Optimized soil sampling location in management zones based on apparent electrical conductivity and landscape attributes

Gabriela K Michelon[a]; Guilherme M Sanches[c]; Igor Q Valente[c]; Claudio L Bazzi[a,b]; Paulo L de Menezes[a]; Lucas R Amaral[c]; Paulo S G Magalhaes[a,b]

[a] Federal University of Technology – Parana, Medianeira, Brazil; [b] CNPq, Brasília, Brazil.
[c] University of Campinas, School of Agricultural Engineering, Campinas, SP - Brazil

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *One of the limiting factors to characterize the soil spatial variability is the need for a dense soil sampling, which prevents the mapping due to the high demand of time and costs. A technique that minimizes the number of samples needed is the use of maps that have prior information on the spatial variability of the soil, allowing the identification of representative sampling points in the field. Management Zones (MZs), a sub-area delineated in the field, where there is relative homogeneity in yield potential, due to similar soil nutrients and environmental effects caused by similar landscape or soil conditions, has been widely accepted in management systems willing to apply precision agriculture techniques. MZs can be delineated using soil apparent electrical conductivity and local relief conditions variability, based on the clustering algorithm. Once the MZs relatively similar in terms of soil characteristics are created, there are no longer the need to take many soil samples to characterize the field. However, remains the question: where the best places are to take a limited number of soil samples within the MZ, where the mean value obtained from these samples represents the overall mean value of the attribute corresponding to the entire zone. This study aimed to evaluate a methodology to define the best locations for soil sampling to represent with a proper resolution of the physic-chemical soil variability. The best location for each soil sample was defined using the Fuzzy C-Means algorithm with some modifications. To evaluate the performance of the propose methodology, one field of 93 ha of sugarcane, was used to delineate the management zones and select the best places for soil sampling. The results obtained using 1 sample per 3 hectares by a guided sample grid was compared with the regular soil sampling, which uses, approximately, 1.3 sample per hectare, evaluating the soil physic and chemical attributes. The results show that, with this approach, it is possible to create targeted sampling grids with high precision for specific nutrients management,*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Suggested citation: Cook, S., Lacoste, M., Evans, F., Ridout, M., Gibberd, M. & Oberthür, T. (2018). An On-Farm Experimental philosophy for farmer-centric digital innovation. In Proceedings of the 14th International Conference on Precision Agriculture. Montreal, CA: International Society of Precision Agriculture.

reducing costs and increasing the sustainability.

Keywords. *Soil apparent electrical conductivity, soil sampling, Fuzzy C-means.*

Introduction

Soil fertility and crop spatial conditions mapping is one of the main procedures to ensure more sustainable production of sugarcane. Intrinsically related to Precision Agriculture (PA), this mapping consists in a detailed soil and yield sampling using modern equipment and techniques (Bullock et al., 2007). The basic principle is related to physical and chemical results of the laboratory analyzes with the geographical positions of the sample. Mapping soil spatial variability is essential for PA management and decision making to efficient agronomic practices to increase profitability of production. However, to ensure a precise mapping, a dense sampling is required; turn the activity sometime unfeasible. To overcome this challenge, several researches for soil sampling improvements have been made (Coelho et al., 2009; Machado et al., 2004), mainly due to answer the question of “what is the more suitable sample grid” to quantify the spatial variability of soil attributes.

One solution to this impasse has been the generation of management zones, defined as a sub-region of the field that presents a combination of limiting factors of productivity and soil attributes are considered similar (Tagarakis et al., 2013). Since each sub-region can be treated differently from the point of view of sampling and management, optimize the use of these resources can be more sustainable (Zhang et al., 2013).

Despite the considerable number of researches indicating the feasibility of using MZs (Bazzi et al., 2015, 2013; Moral et al., 2010), the division of an area in MZs is not a simple task considering that several attributes may have influence on crop yield. Among the variables identified in the literature as potential to generate MZs are elevation (Bazzi et al., 2015; Farid et al., 2016; Peralta et al., 2013; Schenatto et al., 2016), soil electrical conductivity and soil texture (Farid et al., 2016). After delineation of the MZs, the number of samples needed to delineate the field soil variability can be reduced and may vary according to some landscape characteristics.

Based on this context the aim of this study was to evaluate a methodology to define the best locations for soil sampling to represent with a proper resolution the physic-chemical soil variability within a management zone.

Methodology

Experimental field

The experiment was carried out in a commercial sugarcane site located in Ipiranga Mill, São Paulo State, Brazil (47°44'11.29"W 21°49'04.10"S, Fig. 1). The climate in this region is tropical to subtropical, and the mean annual rainfall and temperature are 1560 mm and 22.9 °C, respectively. The soil is classified as a clayey Typic Hapludox, and the clay fraction is dominated by kaolinite and iron and aluminum oxyhydroxides.

Fig 1. Experimental Field.

Soil data

According to (Doerge, 1999), management zone creation should utilize stable data (i.e., static properties that do not change from year to year). For this study digital elevation and the soil apparent electrical conductivity (ECa) was used as layers for the management zones.

ECa data were obtained at two soil depths (0.5 and 1.0 m) using EM38-MK2® (Geonics, Mississauga, Ontario, Canada) connected to a DGPS receiver (Ag142 ™, Trimble Navigation Ltd., Sunnyvale, CA, USA) and pulled by a quadricycle, since they have a good relationship with soil texture characteristics (Sudduth, et al., 2005) and have proven useful for delineation of management zones (Peralta and Costa, 2013; Moral et al., 2010). ECa was measured continuously, in every five rows interval, and later interpolated by kriging.

The attributes clay content, organic matter (OM) and the cation exchange capacity (CEC) were assessed by a sample grid which was used to obtain the best places to collect the elevation and ECa samples. The clay, OM and CEC soil attributes have directly impacts into the spatial and temporal variability of sugarcane yield, and their sample grid was used to see how efficiently this sample grid could be to elevation and ECa too. For this study, only the soil surface layer data (0.00 to 0.20 m) were evaluated.

Statistical and geostatistical analysis

The ECa and soil attributes data were analyzed to remove discrepant values from laboratory errors and field readings. Any input value that deviated from the mean by more than three standard deviations (for a given attribute) was treated as an outlier. To obtain the spatial variability maps of the evaluated attributes in the guided soil sampling (1 sample 3 ha^{-1}), the data were interpolated using ordinary kriging (OK). In the variogram setting, data interpolation was achieved by cross-validation to select the model (exponential, Gaussian or spherical) that best adapted the data and produced the smallest errors. Finally, we validated the results by soil samples from the original sampling grid (excluding the points used in the guided soil sampling). The predict values by OK were compared with the observed values.

Clustering

The SDUM (Software to Definition Management Units) (Bazzi et al., 2013) was used to create management zones via the Fuzzy C-Means method. To create the management zones, was used the digital elevation and ECa of two depths (0.5 and 1.0 m) data interpolated by OK with pixels in an area of 2 x 2 m and 10 neighbors.

The interpolated data was inserted into SDUM as samples transformed to a grid of 50x50 pixels using the software ArcMap 9.3 (to decrease the number of lines of data). After this, the management zones were generated with 2, 3, 4 and 5 classes, utilizing the Fuzzy C-Means algorithm selecting error parameter equal to 0.0001, and weight index equal to 1.3.

To choose the properly number of management zones to represent the field was used the Fuzzy performance index – FPI (Equation 1, Fridgen et al., (2004) and Modified partition entropy index – MPE (Equation 2, Boydell and McBratney, (2002).

$$FPI = 1 - \frac{c}{(c-1)}\left[1 - \sum_{j-1}^{n}\sum_{i=1}^{c}(u_{ij})^2/n\right] \quad (1)$$

where, c – number of clusters;

n – sample size in the whole area (number of observations);

ij u – element ij of the relevance Fuzzy matrix.

$$MPE = \frac{-\sum_{j-1}^{n}\sum_{i=1}^{c}u_{ij}\,(\log u_{ij})\,/n}{\log c} \quad (2)$$

where, c – number of clusters;

n – sample size in the whole area (number of observations);

ij u – element ij of the relevance Fuzzy matrix.

Selecting the best place to take soil samples

To obtain the best locations for soil sampling, were implemented procedural functions with PL/pgSQL language at the PostgreSQL object-relational database management system, using the free spatial extension PostGIS. The functions implemented used the Fuzzy C-Means algorithm with some modifications, which minimizes the sum of squares of errors within each class following some criteria and the data are grouped iteratively to the nearest class using the minimum distance criterion.

The best places were determined using the lower values of index FPI and MPE. For visualization of the results, was used the open source software Quantum GIS, which is a multiplatform geographic information system that allows visualization, editing and analysis of georeferenced data.

Results and discussions

The sampling grid used for soil data collection (clay, organic matter, and cation exchange capacity) has 119 sampling points, with a density of, approximately, 1.3 points ha^{-1} (Fig. 1).

Fig 1. Sample grid used to collect clay, OM and CEC.

From the sampling grid, the definition of management units defined by the attributes of elevation and ECa measured in depth of 0.5 m and 1.0 m was applied. To obtain the non-sampled points, the data were interpolated using the geostatistical method of Ordinary Kriging. With the data interpolated, the area was divided into areas of management, using the definition of two, three, four, and five classes by the Fuzzy C-means method in the SDUM software. The best definition of the area in management zones was the division of two classes, which represented relative efficiency greater than 1 (Table 1), what is considered a valid division by the SDUM software statistics and which has the lower FPI and MPE indexes in relation to the other divisions (3, 4 and 5 classes). These indexes were used to choose the best division, and how closer they are to 0, the better is the division of the area into management zones (Table 2).

Table 1. Results of the Evaluation of Management Units of two classes through the software SDUM.

Class	% Field	N° Samples	Relative Efficiency
1	0.45	1777	1.023
2	0.55	2092	1.023

Table 2. Result obtained from the division of the area into zones with two, three, four and five classes in the software SDUM.

Division into Management Zones	FPI	MPE
Two classes	0.1332	0.0250
Three classes	0.1918	0.0386
Four classes	0.2050	0.0431
Five classes	0.2219	0.0487

It is possible to visualize that 45% of the area is about class 1 of the management zones and the others 55% of the area is about class 2. The relative efficiency value is 1.023 (greater than 1), Table 1. The division into two classes has better delimited classes, lower FPI values, and it has better definitions of classes that have possible sizes to do the localized management, lower MPE value, Table 2 and Fig. 3.

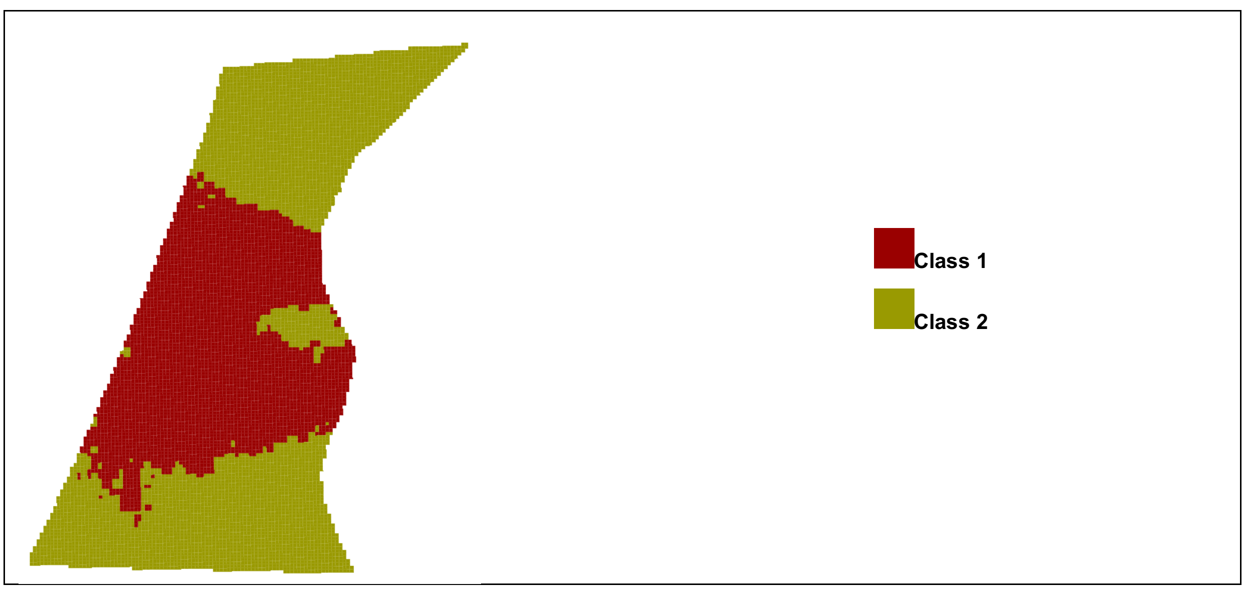


Fig 3. Result of the best division of MZs (two classes), obtained through the software SDUM with the variables of elevation and electrical conductivity of the soil of 1.0 and 0.5 m.

With the division of the area into two classes and the sample grid of 118 points, was applied the algorithm to select the best points to be used as soil sampling grid as an oriented grid. The density used for the algorithm choose samples was 1 point per 3 hectares, which represents 14 points in class 1, and 17 points in class 2 (Fig. 4).

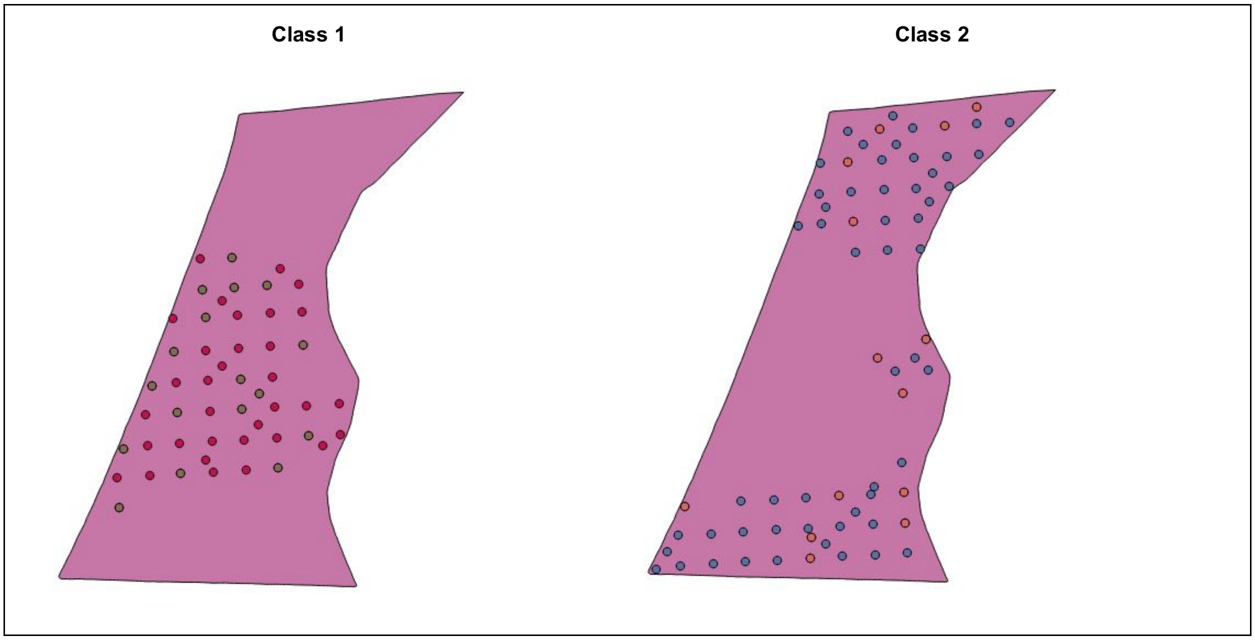


Fig 4. Selected points in brown (class 1) and in pink (class 2).

The histograms of the soil attributes: clay, organic matter, cation exchange capacity, was used to verify the frequency obtained from sampling grid used (Figs. 5, 6 and 7) and to verify the frequency about the selected points by the algorithm obtained from the sample grid using the density of 1 point 3 ha^{-1} (Figs. 8, 9 and 10).

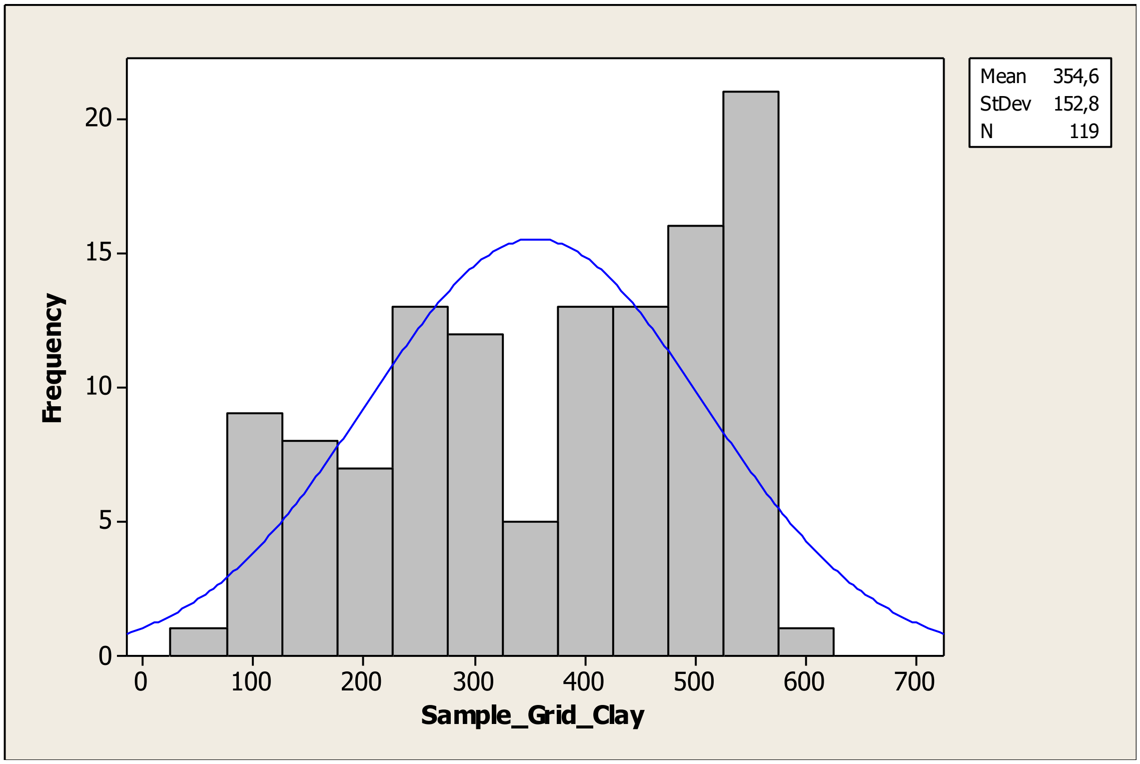


Fig 5. Histogram of of Clay soil attribute using the sample grid with 1.3 points/ha density.

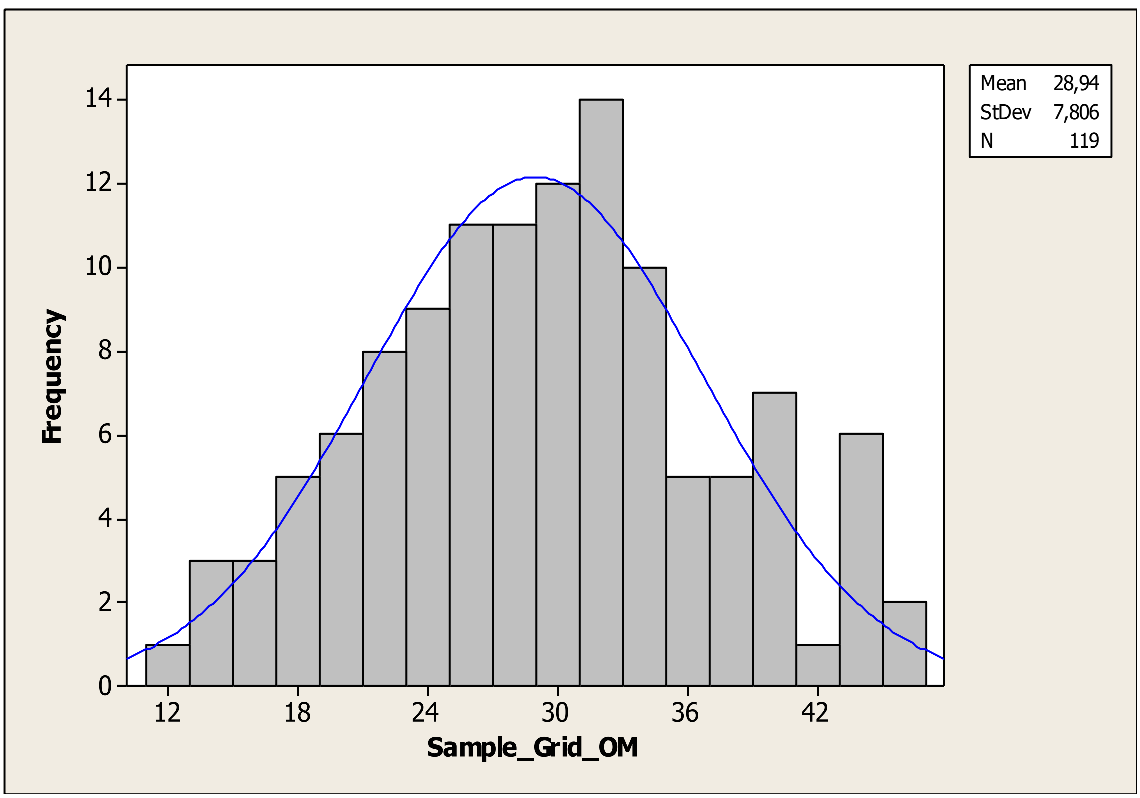


Fig 6. Histogram of OM soil attribute using the sample grid with 1.3 points/ha density.

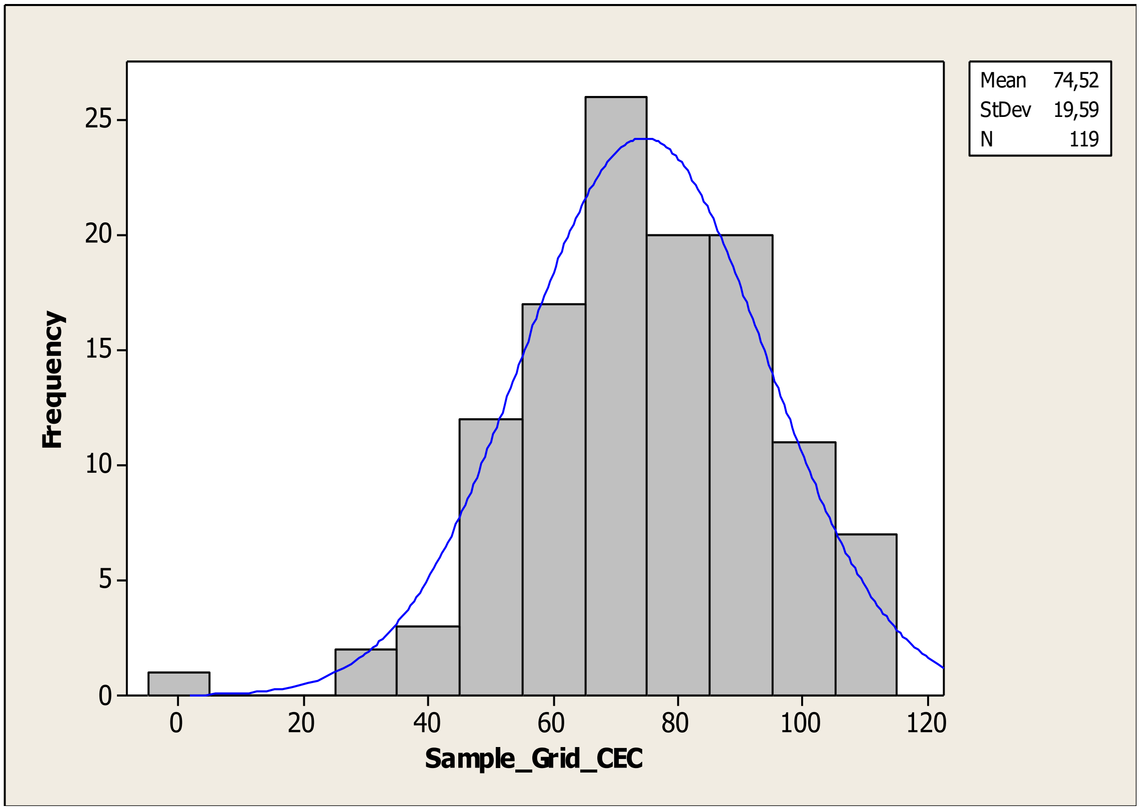


Fig 7. Histogram of CEC soil attribute using the sample grid with 1.3 points/ha density.

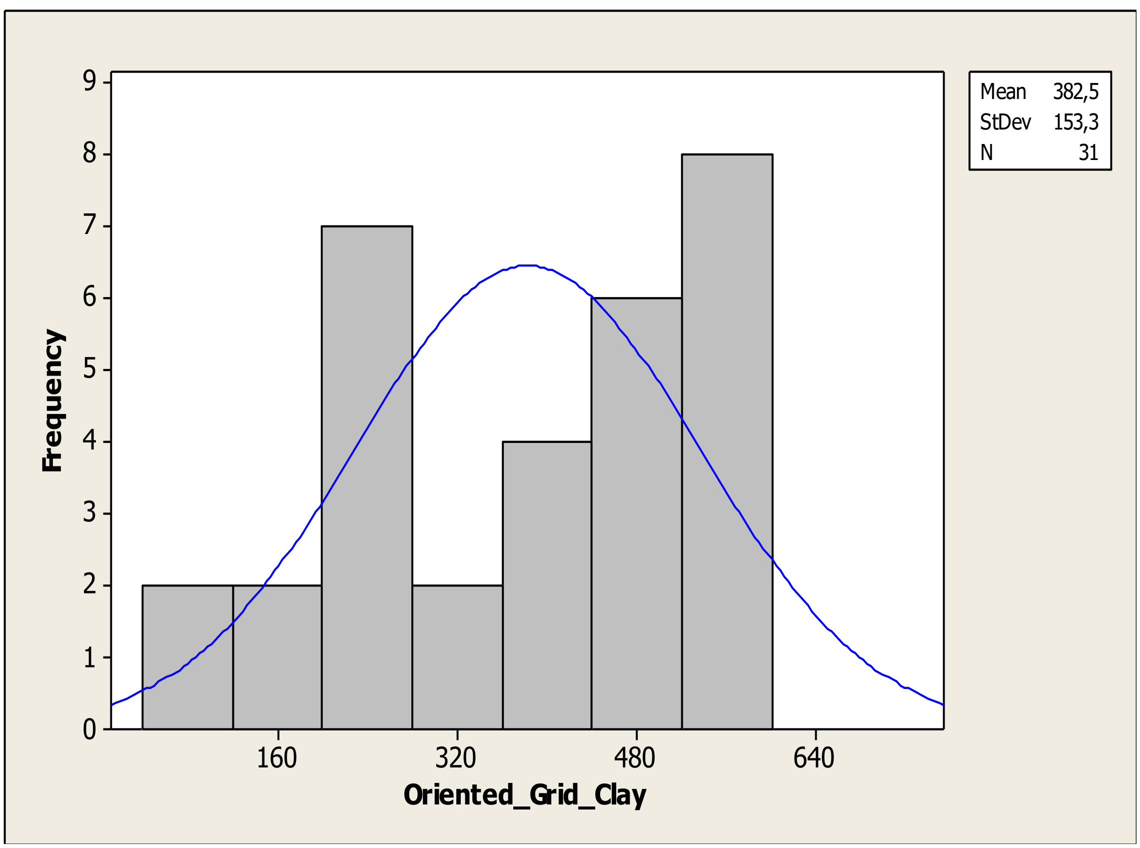


Fig 8. Histogram of Clay soil attribute using the oriented grid by the algorithm with 1 point/3ha density.

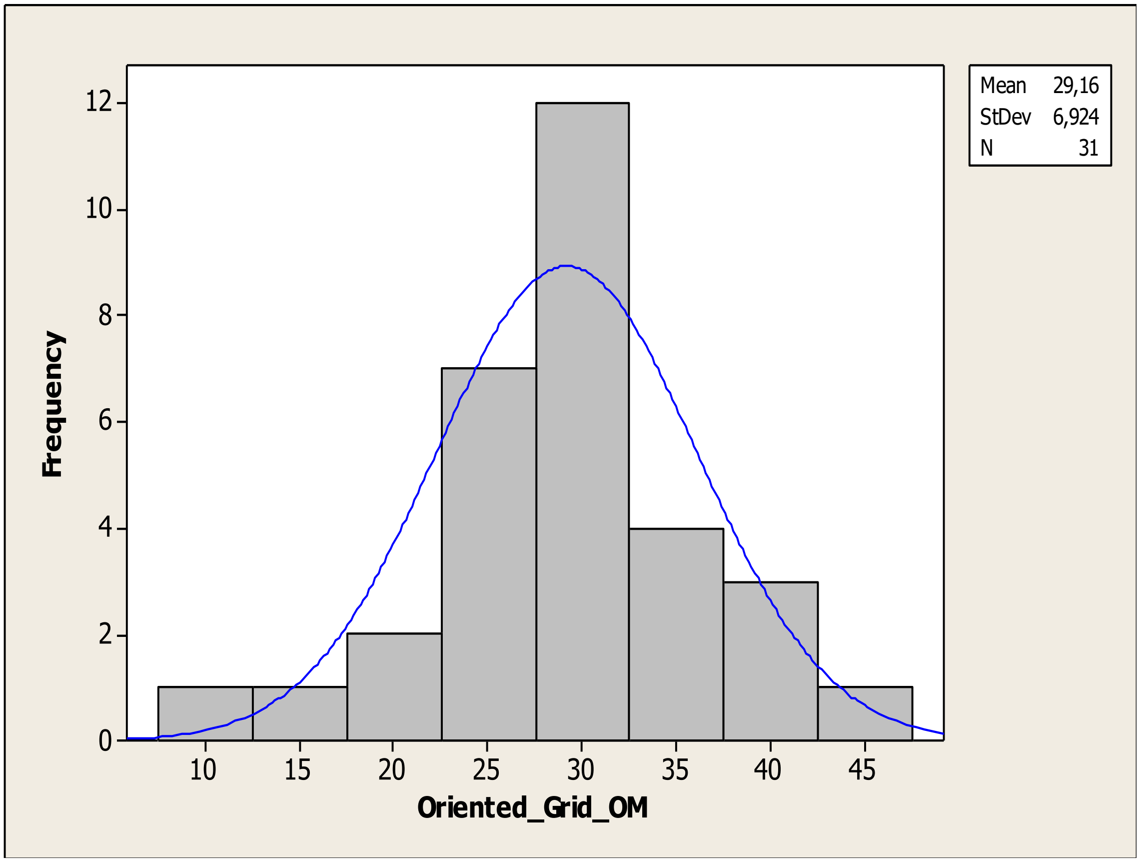


Fig 9. Histogram of OM soil attribute using the oriented grid by the algorithm with 1 point/3ha density.

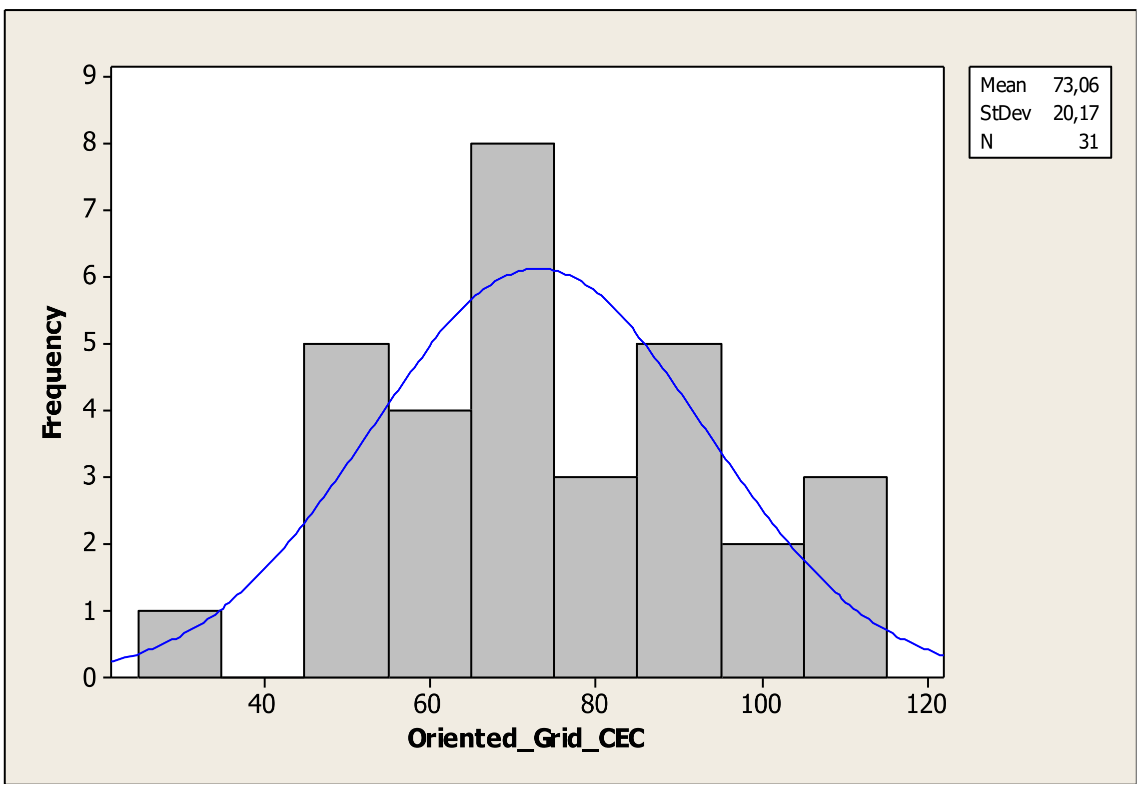


Fig 10. Histogram of CEC soil attribute using the oriented grid by the algorithm with 1 point/3ha density.

Clay histogram (Figures 5 and 8), showed an asymmetry with two peaks, which means that it has two frequencies higher than the others, indicating that the data collected have very different conditions, being necessary the division into two classes to manage the area. The OM histograms (Figures 6 and 9) showed a symmetrical distribution with a good fit.

In the CEC histogram (Figures 7 and 10), there are outliers, which are possibly data values that are very distant from the other data values. The distribution in the histogram has isolated bars at the left end of each, which identifies the outliers in both grids (original and oriented grid).

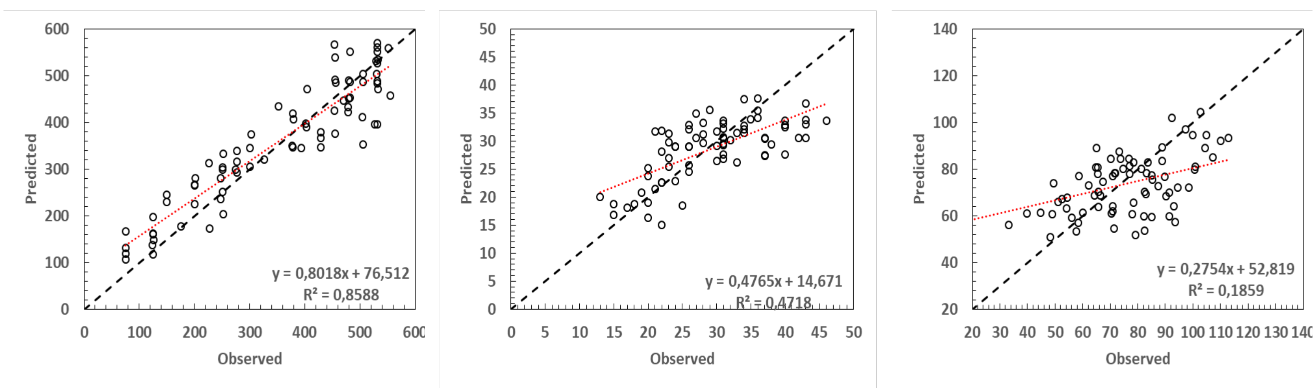
Then, all the histograms for the sample grid (original grid) and oriented grid (grid obtained by the selection places algorithm) shows the similarity frequency distribution for each attribute analyzed, and with very close means values. This represents that the sample/original grid and the oriented grid has similar representation of the total area, and that the oriented grid is a good representation of the sample grid, which allows the manager to use less sample points to do the localized management for the clay, organic matter and cation exchange capacity attributes.

The guided soil sampling was able to reflect the clay content variability in the experimental field (Fig. 11 - a) with high accuracy (R^2 = 0.86). On the other hand, the guided sampling did not satisfactorily reflect the OM and CEC variability (Fig. 11 - b and c, respectively).

The fact that the guided sampling was produced by the ECa and elevation attributes evidences the high correlation of these attributes with the clay content, as reported by (Moore et al., 1993). The variability of physical soil attributes is due to the slope changes that alter pedogenic processes, transport and storage of water in the soil profile (Sanchez et al., 2009), not occurring the same for OM and CEC.

The results show that new landscape attributes should be used for a more precise sampling orientation to be able to reflect the chemical attributes of OM and CEC. The slope and landscape curvatures, like proposed by (Valeriano and Rossetti, 2012), can be a good option to use in soil sampling process. Even more, the ECa can be aid the interpolation methods, as an ancillary information, to obtain better results of the soil variability maps of OM and CEC, as showed by (Sanches et al., 2018).

Despite the results not showed good predictions for OM and CEC, the landscape parameters are a source of information (economically feasible and easily assessed) with great potential to map the clay content variability and aid the site-specific management.



(a) Clay Content **(b) Organic Matter** **(c) Cation Exchange Capacity**

Figure 11. Observed and predicted values of the clay content (a), organic matter (b) and cation exchange capacity (c).

Conclusion

From the original soil sample grid, which reflected the clay content variability in the experimental field with high precision, the algorithm was applied to select the best locations to perform the soil sampling of the elevation attributes and ECa, which are stable attributes and therefore were used to generate management zones, which were better delineated with two classes.

By means of the original sample grid, the oriented grid was obtained and in an analysis of the frequency distribution, it was verified that although OM did not correlate with the attributes of elevation and ECa, we obtained symmetrical distributions between the guided/original grid and oriented grid. As the oriented grid was created with the elevation and ECa layers, it can be concluded through the frequency distributions that both in the guided and original grid sampling, better results can be obtained for a better total representation of the soil data if use more soil attributes.

Thus, the oriented grid could be used to collect the elevation, ECa and clay data as well, because clay was strongly correlated with elevation and ECa. The OM-oriented grid did not have much discrepancy in relation to the original grid, although it correlated little with the attributes used to generate the oriented grid.

Therefore, it is recommended to use more attributes for oriented grid generation to obtain better soil variability maps for OM and CEC. However, with the oriented grid it is possible to perform smaller soil samples and to obtain efficiency as good as the already used grid contributing to reduce costs and increase the sustainability.

Acknowledgements

The authors thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; action 2015/01587) for providing the necessary support.

.

References

Bazzi, C.L., Souza, E.G., Betzek, N.M., 2015. SDUM Software para definição de unidades de manejo: Teoria e Prática, First. ed. PGEAGRI, Cascavél.

Bazzi, C.L., Souza, E.G., Uribe-Opazo, M.A., Nóbrega, L.H.P., Rocha, D.M., 2013. Management zones definition using soil chemical and physical attributes in a soybean area. Eng. Agrícola, Jaboticabal 34, 952–964.

Boydell, B., McBratney, A.B., 2002. Identifying potential within-field management zones from cotton-yield estimates. Precis. Agric. 3, 9–23. doi:10.1023/A:1013318002609

Bullock, D.S., Kitchen, N., Bullock, D.G., 2007. Multidisciplinary Teams: A Necessity for Research in Precision Agriculture Systems. Crop Sci. 47, 1765–1769.

Coelho, E.C., Souza, E.G. de, Uribe-Opazo, M.A., Pinheiro Neto, R., 2009. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. Acta Sci. - Agron. 31, 165–174. doi:10.4025/actasciagron.v31i1.6645

Doerge, T.A., 1999. Management Zone Concepts. Site-Specific Manag. Guidel.

Farid, H.U., Bakhsh, A., Ahmad, N., Ahmad, A., Mahmood-Khan, Z., 2016. Delineating site-specific management zones for precision agriculture. J. Agric. Sci. 154, 273–286.

Fridgen, J.J., Kitchen, N.R., Sudduth, K. a, Drummond, S.T., Wiebold, W.J., Fraisse, C.W., 2004. Management zone analyst (MZA): software for management zone delineation. Agron. J. 96, 100–108. doi:10.2134/agronj2004.1000

Machado, P.L.O.A., Bernardi, A.C.C., Valencia, L.I.O., Molin, J.P., Gimenez, L., SILVA, C.A., ANDRADE, A.G., MADARI, B.E., MEIRELLES, M.S.P., 2004. Mapeamento Da Condutividade Elétrica Do Solo E Relação Com Os Teores De Argila De Um Latossolo Vermelho Sob Plantio Direto De Grãos No Paraná. Congr. Bras. Agric. Precisão- ConBAP 1–10.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Attrib. Predict. Using Terrain Anal. 57, 443–452.

Moral, F.J., Terrón, J.M., Silva, J.R.M. Da, 2010. Delineation of management zones using mobile measurements of soil apparent electrical conductivity and multivariate geostatistical techniques. Soil Tillage Res. 106, 335–343.

doi:10.1016/j.still.2009.12.002

Peralta, N., Costa, J.L., Castro, M., Balzarini, M., 2013. MODELOS DE ELEVACIÓN DIGITAL Y PROFUNDIDAD DE SUELO. Interciencias 38, 418–424.

Sanches, G.M., Magalhães, P.S.G., Remacre, A.Z., Franco, H.C.J., 2018. Potential of apparent soil electrical conductivity to describe the soil pH and improve lime application in a clayey soil. Soil Tillage Res. 175, 217–225. doi:10.1016/j.still.2017.09.010

Sanchez, R.B., Marques Júnior, J., Souza, Z.M., Pereira, G.T., Martins Filho, M.V., 2009. Variabilidade espacial de atributos do solo e de fatores de erosão em diferentes pedoformas. Bragantia 68, 1095–1103. doi:10.1590/S0006-87052009000400030

Schenatto, K., Godoy, E. de S., Bazzi, C.L., Betzek, N.M., Gavioli, A., 2016. Two approach comparison to define crop management zones (MZs). African J. Agric. Res. 11, 3680–3689. doi:10.5897/AJAR2016.11453

Tagarakis, A., Liakos, V., Fountas, S., Koundouras, S., Gemtos, T.A., 2013. Management zones delineation using fuzzy clustering techniques in grapevines. Precis. Agric. 14, 18–39. doi:10.1007/s11119-012-9275-4

Valeriano, M.M., Rossetti, D.F., 2012. Topodata: Brazilian full coverage refinement of SRTM data. Appl. Geogr. 32, 300–309.

Zhang, Z., Lü, X., Lv, N., Chen, J., Feng, B., Li, X.W., 2013. Defining agricultural management zones using Gis techniques: Case study of Drip-irrigated cotton fields. Inf. Technol. J. 12, 6241–6246.