutrient requirements for maize in China based on QUEFTS analysis. F. Crop. Res. 150, 115–125. https://doi.org/http://dx.doi.org/10.1016/j.fcr.2013.06.006

Zingore, S., Murwira, H.K., Delve, R.J., Giller, K.E., 2007. Soil type, management history and current resource allocation: Three dimensions regulating variability in crop productivity on African smallholder farms. F. Crop. Res. 101, 296–305. https://doi.org/10.1016/j.fcr.2006.12.006

Zingore, S., Njoroge, S., Chikowo, R., Kihara, J., Nziguheba, G., Nyamangara, J., 2014. 4R plant nutrient management in African agriculture: an extension handbook for fertilizer management in smallholder farming systems.

Highlights of the paper

- An ex-ante approach for estimating scale of for capturing soil spatial variability at farm level is presented.
- Soil organic carbon is a key soil property that influence crop response in unfertilized maize fields.
- Soil organic carbon display short distance variability in unfertilized smallholder maize fields of the region.
- Fine resolution digital soil maps can be considered in strategies for refining fertilizer recommendation in the region.

List of figures.

Fig. 1: Location of the study area in Siaya and Kakamega counties of western Kenya within , a) administrative boundaries , b) soil types , and c) the 100 km^2 block with the 6.25 km^2 tiles in the agro-ecological zone - LM_1 (c). (FURP trial site in (a), Haplic Nitisol in (b), and UM_1 in (c) fall outside the sampling100 km^2 block and the sampling points.

Fig. 2: The hierarchical sampling strategy following the Land Degradation Sampling Framework (LDSF) with the 100 km^2 block with eight randomly selected tiles measuring 6.25 km^2 with four randomly selected sub-tiles measure 0.25 km^2 with randomly selected maize fields and the plots within the maize field.

Fig. 3: The relative importance (percentage) of the contribution of the eight soil predictors to the explained variance for (a) maize yield and (b) plant biomass from regression analysis across the study area.

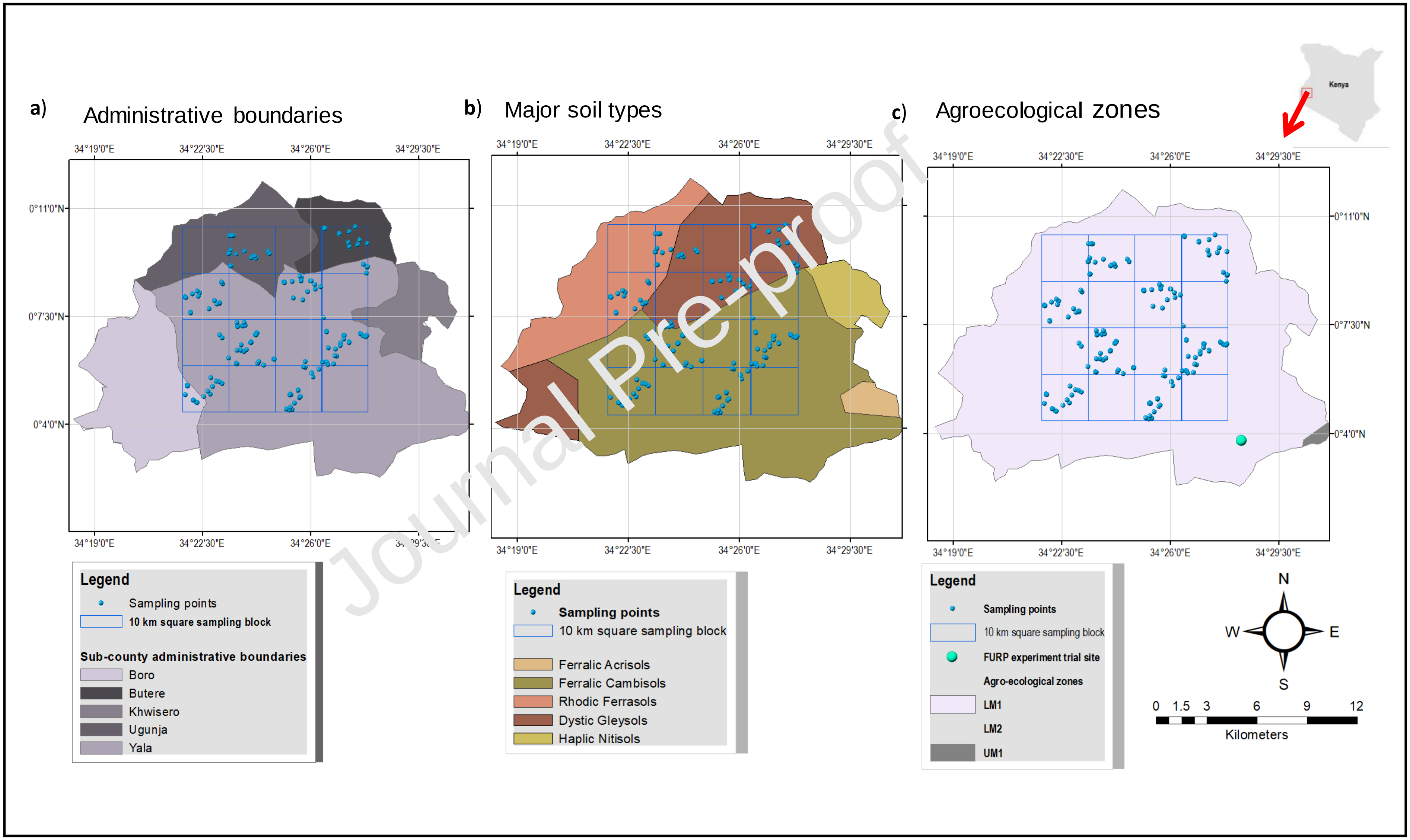
Fig. 4: Graph with the ranking of the soil predictor based on the magnitude of the Akaike Information Criteria (AIC) values for the best regression model from the step-wise multivariate analysis. These values were used to identity the key soil driver that explain variability in maize grain yield or plant biomass.

Fig. 5: The percentage of explained variance for soil organic carbon (SOC), across the mapping units following the hierarchical Land Sampling Degradation Framework - Tile, Sub-tile and Field..

Fig. 6: The percentage of explained variance for crop performance indicators - maize grain yield and plant biomass across the mapping units following the hierarchical Land Sampling Degradation Framework - Tile, sub-tile and field.

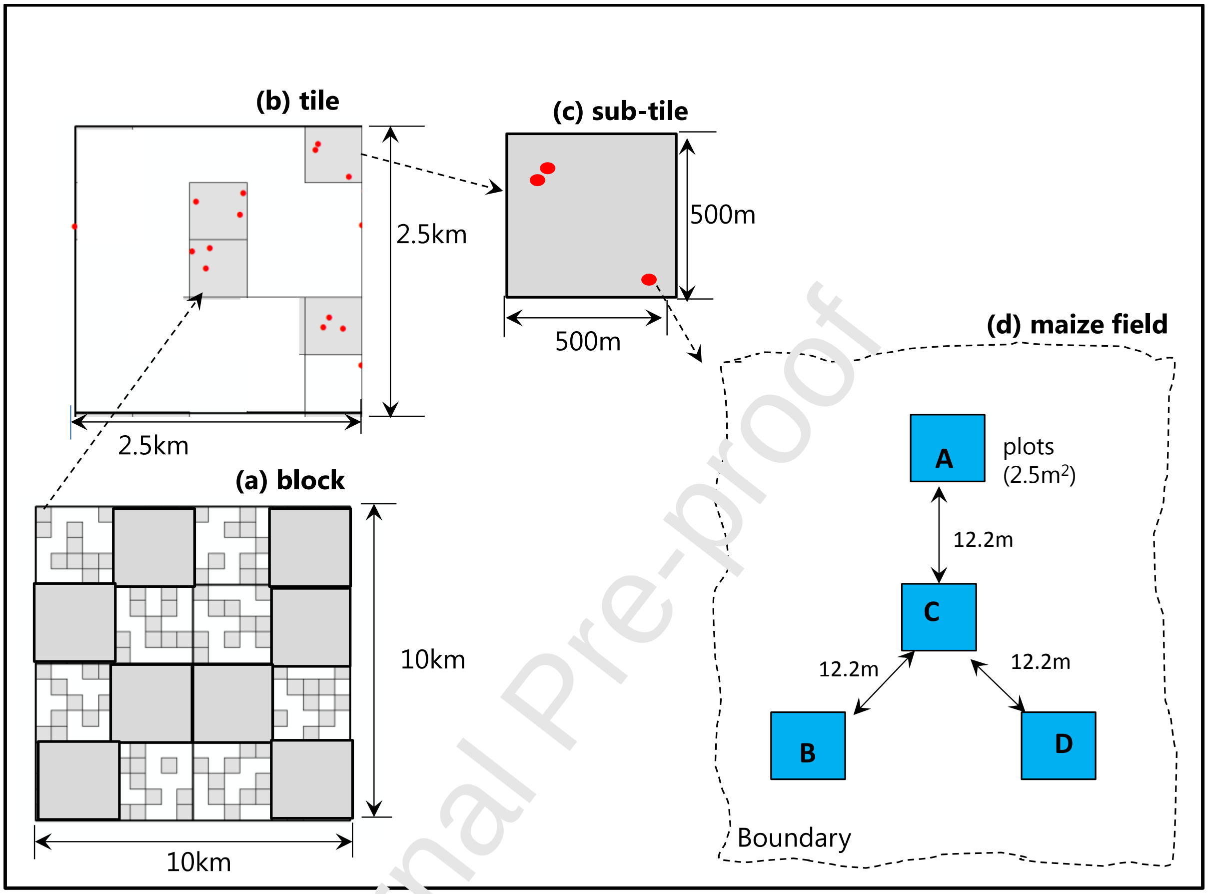
Fig. 7: Decision tree to determine options for providing fertilizer recommendations based on spatial variability in crop response or key soil drivers.

Fig. 1



a) Administrative boundaries
b) Major soil types
c) Agroecological zones
Kenya
34°19'0"E
34°22'30"E
34°26'0"E
34°29'30"E
0°11'0"N
0°7'30"N
0°4'0"N
Legend
Sampling points
10 km square sampling block
Sub-county administrative boundaries
Boro
Butere
Khwisero
Ugunja
Yala
Legend
Sampling points
10 km square sampling block
Ferralic Acrisols
Ferralic Cambisols
Rhodic Ferrasols
Dystic Gleysols
Haplic Nitisols
Legend
Sampling points
10 km square sampling block
FURP experiment trial site
Agro-ecological zones
LM1
LM2
UM1
N
W
E
S
0 1.5 3 6 9 12
Kilometers

Fig. 2


(b) tile
2.5km
2.5km
(c) sub-tile
500m
500m
(d) maize field
(a) block
10km
10km
A
plots
(2.5m²)
12.2m
C
12.2m
12.2m
B
D
Boundary

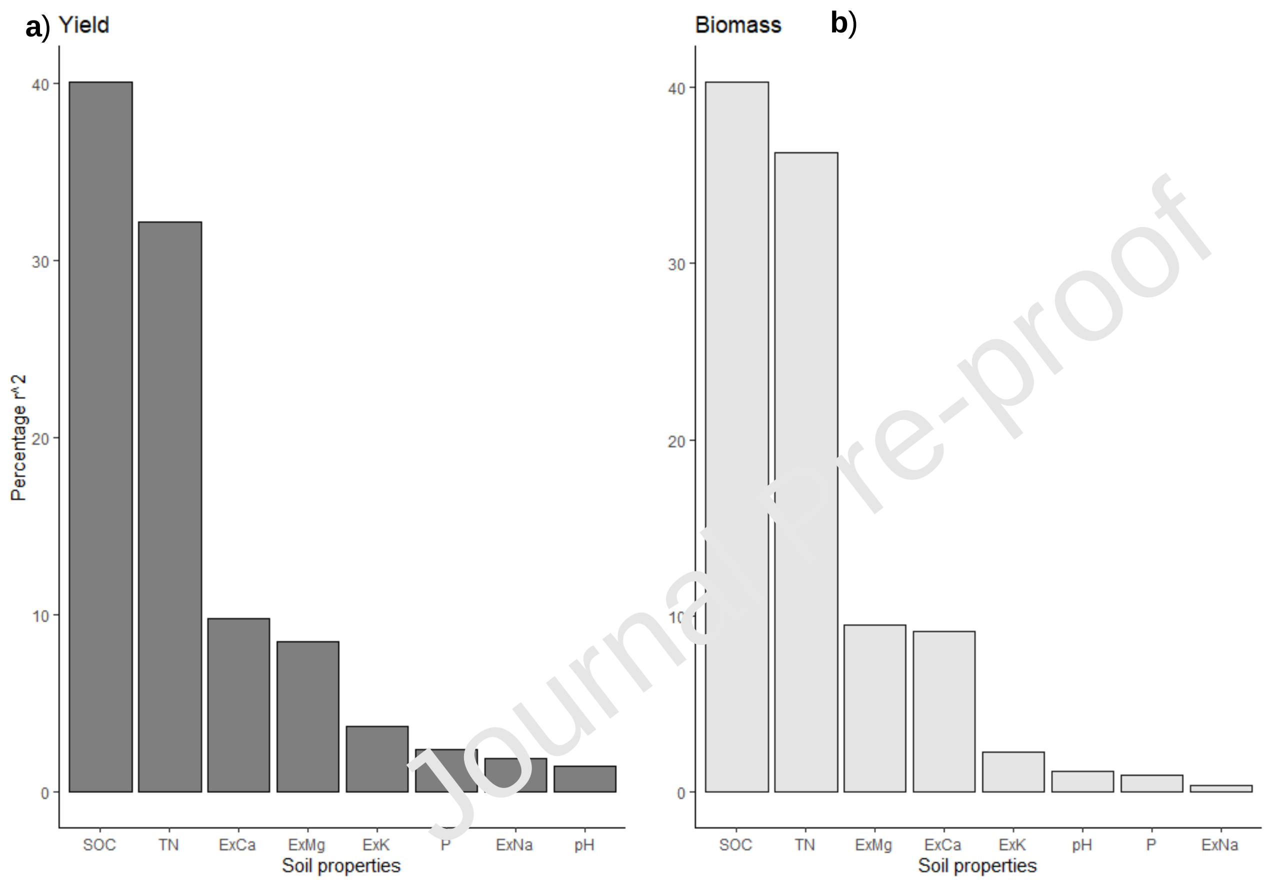


Fig. 3

pH = soil pH, SOC = Soil Organic Carbon, TN = Total Nitrogen, P = Phosphorus, K = Potassium, Ca = Calcium, Mg = Magnesium and Na = Sodium. P, K Ca, Mg and Na were transformed to natural logs

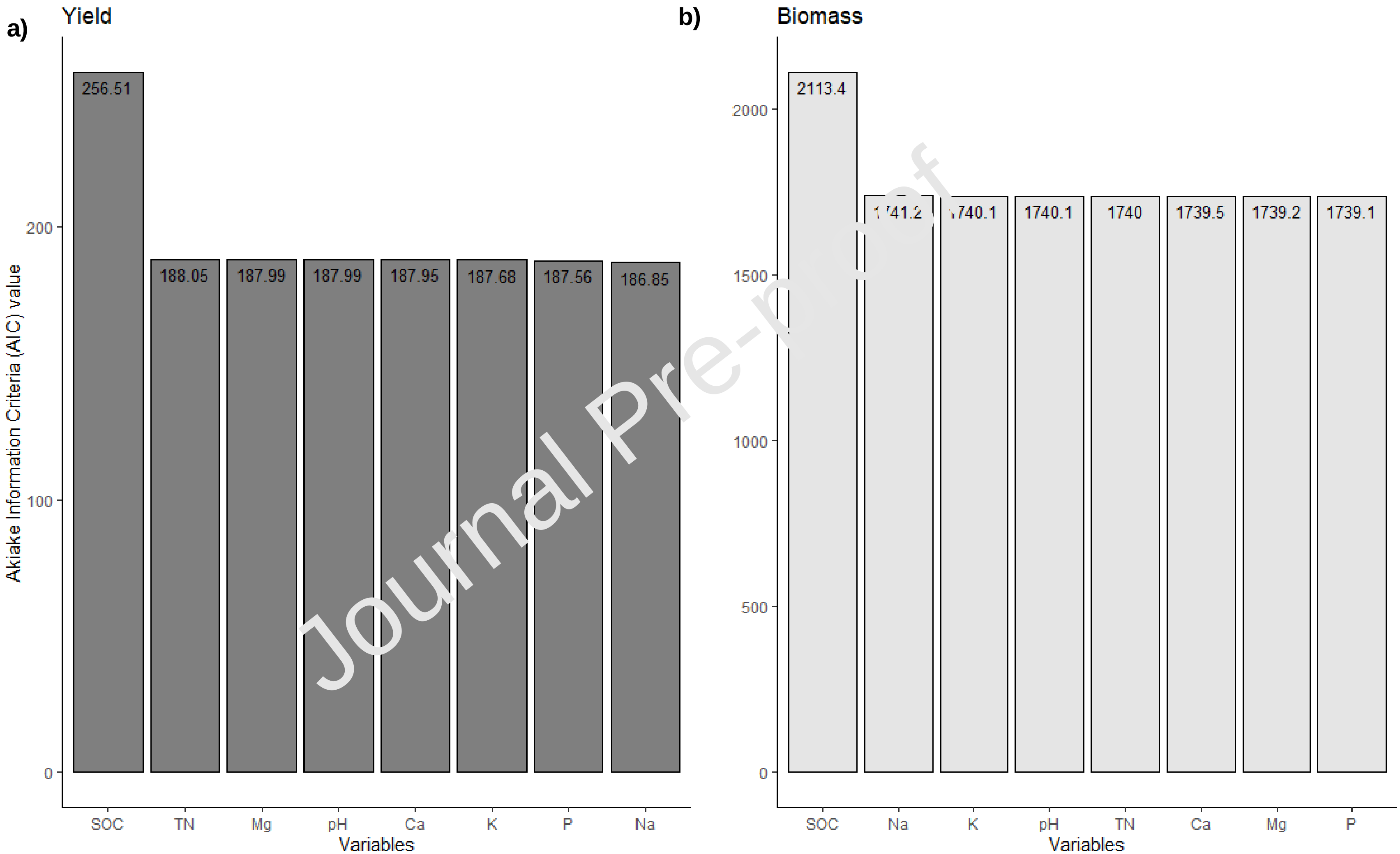
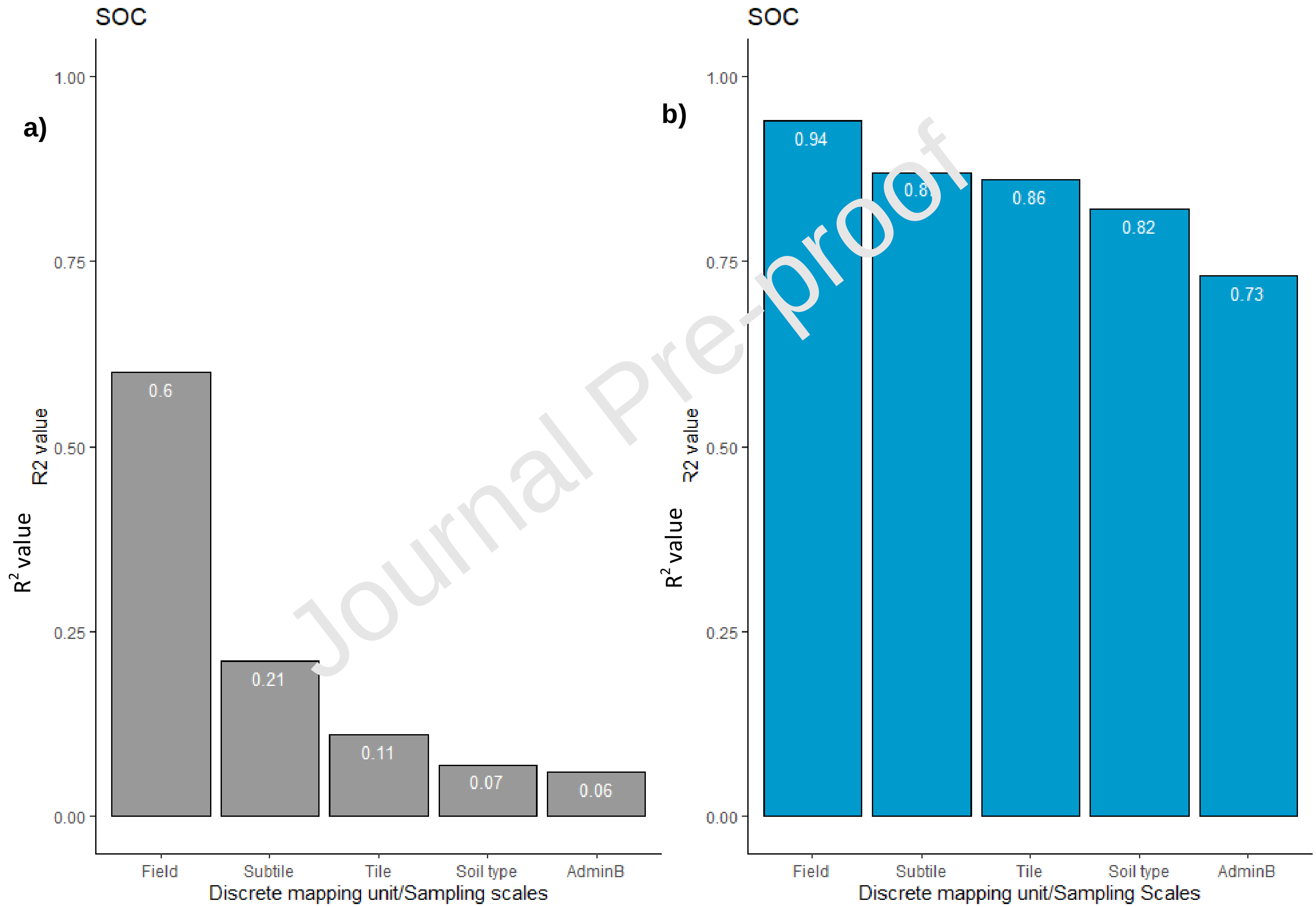

a)
Yield
Akiake Information Criteria (AIC) value
0
100
200
256.51
188.05
187.99
187.99
187.95
187.68
187.56
186.85
SOC
TN
Mg
pH
Ca
K
P
Na
Variables
b)
Biomass
0
500
1000
1500
2000
2113.4
1740.1
1740
1739.5
1739.2
1739.1
SOC
Na
K
pH
TN
Ca
Mg
P
Variables

Fig. 4

pH = soil pH, SOC = Soil Organic Carbon, TN = Total Nitrogen, P = Phosphorus, K = Potassium, Ca = Calcium, Mg = Magnesium and Na = Sodium. P, K Ca, Mg and Na were transformed to natural logs for stepwise multiple linear regression.

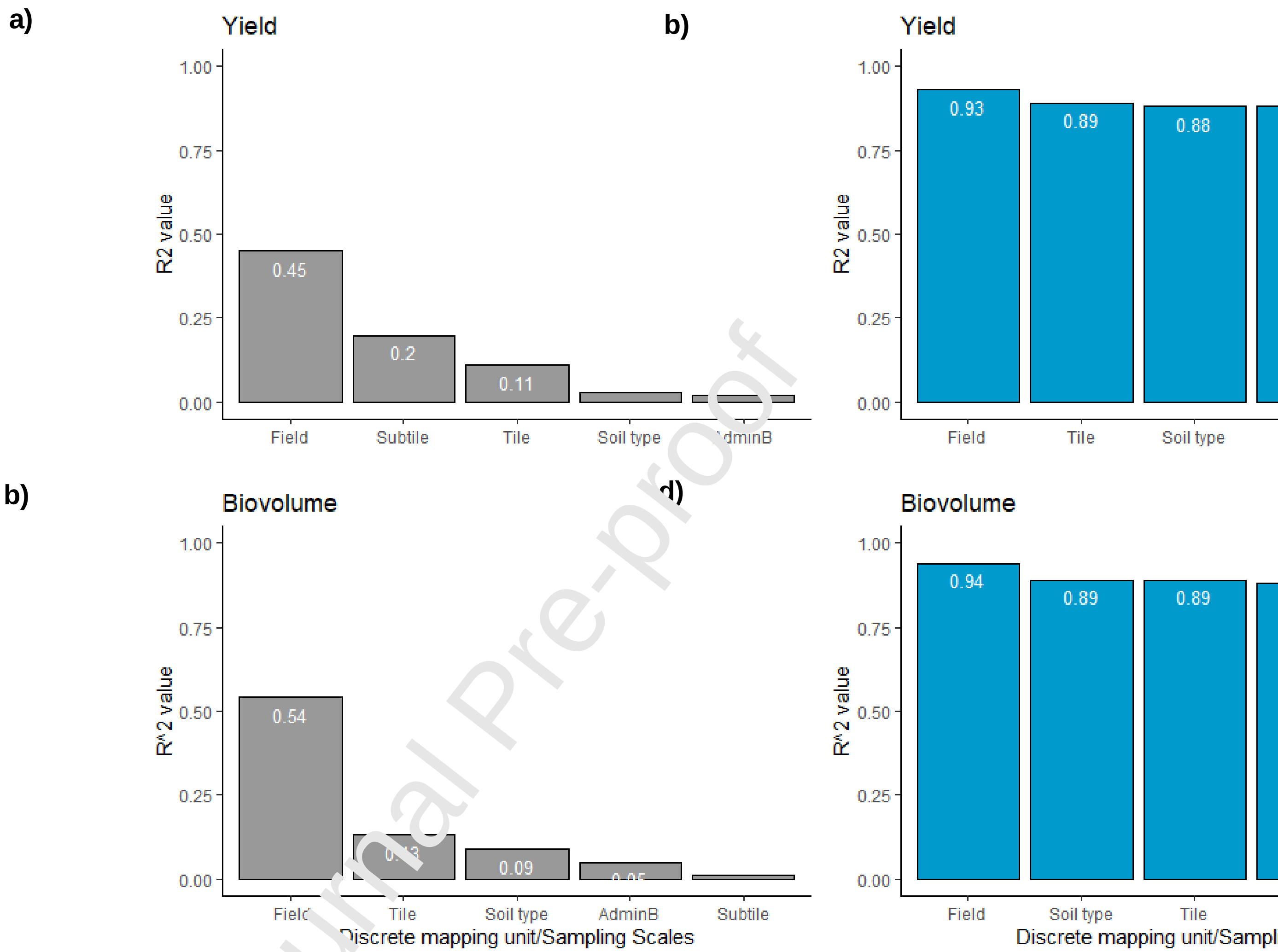
Fig.

5



SOC = Soil Organic Carbon, AdimB= Administrative boundaries

Fig. 6



AdminB = Administrative boundaries

Fig. 7

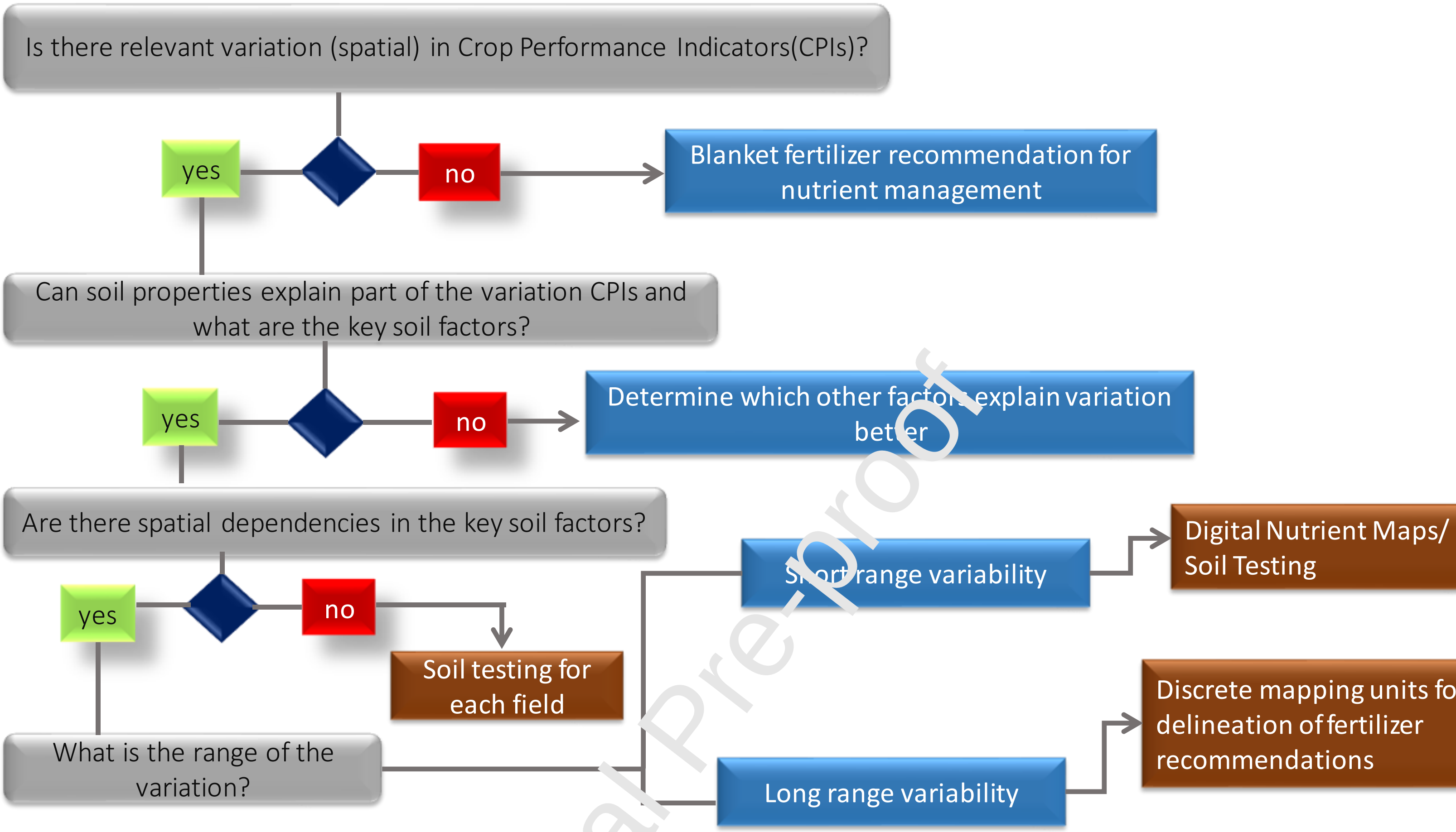

Is there relevant variation (spatial) in Crop Performance Indicators(CPIs)?
yes
no
Blanket fertilizer recommendation for nutrient management
Can soil properties explain part of the variation CPIs and what are the key soil factors?
yes
no
Determine which other factors explain variation better
Are there spatial dependencies in the key soil factors?
yes
no
Soil testing for each field
What is the range of the variation?
Short range variability
Long range variability
Digital Nutrient Maps/ Soil Testing
Discrete mapping units fo delineation of fertilizer recommendations

Table 1: Descriptive statistics of soil properties and crop performance on 203 unfertilized maize plots across a smallholder landscape in Western Kenya.

	Mean	**Median**	**CV (%)**	**Minimum**	**Maximum**
Soil properties					
Soil pH_{water}	5.7	5.6	8.8	4.8	7.4
Total C (%)	1.6	1.6	31.6	0.6	5.2
Total N (%)	0.15	0.14	26.2	0.06	0.3
Mehlich-3 extractable P (mg kg^{-1})	23.0	17.6	74.6	3.7	89.9
Ca (cmol kg^{-1})	5.8	5.0	59.9	0.9	24.5
Mg (cmol kg^{-1})	2.0	1.8	52.1	0.3	6.6
K (cmol kg^{-1})	0.4	0.3	89.3	0.08	2.7
Na (cmol kg^{-1})	0.18	0.16	62.[illegible]	0.01	0.77
Crop performance indicators					
Grain yield (Mg ha^{-1})	3.6	3.20	56.8	0.1	11.3
Plant biovolume (cm^3)	170.4	163.6	43.0	31.0	392.9

Table 2: Pearson pairwise correlation coefficients between soil properties, maize plant bio-volume and grain yield from 197 maize fields across a smallholder landscape in Western Kenya (correlation is significant * at 0.05 and ** at 0.01 level, 2-tailed). P, K Ca, Mg and Na were transformed to natural logs.

	Grain yield	Plant bio-volume
Soil property		
Soil pH	0.08*	0.16*
Total N	0.53**	0.87**
SOC	0.55**	0.89**
P	-0.01	0.02
Ca	0.32*	0.51*
Mg	0.33	0.55
K	0.19	0.26
Na	0.001	0.03
Crop performance indicators		
Plant biovolume	0.55**	1

Abbreviations: N= Nitrogen, P = phosphorus, K = potassium, Ca = Calcium, Mg = magnesium and Na = Sodium.

Table 3: Multivariate regression results of the two models, showing the explained variance r^2 values are the coefficient of determination of the maize yield and plant biovolume for the study area

Crop performance indicator	**Variable**	**Coefficient estimate**	**Standard error**	**t-value**	***p*-value**	
Yield	Intercept	2.429	3.763	0.646	0.5194	
	pH	-0.288	0.396	-0.727	0.4681	
	SOC	3.824	1.367	2.798	0.0057	**
	Total TN	-0.688	1.615	-0.426	0.6704	
	P	-0.319	0.257	-1.244	0.2150	
	K	0.272	0.264	1.031	0.3036	
	Ca	0.586	0.525	1.116	0.2657	
	Mg	-0.545	0.541	-1.008	0.3149	
	Na	-0.216	0.159	-1.354	0.1773	
	r^2 value	**0.324**				
	Adjusted r^2 value	**0.295**				
Biovolume	Intercept	76.671	69.785	1.0990	0.273	
	pH	8.641	7.741	1.1160	0.266	
	SOC	211.335	25.229	8.3780	0.000	***
	Total N	12.672	29.282	0.4330	0.666	
	P	-6.593	5.002	-1.3180	0.189	
	K	0.684	4.966	0.1380	0.891	
	Ca	-1.174	10.022	-0.1170	0.907	
	Mg	-8.432	10.043	-0.8400	0.402	
	Na	-5.602	3.107	-1.8030	0.073	.
	r^2 value	**0.789**				
	Adjusted r^2 value	**0.781**				

level of significance *** = 0.001, ** = 0.01, *= 0.05 and . =0.1

pH = soil pH, N = Nitrogen, P = phosphorus, K = potassium, Ca = Calcium, Mg = magnesium and Na = Sodium. P, K Ca, Mg and Na were transformed to natural log.

Table 4: Spatial dependency of key soil properties, crop performance indicators, and fertilizer response for maize fields in a smallholder Western Kenya in terms of the semi-variogram.

	Semi-variogram model	Nugget: Sill ratio	Range (m)	Spatial dependency
SOC	Exponential	0.60	523	Moderate
Grain yield	Linear	0.24	3291	Strong
Bio-volume	Exponential	0.49	968	Moderate
FR	Spherical	0.50	426	Moderate

SOC = soil organic carbon, FR = fertilizer response