Hierarchical Zoning: Targeted Sampling for Soil Attribute Mapping

Melo[1], D. D.; Cunha[1], I. A.; Brasco[1], T. L.; Oldoni[2], H.; Amaral[1], L. R.

[1] University of Campinas, School of Agricultural Engineering (FEAGRI/UNICAMP), Campinas – SP, Brazil

[2] University of Campinas, Interdisciplinary Center of Energy Planning (NIPE/UNICAMP), Campinas – SP, Brazil

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
The mapping of soil attributes for fertilizer recommendation remains challenging in precision agriculture. Traditionally, this mapping is done through soil sampling in a regular grid, which generally yields good results when done in denser grids. However, due to the high costs associated with sampling and analysis, sparser grids have been adopted, which has not produced good prediction results. Some studies with directed sampling points to obtain more accurate soil maps have been adopted to address the limited sample problem. An interesting methodology involves dividing regions in a field into macro and micro zones of environmental covariate homogeneity and subsequently directing soil sampling points based on the variability of these zones. This technique complements zoning with interpolation and is a promising alternative because it allocates more points where higher attribute variability is expected. In this study, using the fuzzy c-means clustering technique, we divided the agricultural field into macro and micro zones of magnetic susceptibility and historical vegetation index. We then tested two sampling grid densities (one sample/ha and 0.4 samples/ha) and two percentages of sample point targeting (25% and 50%) within these grids, and their effect on mapping available P and K, as well as clay content, in a commercial sugarcane area. We found that directing 50% of soil sampling points in a sparse grid, based on inferred area variability, allowed for more accurate soil maps than those obtained with regular grid sampling. The same pattern of results was not found for directing one sample per hectare, as the regular density with this number of points allows for precise mapping of the tested attributes. Thus, for mapping using sparse grids, directing 50% of soil sampling points based on zones of macro and micro homogeneity of environmental covariates seems attractive for optimizing the mapping of multiple soil attributes compared to regular grid sampling.

Keywords.
Management zones, soil sensing, Digital soil mapping.

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 16th International Conference on Precision Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2024). Title of paper. In Proceedings of the 16th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

Agricultural areas are not uniform, and even in small portions, it is well established that there is soil variability. Therefore, decision-making and the application of amendments and fertilizers should follow the principle of variability, namely variable rate application, to increase productivity potential without depleting natural resources. Mapping soil properties for fertilizer recommendations at varied rates is commonly done using regular grid sampling (Brus 2019). Regular grids emphasize uniform spatial coverage, ensuring a good representation of the entire field with regularly spaced intervals predetermined by samplers. In sampling planning, the number of sample points usually varies according to the financial and/or operational constraints set by those conducting the work. In Brazil, it is known that the size of the sample grid varies widely depending on the region and producer profile (Molin 2017). However, this typical management contradicts ideal grid size recommendations, which should be based on the variability of the attributes of interest rather than financial constraints (Cherubin et al. 2014; Nani et al. 2011). This can compromise decision-making regarding fertilizer application rates throughout the crops (AmaraL and Justina 2019).

Currently, studies are focusing on optimizing sampling grids to achieve more accurate digital soil maps (Baio et al., 2023). Some approaches are based on directing sampling points using auxiliary terrain information that exhibits spatial relationships with the soil variables to be mapped (Pusch et al., 2023; Wang et al., 2023). The complementarity of two or more mapping techniques has proven to be a promising alternative for directing sample points and thus improving map quality. Wang et al. (2023), employing zoning techniques with environmental covariates from DEM-derived terrain data, inferred soil macrovariability, termed as macrozones, identifying areas with more significant heterogeneity. Subsequently, they subdivided these macrozones into microzones and directed sampling points to obtain sampling grids with predictions of higher quality than regular grids. Therefore, dividing the area to be mapped into management zones to infer soil macro and micro variability becomes a promising approach, as these zones represent sub-regions of a field with a certain homogeneity among attributes. In this regard, directing sample points based on the variability of these zones after zoning can lead to more accurate maps. However, relying solely on terrain may be limited in areas with gentle relief, and DEM products have low spatial resolution. In this context, magnetic susceptibility (MS) and vegetation indices (VI) could be substituted for DEM products, as they are covariates related to soil attributes. Another issue is that this method of allocating sample points results in an irregular grid, which can hinder spatial interpolation (Soares 2014). Therefore, it is necessary to study different sizes of sampling grids and the percentage of points to be directed to determine the sampling configurations that yield better mapping results.

There are still many gaps regarding the best methods for creating soil maps for fertilizer prescription and understanding the complementarity of using combined techniques to achieve better prediction results. Therefore, the general objective of this research was to test whether directing sample points based on homogeneous zones determined by environmental covariates allows for producing more accurate digital soil maps compared to traditional maps from regular grids. Additionally, the study aimed to assess whether sample density impacts the outcomes of sample point targeting.

Material and Methods

Study Area

The experiment was conducted in a commercial sugarcane field (20°51'42" S and 47°57'15" W) in the interior of São Paulo, Brazil. The study area covers 71 hectares and has been cultivated for ten years without soil tillage. The area's predominant soil is Ferralsols, and the climate is Aw (tropical savanna) in the Köppen system, with an average annual temperature of 22°C and an average annual precipitation of about 1300 mm.

A dense soil sampling was conducted across the study area using a regular grid of 40 by 40 meters, corresponding to 6 sampling points per hectare, totaling 450 points (Figure 1a). Each sampling point consisted of a composite sample made up of six subsamples collected from a depth of 0 to 25 cm within a five-meter radius from the central point. The composite samples were sent to a commercial soil analysis laboratory to determine the levels of Phosphorus (P) and Potassium (K) and the soil texture (clay content). The descriptive statistics of the analyses are presented in Table 1.

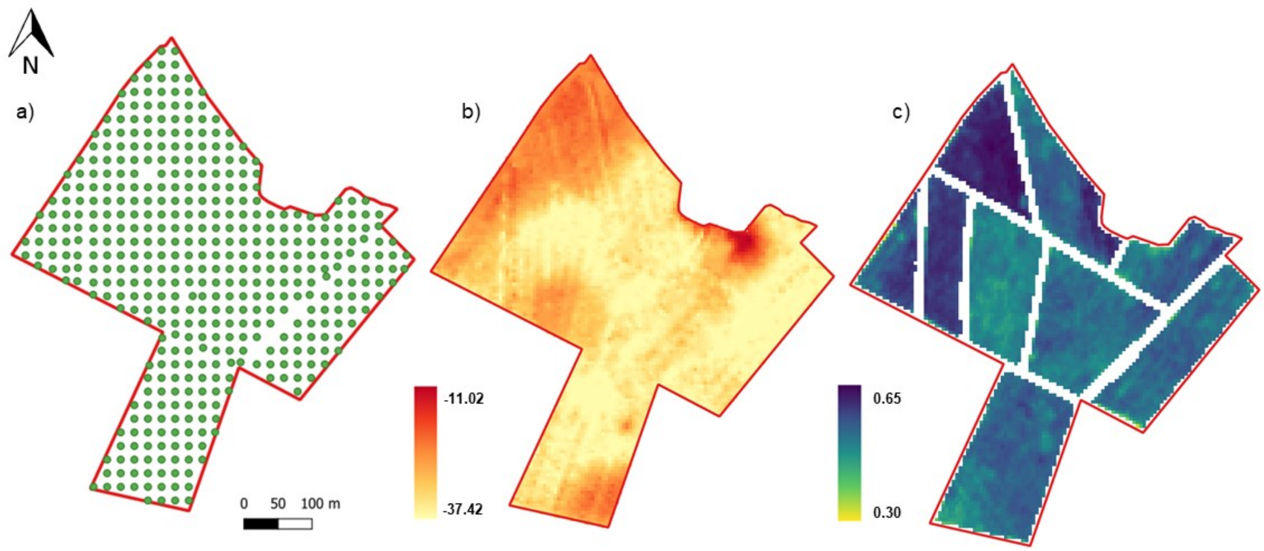

FIGURE 1. a) Total sampling points in a dense grid of 6 samples/ha; b) Soil magnetic susceptibility map; c) Synthetic map of the 5-year historical average of the peak vegetative EVI.

Table 1. Descriptive statistics of the complete sampling for the study area's phosphorus, potassium, and clay content.

	Mean	Median	S.Desviation	C.V	Minimum	Maximum
Clay	530.16	531.50	36.14	6.82	426.00	626.00
K	1.42	1.30	0.37	26.07	0.80	3.30
P	15.06	14.00	6.76	44.90	6.00	45.00

Ambient Variables

The methodology for sample optimization studied here aimed to infer the variability of the study area through the macro and micro spatial variability quantified by environmental covariates. The division into macro-regions sought to identify homogeneous soil zones and subdivide them into micro-regions representing the crop's response to soil variability, thereby directing sampling points. Soil Apparent Magnetic Susceptibility (MS) was the covariate used to delimit the macro-regions. MS is a physical property dependent on the magnetic moments of atoms. In soil, MS is influenced by the amount of magnetic minerals present, primarily controlled by concentrations of magnetite and maghemite (Maher 1986). The application of MS in soil has been studied in agriculture as it allows for identifying and characterizing homogeneous areas (Matias et al. 2013), which can be applied in management zones due to its correlation with agronomic properties (Matos et al. 2023). MS data were obtained at a depth of 37.5 cm using the EM38MK sensor (Geonics), which was moved across the area, collecting data every second (~ five meters) along its path, with passes spaced 15 meters apart. These data were processed using ordinary kriging to create continuous rasters with a spatial resolution of 10 meters (Figure 1b).

To obtain micro-regions we used a synthetic image comprising a time series. Thus, we were able to infer about vegetation homogeneity. Remote sensing was adopted to characterize plant variables using the Enhanced Vegetation Index (EVI). EVI (Equation 1) minimizes the effects of soil and atmosphere, proving effective in high plant vigor situations and is used to infer crop production (Justice et al. 1998). Images from Sentinel-2 over the last five years (2018-2023) were utilized, with an annual image corresponding to the crop's vegetative peak, typically occurring in March each year. These images were composited to create a synthetic image with the average value from the five images (Figure 1c). The boundaries of each field were outlined ten meters inward from the area to reduce carrier reflectance effects in the synthetic image.

$$EVI = 2.5 * (NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1) \qquad \text{Eq. 1}$$

NIR – Near Infrared reflectance (Band 08 of Sentinel-2)

RED – Red reflectance (Band 04 of Sentinel-2)

BLUE – Blue reflectance (Band 02 of Sentinel-2)

Inference of Soil Macro Homogeneity

For the creation of homogeneous MS macrozones, the areas were divided using fuzzy C-Means clustering analysis (Cordoba et al. 2016; Oldoni et al. 2019). The selection of the optimal number of zones was based on the simultaneous analysis of the Fuzzy Performance Index (FPI) (Mcbratney and Moore 1985) and the Modified Partition Entropy (MPE) (Boydell and Mcbratney 2002). The FPI reflects the degree of segregation between observations and the formed clusters, while the MPE expresses the degree of disorder within the clusters, providing a comprehensive view of the resulting structure. The appropriate number of clusters was determined based on the point of the lowest value on the curve, which indicated stabilization in the spatial heterogeneity of the data. This resulted in a smooth curve, suggesting that the marginal gain from increasing the number of zones became unnecessary.

Determining the Optimal Number of Points per Homogeneous Macrozone

The optimal number of points to be directed per zone was determined based on the variability of Magnetic Susceptibility (MS) and the size of each zone. Therefore, MS rasters were cropped for each zone, and using the Coefficient of Variation (CV) - a statistical measure that explains the percentage of pixel variation relative to the mean - the proportional number of samples was determined based on the area size. Essentially, larger areas with higher CV received more points per zone. The formula used to determine the number of points was adapted from Wang et al. (2023) (Equation 2), which starts from an arbitrarily defined number of samples, typically based on practical field constraints such as budgetary considerations for surveying.

In this context, four sampling density scenarios were explored: testing the density of one sample per hectare, often considered a desired density to achieve, and the density of one sample every 2.5 hectares (0.4 samples per ha), commonly adopted in-field practices to optimize survey costs involving sampling and laboratory analysis. Both sampling densities directed 25% and 50% of points using this strategy, with the remaining 75% and 50% following a regular grid pattern.

$$SNh = SN \cdot Ah \cdot CVh / \sum I \;\; Ah \cdot CVh \qquad \text{Eq. 2}$$

SNh – Number of points in each macrozone;

SN – Number of samples to be directed over the entire area, arbitrarily defined;

Ah – Area of the macrozone under study;

CVh – Coefficient of variation of MS in each macrozone

Inference of Soil Micro Homogeneity

For the subdivision of homogeneous zones into smaller regions to indicate the sampling locations for each soil sample, referred as microvariability, the macrozones were divided into microzones. The technique employed utilized only the synthetic EVI image. The Fuzzy C-means clustering method (FCM) (Bezedek et al. 1984) was applied to this image to create homogeneous microzones (Cordoba et al. 2016). This approach allowed subdividing the macro homogeneous zones derived from MS into homogeneous EVI microzones. The microzones in each macrozone was determined based on the number of directed points for the study area, calculated using Equation 2.

Determination of the location of additional points

Following the methodology adapted from Wang et al. (2023), where each macrozone accommodated a certain amount of sampling points equal to the total number of microzones, each microzone received one directed point. Therefore, the pixels within the zones generated by the Fuzzy-C-Means cluster algorithm has dynamic classification and may have different degrees of membership in each cluster (microzone) for each pixel on the map. Thus, to sample the location that is most likely to represent the respective microzone, the pixel with a membership value closest to 1 within each microzone was adopted as the sample selection point within the previously densely sampled grid.

Interpolation

After directing the soil sampling points at each percentile level (25% and 50%), based on each sampling density (1 sample/ha and 0.4 samples/ha), according to homogeneous soil macrozones, the values of the selected soil samples were integrated into the equivalent regular grid (Figure 2). Subsequently, these values underwent a data interpolation process to generate maps. Ordinary kriging was employed for this purpose, following all the principles of variographic modeling as suggested by Oliver and Webster (2014).

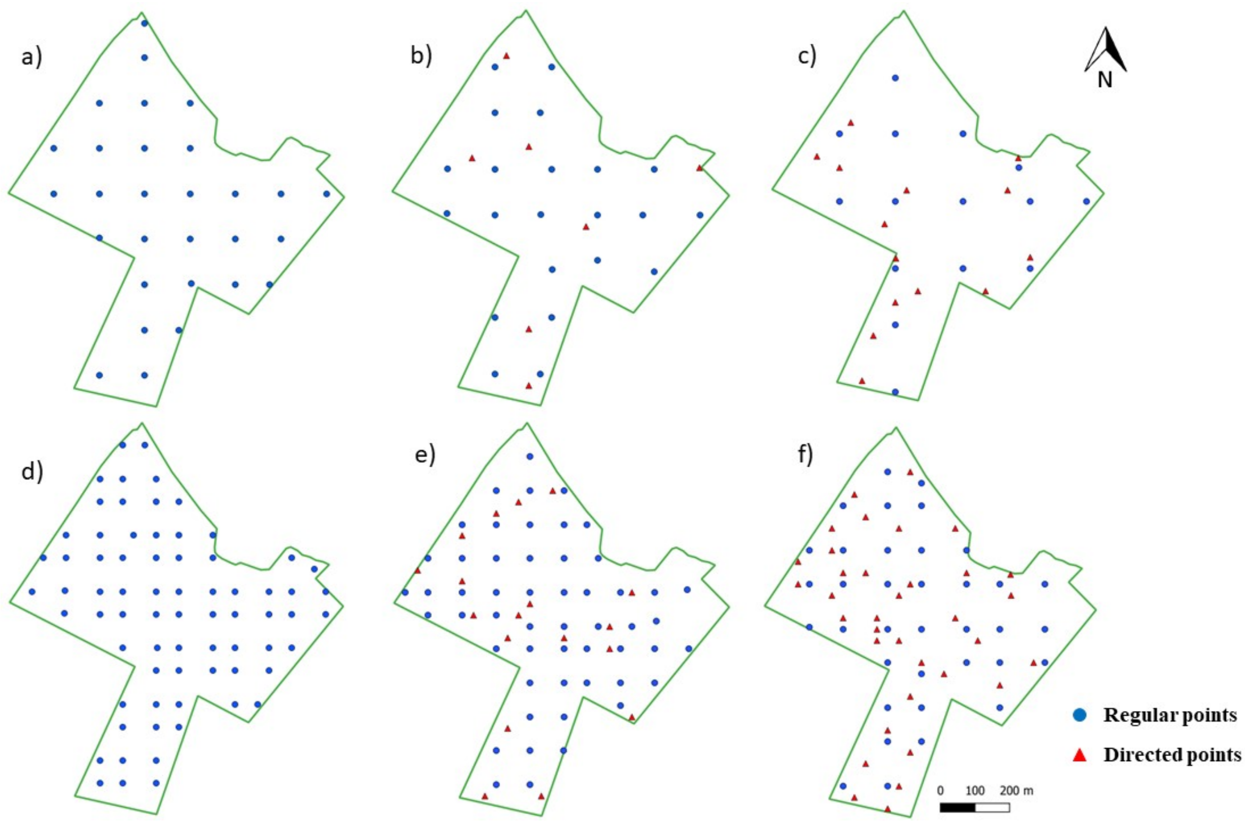

Figure 2. a) Regular grid 0.4 samples/ha; b) Grid with 25% of sampling points directed at 0.4 samples/ha density; c) Grid with 50% of sampling points directed at 0.4 samples/ha density; d) Regular grid 1 sample/ha; e) Grid with 25% of sampling points directed at 1 sample/ha density; f) Grid with 50% of sampling points directed at 1 sample/ha density.

Analysis Approach:

An external validation procedure was adopted to assess whether the sample optimization based on the proposed methodology improves the quality of soil fertility maps compared to a regular sampling grid with the same number of samples. Therefore, the sampling points not used in constructing the grids served as the reference, resulting in 369 and 264 validation points for the 0.4 samples/ha and 1.0 samples/ha grids, respectively. The interpolated values of different soil properties were extracted at the coordinates of these external validation points. Thus, the actual

values were compared with the predicted values using Root Mean Square Error (RMSE) and Spearman's correlation coefficient (r).
Additionally, adjustments to the variograms were explored to identify if increasing the percentage of directed points improved the capture of spatial variability. Furthermore, the Moran's Index was also calculated to infer if the spatial clustering of soil properties aligned with the proposed methodology.

Results and Discussion

Mapping Clay Content

1 sample/hectare

The sample configuration with 50% of directed sampling points at one sample per hectare density yielded the best prediction results for clay content in the experimental area (Table 2). The RMSE for clay content showed slightly better results for the grid with 50% directed points, with a value of 31.82 g/kg, while the error for the regular grid was 31.95 g/kg, followed by 33.79 g/kg for the grid with 25% directed points. However, these values are very close, indicating little difference between the sampling configurations. This similarity may be associated with soil homogeneity within the study area, as evidenced by the variability scale of complete grid sampling (Table 1), showing low data variability with a coefficient of variation of about 6%, and a clay content ranging from 426 to 626 g/kg. Despite the differences in grid types, both 1 sample/ha grids delivered satisfactory results for mapping this attribute (Figure 3). Studies on increasing the number of soil sampling points in agricultural fields have shown that uncertainty in kriging mapping tends to decrease as the number of sampling points increases (Amaral and Justinna, 2019; Cherubin et, al. 2014; 2015). Our results support these findings, as the density of 1 sample per hectare, regardless of grid type, accurately characterized soil variations, suggesting that the regular grid can be used for clay mapping, as it requires less operational effort and delivers satisfactory results at this density.

Table 2. Soil clay content prediction performance for the validation samples based on different sampling point allocation methods at two sampling densities. The best results are in bold.

	25%		50%		Regular	
	RMSE	**r**	**RMSE**	**r**	**RMSE**	**r**
1 samples/ha	33.79	0.45	**31.82**	**0.52**	31.95	0.51
0.4 samples /ha	31.91	0.44	**35.65**	**0.39**	34.39	0.40

RMSE: Root Mean Square Error; r: Spearman's correlation coefficient

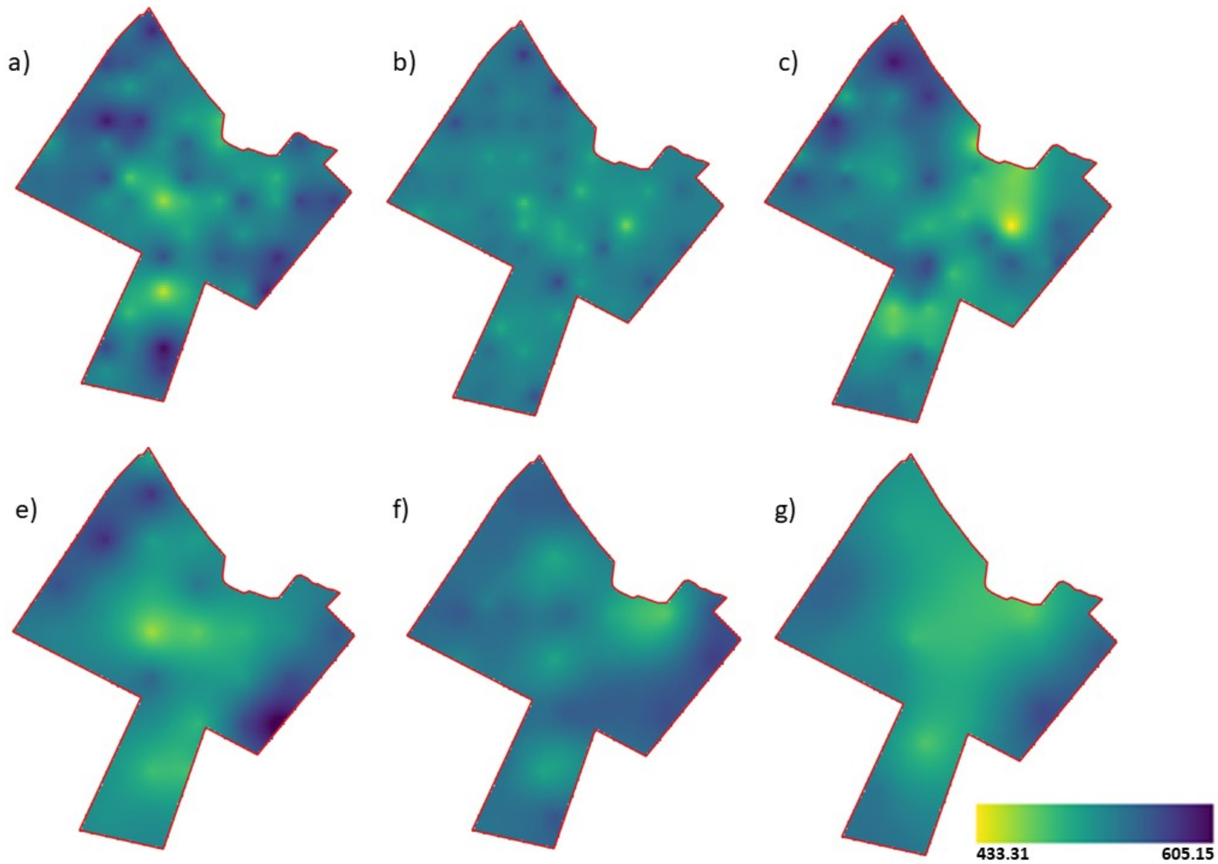

Figure 3. a) Clay map derived from the regular grid with a density of 1 sample/ha; b) Clay map derived from the grid with 25% of the sampling points directed with a density of 1 sample/ha; c) Clay map derived from the grid with 50% of the sampling points directed with a density of 1 sample/ha; d) Clay map derived from the regular grid with a density of 0.4 sample/ha; e) Clay map derived from the grid with 25% of the sampling points directed with a density of 0.4 sample/ha; f) Clay map derived from the grid with 50% of the sampling points directed with a density of 0.4 sample/ha.

0.4 sample/ha

On the other hand, when considering the characterization of clay through the grid with 0.4 sampling points per hectare, directing 25% of the sampling points proved more effective for mapping (Table 2). The reduction in sampling uncertainty was more pronounced in this configuration, allowing for a more precise capture of soil variability when points were directed. Although results were more favorable with 25% of the sampling points directed, the analysis of the correlation between covariates and the attribute in question suggests that the improvement was not directly due to the directing method, as there was no clear correlation between the covariates and the attribute (Table 3). In the regular grid of 0.4 samples per hectare, the distance between sampling points is approximately 158 meters. Examining the variogram parameters for interpolation of the grid with all sampling points, it becomes evident that the range is 220 meters, indicating low spatial dependence even with a sampling density of 6 points per hectare (Table 4). This pattern is corroborated by Moran's index, which also suggests reduced spatial clustering (Table 4). Previous studies have highlighted that the distribution of random points within the regular sampling grid tends to result in lower mapping errors than a completely regular grid (Baio et al., 2023). The results of this study align with these findings, as even in the absence of a clear relationship between the attribute and covariates, directing points proved effective by maintaining spatial coverage in 75% of the regular points, while the 25% directed points allowed for shorter distances, thus improving interpolation quality. However, directing 50% of the points resulted in loss of equidistance between points, yielding the worst results among the three configurations.

The low clay variability in this area makes directing 50% of the sampling points inefficient, as the methodology requires attribute variability and correlation between it and the covariates, which was not found in this study. Therefore, directing 25% can provide superior results while adopting a more conservative approach, as it initially maintains spatial coverage of the area and provides points at closer distances, favoring the capture of variability. Thus, directing points becomes more advisable for mapping clay in sparse grids than using a regular sampling grid.

Table 3. Spearman correlation between soil attributes and environmental variables used for delimitation of macro and microzones.

	P	K	Clay	MS	EVI
P	1.00				
K	0.170	1.00			
Clay	-0.003	0.086	1.00		
MS	-0.370	-0.001	0.140	1.00	
EVI	-0.350	-0.059	0.012	0.310	1.00

Table 4. Variogram parameters (range – m (A), nugget effect (C0), and sill (C1)); and cross-validation (RMSE and r² between predicted and observed) for soil attributes phosphorus (P), potassium (K), and clay; spatial clustering measured by Moran's index and its significance.

	C0	C1	A	RMSE	r	model	Moran	significance
P	12	30	400	4.78	0.70	Exp	0.46	0.001
K	0.04	0.1	395	0.29	0.57	Exp	0.29	0.001
Clay	200	1000	220	26.06	0.68	Exp	0.36	0.001

Phosphorus Mapping

1 sample ha

A regular grid sampling resulted in the best mapping of P, followed by the grid with 25% of directed sampling points and subsequently by the grid with 50% at a density of 1 sample per hectare. The P maps demonstrate the grids' ability to map the nutrient, where the regular grid allows for well-defined delineation of regions with low and high concentrations of the element, a similar behavior observed for the 25% grid but with a smoother representation of these regions. On the other hand, the 50% directed points grid generated a completely smooth map, making it unable to identify regions of high and low attribute availability (Figure 3a, 3b, and 3c). The best variogram adjustments, as well as the largest clusters, were observed for phosphorus in the grid with total points (Table 4), a behavior that is not commonly observed, as phosphorus often exhibits a highly random distribution, as reported in several studies (López-Castañeda et al. 2022; Bortega et al. 2013; Silva and Chaves, 2001). This farm has a history of sugarcane cultivation, with 10 years without area renovation (replanting), where fertilization with this nutrient is applied in low doses along the ratoons, aiming to maintain soil fertility without depleting it, contributing to a certain stability in soil nutrient levels due to natural conditions. For the grid with one sample per hectare, the best RMSE and correlation results between predicted and observed values were associated with the regular grid (Table 5), as expected, since this density satisfactorily supports digital mapping through ordinary kriging (Cherubins et al. 2014; AmaraL and Justinna 2019). Additionally, the number of sample points (71) meets the recommendations for variogram adjustments (Oliver and Webster 2014). Therefore, similar to what was observed for clay, when the sample density used is sufficient to capture property variability, zoning with covariates to direct sampling does not contribute to modeling variability and, consequently, improving predictions, even if the covariates show high correlations with the soil attribute (Table 3). Thus, similarly to clay, regular grid sampling is recommended for dense mapping as it accurately maps the attribute without the need for greater computational and operational effort.

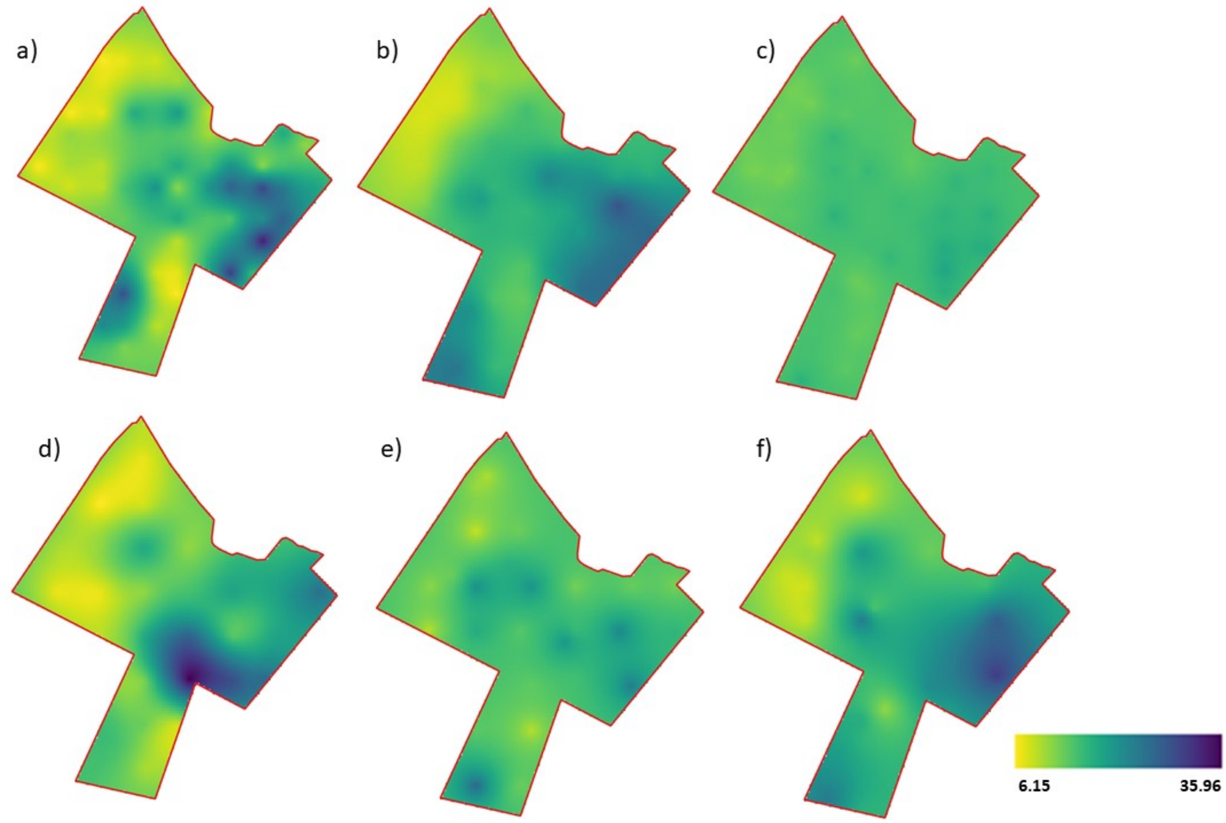

Figure 3. a) P map derived from the regular grid with a density of 1 sample/ha; b) P map derived from the grid with 25% of the sampling points directed with a density of 1 sample/ha; c) P map derived from the grid with 50% of the sampling points directed with a density of 1 sample/ha; d) P map derived from the regular grid with a density of 0.4 sample/ha; e) P map derived from the grid with 25% of the sampling points directed with a density of 0.4 sample/ha; f) P map derived from the grid with 50% of the sampling points directed with a density of 0.4 sample/ha.

Tabela-5. Performance in mapping phosphorus (P) content for validation samples based on different sampling point allocation methods at two sampling densities. The best results are in bold.

	25%		**50%**		**Regular**	
	RMSE	**r**	**RMSE**	**r**	**RMSE**	**r**
1 samples/ha	5.59	0.67	6.28	0.65	**5.56**	**0.68**
0.4 samples /ha	5.96	0.58	**5.20**	**0.71**	6.54	0.55

RMSE: Root Mean Square Error; r: Spearman's correlation coefficient.

0.4 sample ha

However, for the sparse grid of 0.4 sample points per hectare, directing 50% of the points using the proposed methodology resulted in a more significant reduction in sampling error, followed by directing 25% of the points and the regular grid (Table 5). Directional sampling proved effective in this case as it adhered to the principle of spatial coverage, allowing for point allocation at shorter distances, which favored variogram adjustments. Directional sampling enabled better identification of areas with pronounced variations (Figure 3d, 3e, and 3f), unlike the regular grid, which tends to smooth estimates due to the absence of points at shorter lags. Another point to highlight was the correlation between the measured variables and the target attribute (Table 3). Since the methodology aims to capture attribute variability based on covariate variability, correlation becomes key to success. Thus, obtaining a wider range of values for the target variable becomes possible, which tends to improve modeling.

Finally, as the clustering technique used in zone creation aims to minimize the variability of covariates within each zone and maximize differentiation between clusters, there must be variability for this objective to be achieved. This can be observed in soil sampling data, where the highest variability among attributes is found for phosphorus (Table 1). Therefore, when using this method, a preliminary exploration of correlations between variables is necessary, aiming to use one or more auxiliary variables related to the variable being mapped (Push et al., 2023). Plant, soil, topography, and even management covariates may influence the spatial distribution of soil chemical properties, and thus, the dataset needs to be thoroughly explored in each specific situation. In this case, a two-stage sampling plan could be an option (Szatmári et al. 2019). If these observations are met, directed grid sampling for mapping P at low sample densities is recommended, as it tends to reduce passive sampling errors associated with regular grids.

Potassium Mapping

1 sample ha

The potassium (K) mapping with regular sample points for one sample per hectare grid provided better predictive results, followed by the grids with 50% and 25% of directed points (Table 6). These results are evident when observing the maps of the areas where the smoothing of the maps occurred gradually with the increase in the number of directed points (Figure 4a, 4b, and 4c). The Moran's index shows low clustering among samples (Table 4) for the total points grid (6 samples per hectare), and cross-validation parameters indicate low spatial dependence (Table 4), meaning high nugget effect and low contribution, which suggests that a significant portion of the data variability is not related to its location. Thus, the poorer results for K seem to be associated with the poorer fitting of the variograms (Table 4). Sugarcane cultivation requires large amounts of this nutrient (Cantarella et al. 2022), which is generally applied at fixed rates across the areas. Another point to highlight is that sugarcane can absorb more of this nutrient than necessary for its optimal growth and production (luxury consumption), significantly reducing the element's levels in the soil (Otto et al. 2019). Thus, over the fertilization cycles, the spatial distribution of this attribute becomes more random due to management practices and nutrient extraction by the crop, leading to diminished gains when using combined covariate optimization sampling methods. Therefore, in this case, using regular grids would be more appropriate as it allows for more precise mapping and requires less effort than directional methods.

Tabela-6. Performance in mapping potassium (K) content for validation samples based on different sampling point allocation methods at two sampling densities. The best results are in bold.

	25%		50%		Regular	
	RMSE	**r**	**RMSE**	**r**	**RMSE**	**r**
1 samples/ha	0,45	0,45	0,35	0,40	**0,34**	**0,49**
0.4 samples /ha	0,37	0,12	**0,35**	**0,5**	0,36	0,24

RMSE: Root Mean Square Error; r: Spearman's correlation coefficient

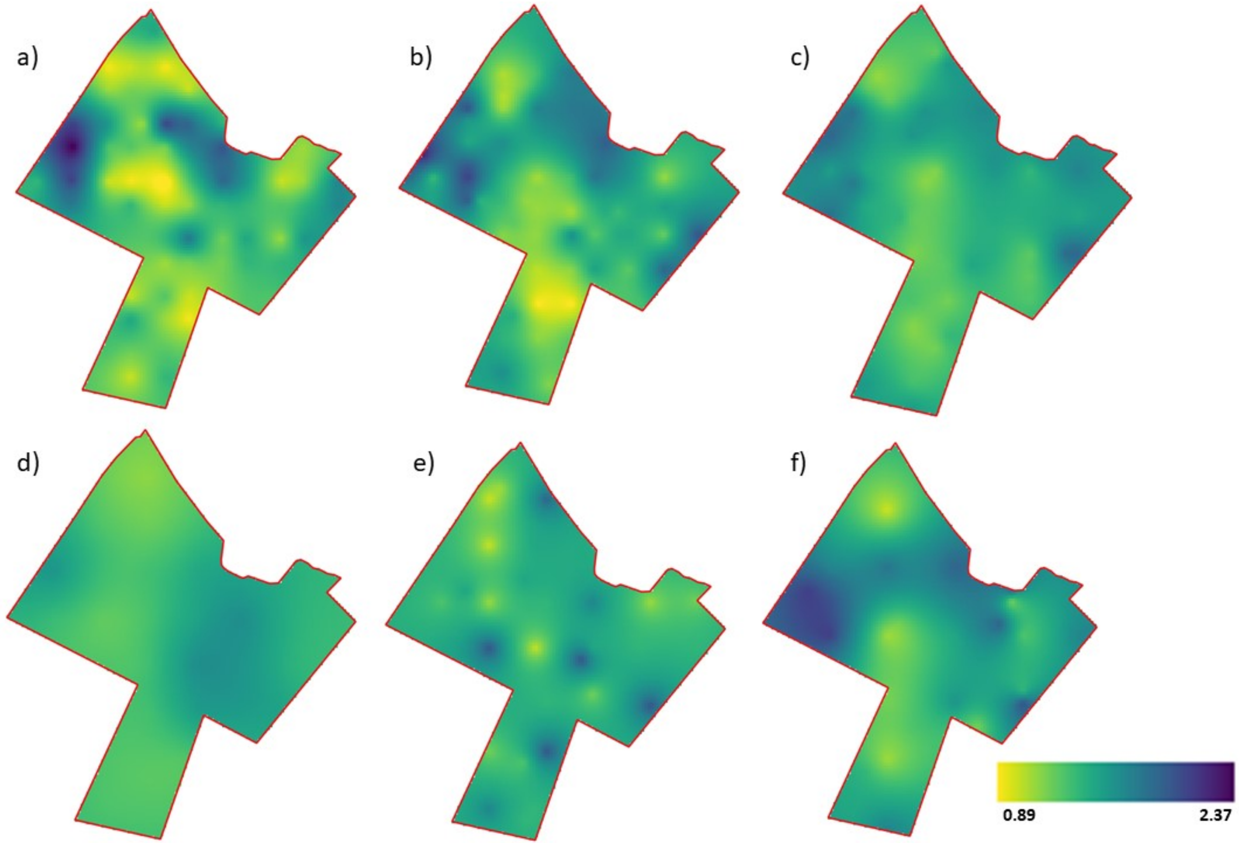

Figure 4. a) K map derived from the regular grid with a density of 1 sample/ha; b) K map derived from the grid with 25% of the sampling points directed with a density of 1 sample/ha; c) K map derived from the grid with 50% of the sampling points directed with a density of 1 sample/ha; d) K map derived from the regular grid with a density of 0.4 sample/ha; e) K map derived from the grid with 25% of the sampling points directed with a density of 0.4 sample/ha; f) K map derived from the grid with 50% of the sampling points directed with a density of 0.4 sample/ha.

0.4 sample ha

However, sample targeting yielded the best prediction results for the sparsest grid (0.4 samples per hectare). The allocation of points resulted in better prediction performance (RMSE and r), with 50% of sample points allocated (Table 6), performing better than the 1 point per hectare sampling. The result from the regular grids and those with 25% of points allocated was unsatisfactory, as the observed spatial dependence is low. Therefore, these grids have greater coverage of sample points in the area, which contributes to generating unsatisfactory results by losing the equidistance between point pairs, thereby not capturing the spatial dependence of K as well as the 50% directed grid (r = 0.5), as can be seen in figures (4d, 4e, and 4f). Previous studies have shown that very sparse regular grids are not recommended as they lose the ability to capture spatial dependence (Cherubin et al. 2014). However, using covariates diminishes the effect of reduced soil sample density (Teixeira et al. 2017). In this research, we confirm these findings, as allocating points at short distances through auxiliary information for zoning improved predictions due to capturing this low spatial dependence through better variogram adjustments. Authors like Wang et al. (2023) have also reported that directing sample points using fuzzy methodology leads to lower sampling errors compared to other targeting methods for mapping K, especially with a limited number of soil samples. Therefore, directing 50% of sample points based on soil moisture and historical EVI can be a viable alternative to obtain more accurate digital soil maps for characterizing K in sparsely sampled grids.

Conclusion

For mapping using directed grids with one sample per hectare, we observed a slight predictive gain only in clay content compared to the regular grid. However, even for this attribute, there are no significant gains with allocation, as the regular grid at this density can accurately capture spatial dependence. Another point is that targeting can lead to loss of equidistance and may even impair mapping results. Therefore, in sampling situations with a density of one sample per hectare for mapping the tested attributes, using a regular grid continues to be recommended, as this density is sufficient to capture soil variability.

Although mapping with a sparse grid of 0.4 samples per hectare produces poor predictive results, sample targeting has been shown to reduce interpolation errors compared to a regular grid. In this study, we tested two chemical and one physical attribute, finding that targeting in a sparse grid was more effective. Spatial dependence, high variability of the target attribute, and correlation between covariates and mapped attributes result in more accurate maps when 50% and 25% of points are targeted compared to a regular grid. However, without prior knowledge of these parameters, agronomic knowledge can serve as a valuable indicator, as targeting aims to capture soil variability based on covariate variability. Thus, when mapping in a sparse grid, we recommend targeting 50% of sample points, as this helps mitigate prediction errors associated with low-density sampling.

Acknowledgments

The authors would like to thank the owner of the Aliada farm. We thank São Paulo Research Foundation - FAPESP for the scholarship granted to the first author (Process number 2023/02592-4). We also acknowledge the support from CNPq-Brazil with the Level-2 Productivity Grant to the fourth author (306867/2022-2) and FAPESP for financial support for this research (Process number 2022/03160-8).

References

Amaral, L. R. D., & Justina, D. D. D. (2019). Spatial dependence degree and sampling neighborhood influence on interpolation process for fertilizer prescription maps. *Engenharia Agrícola,* 39, 85-95.

Baio, F. H. R., Alixame, D., Neves, D. C., Teodoro, L. P. R., da Silva Júnior, C. A., Shiratsuchi, L. S., ... & Teodoro, P. E. (2023). Adding random points to sampling grids to improve the quality of soil fertility maps. *Precision Agriculture*, 24(5), 2081-2097.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3), 191-203.

Boydell, B., & McBratney, A. B. (2002). Identifying potential within-field management zones from cotton-yield estimates. *Precision agriculture*, 3, 9-23.

Brus, D. J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma,* 338, 464-480.

Cherubin, M. R., Santi, A. L., Eitelwein, M. T., Amado, T. J. C., Simon, D. H., & Damian, J. M. (2015). Dimensão da malha amostral para caracterização da variabilidade espacial de fósforo e potássio em Latossolo Vermelho. *Pesquisa Agropecuária Brasileira*, 50, 168-177.

Cherubin, M. R., Santi, A. L., Eitelwein, M. T., Menegol, D. R., Ros, C. O. D., Pias, O. H. D. C., & Berghetti, J. (2014). Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. *Ciência Rural*, *44*, 425-432..

Cantarella, H.; Raij, B. van; Quaggio, J. A.; Boaretto, R. M., Mattos, D. (2023) Recomendação de adubação e calagem para o Estado de São Paulo. 2ª ed. *Instituto Agrônomico de Campinas*

-*IAC.Campinas*. (Boletim Técnico, 100).

Córdoba, M. A., Bruno, C. I., Costa, J. L., Peralta, N. R., & Balzarini, M. G. (2016). Protocol for multivariate homogeneous zone delineation in precision agriculture. *Biosystems engineering*, *143*, 95-107.

Farid, H. U., Bakhsh, A., Ahmad, N., Ahmad, A., & Mahmood-Khan, Z. (2016). Delineating site-specific management zones for precision agriculture. *The Journal of agricultural science*, *154*(2), 273-286.

Matias, S. S. R., Marques Júnior, J., Siqueira, D. S., & Pereira, G. T. (2013). Modelos de paisagem e susceptibilidade magnética na identificação e caracterização do solo. *Pesquisa Agropecuária Tropical*, *43*, 93-103.

Matos, A. P. D., Matias, S. S. R., Nunes, R. K. L., Morais, E. M., & Tavares Filho, G. S. (2023). Soil management of limed areas cultivated with banana identified by magnetic susceptibility. *Revista Ceres*, *70*, 17-24.

McBratney, A. B., & Moore, A. W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and forest meteorology*, *35*(1-4), 165-185.

Molin, J. P. (2017). Agricultura de precisão: números do mercado brasileiro. *Boletim Técnico*, *3*, 7.

Karimi, A., Haghnia, G. H., Ayoubi, S., & Safari, T. (2017). Impacts of geology and land use on magnetic susceptibility and selected heavy metals in surface soils of Mashhad plain, northeastern Iran. *Journal of Applied Geophysics*, *138*, 127-134.

Nanni, M. R., Povh, F. P., Demattê, J. A. M., Oliveira, R. B. D., Chicati, M. L., & Cezar, E. (2011). Optimum size in grid soil sampling for variable rate application in site-specific management. *Scientia Agricola*, *68*, 386-392.

Oldoni, H., Terra, V. S. S., Timm, L. C., Júnior, C. R., & Monteiro, A. B. (2019). Delineation of management zones in a peach orchard using multivariate and geostatistical analyses. *Soil and Tillage Research*, *191*, 1-10.

Otto, R., Altarugio, L., & Sanches, G. (2019). Atualizações sobre exigências nutricionais da cana-de-açúcar para fins de manejo da adubação. *Informações Agronômicas*, (3), 1-10.

Oliver, M. A., & Webster, R. (2014). A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, *113*, 56-69.

Pusch, M., Samuel-Rosa, A., Magalhães, P. S. G., & do Amaral, L. R. (2023). Covariates in sample planning optimization for digital soil fertility mapping in agricultural areas. *Geoderma*, *429*, 116-252.

Soares, A. Geostatica para as ciencias da terra e do ambiente. 3° Edição. IST Press; (1 janeiro 2000).

Szatmári, G., László, P., Takács, K., Szabó, J., Bakacsi, Z., Koós, S., & Pásztor, L. (2019). Optimization of second-phase sampling for multivariate soil mapping purposes: Case study from a wine region, Hungary. *Geoderma*, *352*, 373-384.

Wang, Y., Qi, Q., Bao, Z., Wu, L., Geng, Q., & Wang, J. (2023). A novel sampling design considering the local heterogeneity of soil for farm field-level mapping with multiple soil properties. *Precision Agriculture*, *24*(1), 1-22.