RMAPS: AN INTEGRATED TOOL TO DELIMITATE HOMOGENEOUS MANAGEMENT ZONES

Edwin Erazo-Mesa[1]; Carlos Mosquera[2]; Oscar Ochoa[2]

1 Escuela EIDENAR, Facultad de Ingeniería, Universidad del Valle, Calle 13 # 100-00, Edificio E39, Postal code: 760032, Cali, Colombia. E-mail: osvaldo.erazo@correounivalle.edu.co

2 AgroAP CI. Carrera 100 # 11-60 Torre Valle de Lili – Oficina 819, Cali, Colombia. E-mail: carlosmosquera@agroap.com; oscar.ochoab@gmail.com

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

Abstract.

We present rmaps, an R package that integrates soil and crop yield spatial variability using geostatistical methods and one-hidden-layer perceptron (OHLP) modeling, to identify how input parameters influence crop yield and delimit homogenous zones. rmaps were tested using three synthetic datasets and one sugarcane dataset. The synthetic datasets consisted of 21 randomized, linear, and random-linear parameters with 2000 samples, 20 used as inputs and the remaining one as the output parameter. The sugarcane dataset consisted of 54 soil samples where physical and chemical properties were measured in the laboratory, drainage was used as a binary parameter, and yield data were collected from 2015–2016. The results of the synthetic dataset showed that the OHLP, implemented in rmaps, could identify the linear relationship between the input and output parameters when these parameters are linearly related ($R^2 = 0.912, p < 0.001$). In contrast, rmaps could not identify any relationships among parameters for the randomized dataset ($R^2 = 0.001, p > 0.05$). Regarding the sugarcane dataset, drainage was the parameter that mainly explained changes in sugarcane yield with a relative importance, compared to the remaining input parameter of 26.722% ($p < 0.05$). The relevance metric map showed homogeneous zones where drainage and soil properties can be managed differentially to increase sugarcane yield. We conclude that rmaps permit the identification of relevant input parameters for improving crop yields and displaying them in homogeneous management zones.

Keywords.
Artificial intelligence; artificial neural networks; R package; precision agriculture; relevance metric maps

Introduction

The use of management zones (MZs) is one of the most studied methods used in precision agriculture (PA) to optimize crop yield and address soil spatial variability. These are defined as subareas within a field with similar soil, plant, topographic, and climate characteristics in which PA practices are carried out in an uninform manner (Guastaferro et al., 2010; Nawar et al., 2017). MZs vary in number per field, shape, and extent, from field to field, and throughout time. An MZ can cover a few square meters to several hectares. Fertilization, irrigation, seeding, and weed control practices are carried out by zone. Although ancient farmers fractioned fields according to their fertility, this method was developed from the opportunity to use geospatial technology to the traditional uniform management applied in agriculture before the PA (Robert, 2002).

Considering that soil, plants, topography, and climate interact in an unknown manner to return a crop yield, parameters from these factors are usually measured and estimated to delimit MZs. Some approaches are focused on soil properties, others are focused on plant parameters, with yield being the primary parameter, and the remaining approaches are focused on a combination of these two parameters (Ortega & Santibáñez, 2007). These parameters, measured with sensors or manually in the field, processed in the laboratory and spatialized with geostatistics, are sorted in georeferenced layers and incorporated into empirical or stochastic models. Among the models reported to delineate MZs are cluster (k and c-means), Rasch, coefficient of variation, and principal component models (Gavioli et al., 2019). The collection, filtering, selection, grouping, and mapping of MZs are required (Santos & Saraiva, 2015).

Because of its ability to provide reasonable responses to highly complex problems, artificial intelligence is broadly used in PAs (Shaikh et al., 2022). Regarding MZs, Von Hebel et al. (2021) used the ML technique fuzzy c-means to delineate two MZs in six potato fields using the NDVI and apparent electrical conductivity. The authors found that the yield did not significantly differ between the two zones. Gallardo-Romero et al. (2023) found that the machine learning (ML) algorithms used to predict three MZs in corn were accurate. Gallardo-Romero et al. (2023) arbitrarily defined the number of MZs based on the high, medium, and low yield values. Ohana-Levi et al. (2019) analyzed the influence of soil, plant, and topographic parameters on crop yield to define MZs. Using the weights of each input parameter, Ohana-Levi et al. (2019) delineated four MZs that are not associated with an agronomical task in the field. (Bai et al., 2023) used a multiscale segmentation method to delineate salinity management zones. This study also lacks specific recommendations for salinity management for each delimited zone.

According to the above studies, we determined that MZs do not integrate the effects of soil, topography, and other influential factors on crop yield; the number of MZs, usually three, is defined by the authors; the most commonly used methods for delineating MZs are traditional (k-means) and ML clustering (fuzzy c-means); and each MZ zone in the map is represented by a single parameter, usually yield, vegetation index, or nitrogen, hindering several tasks at the same time. Erazo et al. (2015) proposed a method to identify limiting crop factors using the internal structure of a trained artificial neural network (ANN) and spatialized this result through relevance metric maps. These maps show homogeneous areas that can be analogized to MZs and fill the mentioned gaps. This study develops the methodological framework presented by Erazo et al. (2015) in a set of R functions compiled in the R package rmaps.

Methods and data

Conceptualization and methodological framework of rmaps

Crop yield can be seen as the result of interactions among soil, plants, topography, agronomic tasks, and climate factors. Each of these factors is composed of parameters such as bulk density, soil organic matter, NDVI, slope, or fertilization, which influence the yield to a variable degree. In an agricultural field, the climate is relatively uniform, and several agronomic tasks are assumed to be homogeneous, by which soil, plant, and topographic parameters are spatially variable. Information and communication technologies and spatial soil sampling permit accounting with digital and successive layers composed of pixels with values of these parameters. In a particular pixel, crop yield is a function dependent on soil, plant, and topographic parameters.

Relevance Metric Maps (rmaps) is an R package for analyzing soil, topography, climate, plant, and crop yield spatial variability and delimitating homogenous zones. rmaps are based on the relevance metric methods reported in Satizábal & Pérez-Uribe (2007), and the spatialization of these methods is based on the use of training and testing datasets of ANNs reported by Zurada et al. (1997). rmaps perform four main processes: geostatistical analysis, multilayer perceptron training, relevance metric analysis, and relevance metric mapping. Geostatistical analysis is composed of eight functions, and the other processes implement one function (Table 1). Geostatistical functions can be run sequentially through function computemaps or individually. The R maps were built in R (version 4.4.0) via the sf, gstat, and neuralnet packages.

Table 1. Description of the functions of the rmaps package

Process	Function	Description
Geostatistics	checkdbase	Checks structure of georeferenced database
	transcoord	Transforms geographical or projected coordinates
	duplicates	Removes duplicates points from georeferenced database
	outliers	Identifies and remove spatial outliers from georeferenced database
	statistic	Computes descriptive statistics
	tr	Transforms and back-transform numerical parameters
	experimental	Computes experimental semivariogram
	theoretical	Computes theoretical best-adjusted model of semivariogram
	Interpolate	Interpolates spatial data using Kriging or IDW methods
Multilayer perceptron training	architecture	Trains and finds the best architecture of artificial neural network
Relevance metric analysis	relevancemetric	Computes relevance metric based on artificial neural network architecture
Relevance metric maps	relevancemaps	Computes relevance metrics maps

From georeferenced data with soil, plant, management, climate, and crop yield parameters, as shown in Fig. 1, rmaps transforms geographic to projected coordinates; identifies spatial duplicate and outlier samples; computes statistical descriptive parameters; identifies and removes (when applicable) first- and second-order spatial trends; models the spatial correlation structure (variogram); and interpolates and performs cross-validation. rmaps trains a one-hidden-layer perceptron (OHLP) and selects the number of hidden neurons with the lowest training error on the dataset resulting from spatializing with geostatistical input and output parameters. Then, perturbance methods applied to the trained OHLP weights and associated with interpolated pixels produce relevance metric maps with positive and negative impacts of the input parameters on the crop yield. These relevance metric maps, homologated with homogeneous management zones, contain the spatial distribution of the most influential input parameters on crop yield.

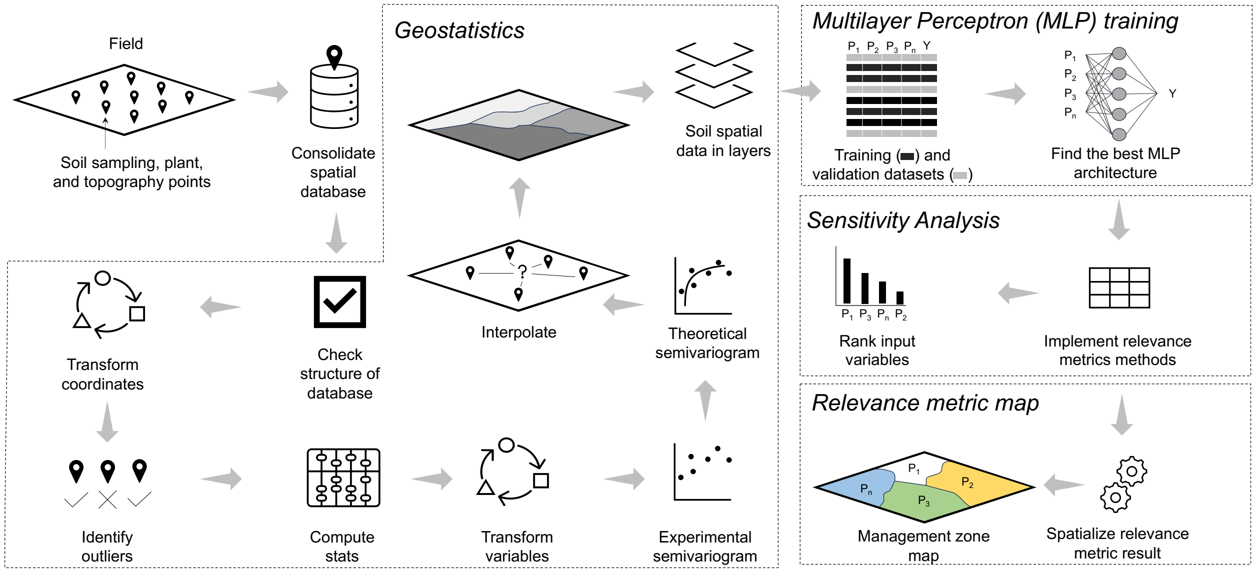

Fig 1. Methodological framework of rmaps.

The geostatistical framework of the rmaps is described in Bivand et al. (2008). The ANN used for rmaps a is a feedforward multilayer perceptron with a backpropagation training algorithm available in the neuralnet R package through the function neuralnet (Fritsch et al., 2019). Although rmaps is designed to train and test multilayer perceptron nets, in this study, an OHLP was implemented because relevance metric methods apply for a one-hidden-layer perceptron. Among the limitations of rmaps are that their implementation depends on georeferenced samples in the field, which impedes the direct addition of a raster layer to remote sensing products. To simultaneously perform geostatistical analysis for several parameters, these parameters must be sampled in a unique spatial sample grid. rmaps are focused on analyzing the spatial variability of crop factors per crop cycle, i.e., they do not include a spatiotemporal analysis of the crop. Moreover, the use of the OHLP for MZ computation restricts the use of more hidden layers, weakening the model accuracy in some cases.

Data

Synthetic datasets

Synthetic datasets were created to verify the robustness of the architecture and relevancemetric rmaps functions for identifying correlations between inputs and output parameters. We expected to find a weak correlation between the inputs and output parameters for the random dataset. For linear and linear-random relationships, a strong correlation is expected to be found with all and with one input parameter, respectively. These datasets were built in R with 21 parameters and 2000 samples each, of which 20 parameters corresponded to inputs and the remaining parameter corresponded to the output parameter.

Samples for the 20 input parameters for all datasets and the output parameter of the random dataset were randomly extracted from the uniform distribution varying between -1 and 1. The output samples of the linear dataset resulted from the sum of each of the input samples plus a random value from the uniform distribution between –1 and 1. The output parameter for the linear-random dataset corresponded to the 20th parameter multiplied by 2. Therefore, the number of input parameter samples for all datasets varied between –1 and 1, the number of output samples for the linear dataset varied between -9 and 9, and the number of output samples for the linear-random dataset varied between –2 and 2.

Sugarcane dataset

This consisted of 54 soil samples spatially distributed across 100 ha taken at the sugarcane Farm Churimal (Valle del Cauca, Colombia) in 2019. Three soil physical and nine chemical properties were measured in the laboratory as follows. The percentages of sand, silt, and clay, pH, percentage of organic matter (OM), phosphorus (P, meq 100 g^{-1}), calcium (Ca, cmol kg^{-1}), magnesium (Mg, cmol kg^{-1}), potassium (K, cmol kg^{-1}), cation exchange capacity (CEC, cmol kg^{-1}), ratio of calcium magnesium (Ca/Mg), and ratio of calcium plus magnesium to potassium [(Ca+Mg)/K] were determined. The presence or absence of drainage technology was also recorded in the plots as an input parameter, expressed in binary numbers (0 for absence and 1 for presence of drainage). The sugarcane yield during the 2015-2016 season was recorded with a yield monitor in units of t ha^{-1}.

rmaps testing

The function architecture trains an OHLP and selects the optimum number of neurons in the hidden layer. This function was tested by training 20 OHLP architectures and selecting the one with the lowest root mean square error (RMSE). The architectures consisted of 1 to 20 neurons in the hidden layer. Each architecture was trained 50 times, for a total of 1000 OHLPs trained per dataset. The synthetic datasets used to train the OHLPs were split 65% for training and 35% for testing. The first 20 columns and the last column were used as the inputs and outputs of the OHLP model, respectively. In the training and testing stages, the RMSE was computed, taking the observed data as the output parameter from the dataset and the simulated data as the parameter outputted from the trained OHLP. In addition to the RMSE, the coefficient of determination (R^2) was computed for testing.

Function computemaps were used to perform geostatistical analysis of spatial soil sampling and yield data for sugarcane crops in the Churimal Farm. Geostatistical analysis revealed spatial outliers, computed descriptive statistics, experimental and theoretical semivariograms, and interpolation with kriging or inverse distance weighting. The RMSE, mean error (ME), mean absolute error (MAE), standardized RMSE (RMNSE), normalized RMSE (RMSEr), and coefficient of correlation (r) were computed for cross-validation (Hengl, 2009). Once spatialized, OHLPs with 1 to 13 hidden neurons were trained and tested 100 times. The best OHLP architecture had the lowest RMSE and the highest R^2. The function relevance metric was used to compute the relative importance of each input parameter to the crop yield, and the Wilcoxon test was used to determine whether the input parameters significantly differed from the remaining parameters. The relevance metric results were spatialized through functional relevance maps to determine the effects of positive impact parameters on yield and to delimit homogeneous management zones. All computations and figures were coded in R. All values reported in intervals (± symbols) in the results refer to the average and uncertainty, computed with the mean and standard deviation.

Results and discussion

rmaps testing with synthetic datasets

The output parameter values for the random, linear, and linear-random datasets varied from –0.999 to 0.999, –8.664 to 8.955, and –1.999 and 1.999, respectively, with average values of 0.003, –0.065, and –0.037, respectively. The Kolmogorov–Smirnov test for each sample revealed that random and linear random data were obtained from a uniform distribution, and linear data were obtained from a normal distribution. Regarding the distribution of input data, ANNs have been demonstrated to accurately predict a variety of data distributions, noise, nonlinear problems, and learning from the probabilistic distribution of input data (Yuan et al., 2020).

The function architecture trained 1000 OHLP with random, linear, and linear-random datasets in 0.57, 0.12, and 0.12 h, respectively. The processing time of OHLPs depends on the difficulty of finding relationships between inputs and output parameters, as discussed by Shamir (2016). The architectures with the lowest RMSEs for the random, linear, and linear-random datasets were 1, 11, and 3 neurons in the hidden layer, respectively, with average RMSEs of 0.582 ± 0.003, 0.918 ± 0.141, and 0.127 ± 0.056, respectively. The RMSE as a function of the number of neurons in the hidden layer varied differently according to the dataset. For the random dataset, more neurons had a higher RMSE; for the linear datasets, more neurons had a lower RMSE of up to 11 neurons, from which the RMSE increased slightly; and for the linear-random datasets, more neurons had a lower RMSE of up to 3 neurons, from which the RMSE increased to 12 neurons and then remained relatively constant. Similar to that reported by D'souza et al. (2020), the OHLP architecture selected corresponds to an optimization process where the OHLP prediction accuracy can be improved or the processing time, which maintains the accuracy, can be reduced.

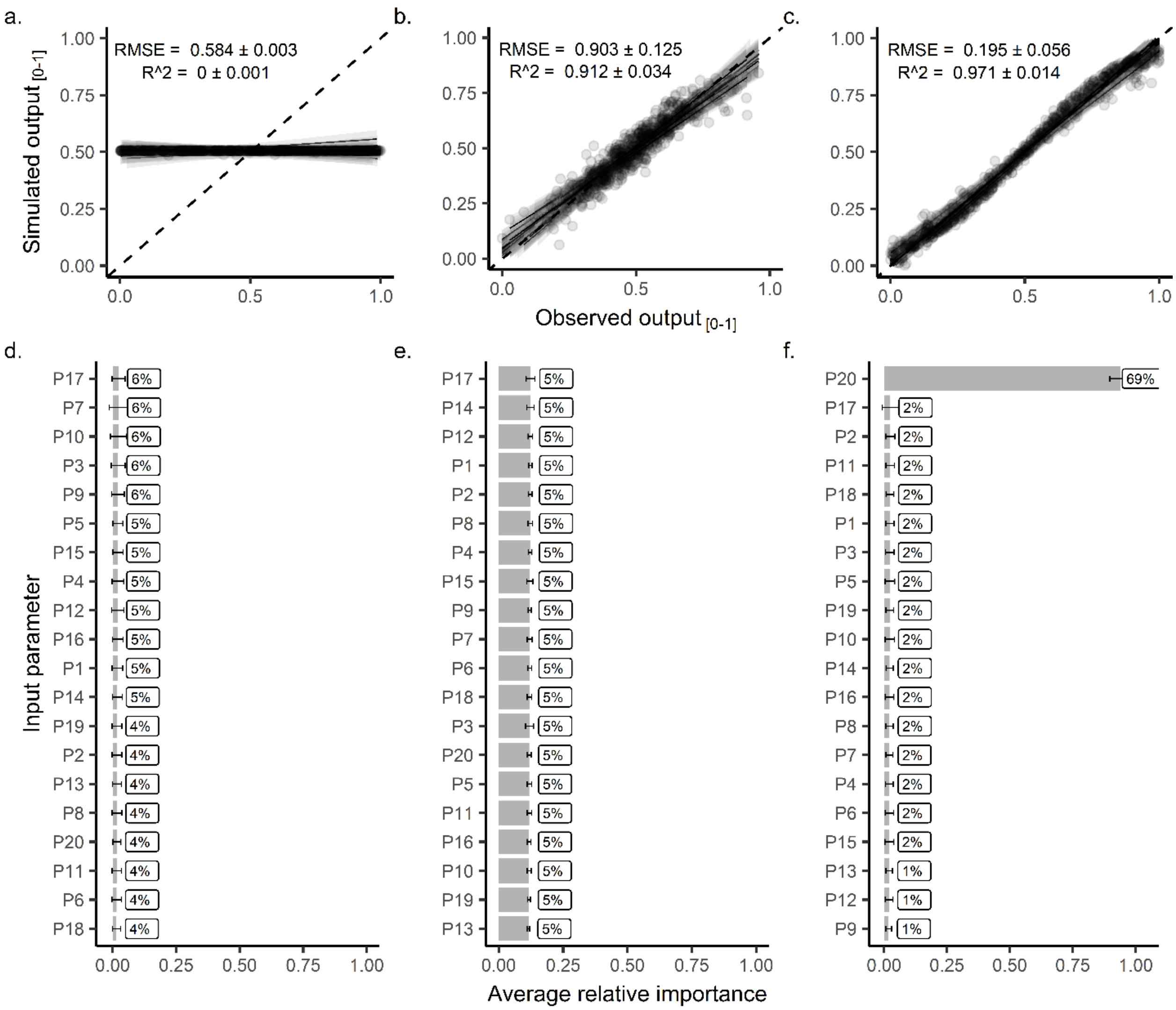

Fig 2. Performance of one-hidden layer perceptron testing and relevance metric analysis of inputs on output parameters with random (a, d), linear (b, e), and linear-random (c, f) datasets.

The testing and relevance metric analysis of the trained OHLPs for the three datasets are shown in Fig. 2. OHLP poorly predicted the output parameter in the random dataset, which was expected because of the null relationship between the inputs and the output parameters (Fig. 2a). This is the result of the low relative importance of inputs to the output in the random dataset (Fig. 2d). The accuracy and relative importance of inputs on output increased significantly in linear and linear-random datasets, where input parameters were strongly correlated with the output (Figs. 2b, e, c, and f). In the linear dataset, an R^2 of 0.912 and a relative importance of 0.120 indicate a strong correlation between the inputs and output (Figs. 2b and 2e). Because the output is the sum of all inputs, the relative importance was equally distributed among all inputs with the same percentage (Fig. 2e). In the linear-random dataset, the strong correlation between the inputs and the output (Fig. 2 c) is due to the output being a linear function of parameter 20 (P20), which was significantly different from the others according to the Wilcoxon test (Fig. 2f). The remaining parameters acted as noise in the dataset (Fig. 2f).

Although the above-described ANN capabilities are fully documented (Salah & Hannan, 2020), the results of the tested rmaps functions architecture and relevancemetric suggest that these were correctly parameterized, and internal processes such as the transformation and back-transformation of parameters between 0 and 1 (function tr), dataset splitting, the selection of the OHLP architecture, and relevance metric methods were properly coded. Regarding relevance metric methods, recent studies have reported how to extract information from the internal structure of multilayer neural networks (Jeczmionek & Kowalski, 2022), suggesting an opportunity to improve rmaps.

RMpas testing with the sugarcane dataset

Function computemaps can be used to geostatiscally analyze sugarcane dataset parameters. The function outliers removed 10% of the spatial outliers using the method described in Lu et al. (2003). The function statistics computed descriptive statistics for the parameters. The sugarcane yield in the Churimal for 2015-2016 varied between 105.25 and 120.38 t ha^{-1}, with an average of 112.97 t ha^{-1}, and followed a normal distribution. The circular and spherical semivariograms, fitted with theoretical functions, were better adjusted to the experimental semivariance for 11 of the 13 parameters. The ratio of nugget to total semivariance for pH was 0.865, suggesting weak spatial dependence, and the remaining parameter was lower than 0.272, indicating strong spatial dependence (Cambardella et al., 1994). Function computemaps were parameterized to generate spatially coincident maps of inputs and output parameters with a pixel size of 20 × 20 m.

Cross-validation of the interpolated parameters, computed through function interpolation, indicates a high correspondence between the interpolated and observed data for sand, silt, clay, OM, P, Ca, Mg, CEC, and Ca/Mg (Table 2). This contrasts with the findings for pH, K, (Ca+Mg/K), and yield, where the interpolation method weakly predicts the observed data in the field (Table 2). The spatial variability map shows that the yield was lower than 110 t ha^{-1} in the upper central part and greater than 115 t ha^{-1} from the central part downward. Although several factors explain the low interpolation accuracy in the context of soil and crop spatial variability (Lin et al., 2005; Peukert et al., 2012), we attribute weak interpolation predictions to the low number of spatial points sampled.

Table 2. Interpolation performance for the input and output parameters

Parameter	Method	Mean	RMSE	ME	MAE	RMNSE	RMSEr	r
Sand	Kriging	21.00	11.45	–0.37	8.97	0.17	72.52	0.69
Silt	Kriging	27.02	4.31	0.13	3.46	0.31	67.98	0.73
Clay	Kriging	53.94	11.22	0.35	9.09	0.09	69.88	0.71

Parameter	Method	Mean	RMSE	ME	MAE	RMNSE	RMSEr	r
pH	IDW	6.56	0.34	–0.01	0.28	-	97.02	0.20
OM	Kriging	2.36	0.46	–0.01	0.37	2.32	73.05	0.68
P	Kriging	16.35	6.16	–0.39	5.08	0.15	74.76	0.66
Ca	Kriging	13.09	2.28	0.08	1.76	0.57	70.54	0.71
Mg	Kriging	11.40	2.31	0.06	1.89	0.40	54.10	0.84
K	Kriging	0.48	0.12	0.00	0.09	8.16	92.37	0.36
CEC	Kriging	25.68	3.85	0.24	3.19	0.20	53.33	0.84
Ca/Mg	Kriging	1.28	0.18	–0.01	0.15	4.18	53.05	0.85
(Ca+Mg)/K	Kriging	55.12	13.60	–0.70	11.35	0.07	84.82	0.52
Yield	Kriging	112.97	3.11	–0.08	2.57	0.29	89.65	0.42

The OHLP training results indicate that the architecture with 13 neurons in the hidden layer predicted the lowest RMSE (0.853 ± 0.108 t ha^{-1}) and highest R^2 (0.853 ± 0.038) (Fig. 3a). The RMSE as a function of the number of neurons suggests that the greater the number of neural processing units is, the lower the error prediction, similar to that found by Çolak (2021). Relevance metric analysis of the sugarcane dataset indicated that drainage, silt, and sand had a strong influence on yield during the 2015-2016 crop season, with relative importance values of 26.722 ($p < 0.05$), 8.220 ($p < 0.05$), and 7.754%, respectively (Fig. 3b). This result agrees with that of Deng & Bailey (2020), who found that simulated drainage practices led to crop benefits.

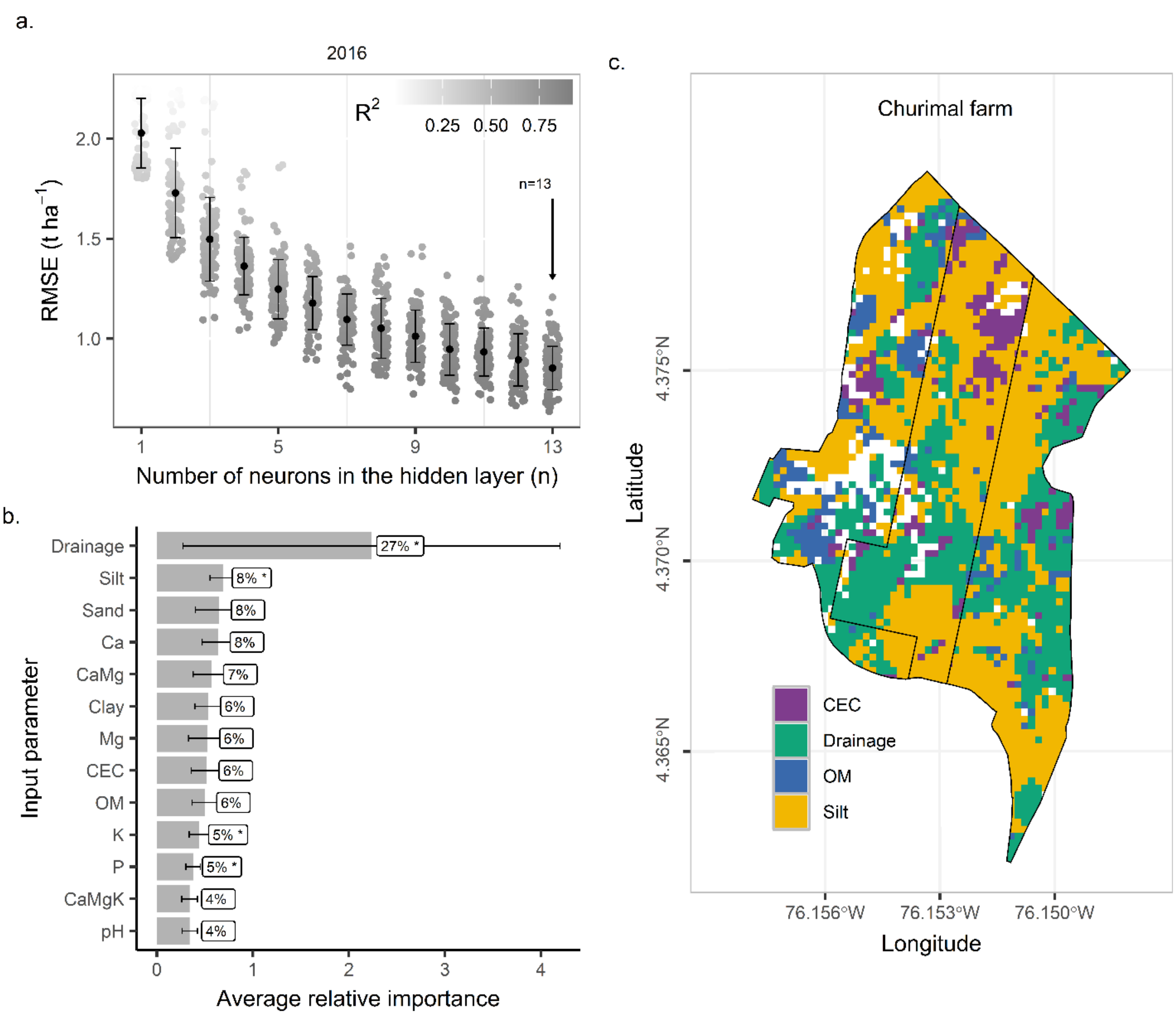

Fig 3. Performance of one-hidden layer perceptron training (a), relevance metric analysis (b) of inputs on output parameters and homogeneous management zones delimited with the sugarcane dataset (c).

The functional relevance maps were spatially distributed for the relevance metric analysis on a map with four zones (Fig. 3c). Silt (47.866% of the area), drainage (29.703%), CEC (7.853%), and OM (7.750%) occupied 97.172% of the area of the Churimal Farm. Silt and drainage were more spatially continuous than OM and CEC were (Fig. 3c). By extrapolating to the spatial context the interpretation of relevance metric methods (Satizábal & Pérez-Uribe, 2007), we affirm that improving drainage, increasing silt (which is complex but can be interpreted as balancing the other soil texture fractions), CEC, and OM in the zones shown in Fig. 3c could increase sugarcane yield.

Clustering is the method most commonly used in studies to delimit MZs (Gavioli et al., 2019). MZs delimited by clustering lack physical interpretation. We present a disruptive way to compute and interpret homogeneous management zones in precision agriculture because management zones spatialized with relevance metric methods are associated with input parameters that influence crop yield. The agronomic recommendations for the findings of our study are as follows: for drainage MZs, farmers could install drainage pipes and increase the maintenance of open channels, pumping, and land levelling (Castellano et al., 2019). For OM and CEC MZs, farmers can fertilize sugarcane crops with organic fertilizers, mulch, and other organic amendments without tillage (Chenu et al., 2019). For the silt MZ, farmers could add silt-rich stones to the soil to modify the soil texture (Zhang et al., 2020).

Conclusions

rmaps use the influence of input factors on yield crops to delimit homogeneous management zones in precision agriculture. Specific agronomical tasks can be recommended based on the use of rmaps to delimit homogeneous management zones, generating a potential improvement in yield for the following seasons.

Acknowledgments

We thank Dr. Daniel Jiménez R. (CGIAR), who, through his work on artificial intelligence in agriculture, encouraged the conceptual framework of relevance metric maps presented here.

References

Bai, J., Wang, N., Hu, B., Feng, C., Wang, Y., Peng, J., & Shi, Z. (2023). Integrating multisource information to delineate oasis farmland salinity management zones in southern Xinjiang, China. *Agricultural Water Management*, *289*, 108559. https://doi.org/10.1016/j.agwat.2023.108559

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Springer New York. https://doi.org/10.1007/978-0-387-78171-6

Cambardella, C. A., Moorman, T. B., Novak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F., & Konopka, A. E. (1994). Field-Scale Variability of Soil Properties in Central Iowa Soils. *Soil Science Society of America Journal*, *58*(5), 1501–1511. https://doi.org/10.2136/sssaj1994.03615995005800050033x

Castellano, M. J., Archontoulis, S. V., Helmers, M. J., Poffenbarger, H. J., & Six, J. (2019). Sustainable intensification of agricultural drainage. *Nature Sustainability*, *2*(10), 914–921. https://doi.org/10.1038/s41893-019-0393-0

Chenu, C., Angers, D. A., Barré, P., Derrien, D., Arrouays, D., & Balesdent, J. (2019). Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil and Tillage Research*, *188*, 41–52. https://doi.org/10.1016/j.still.2018.04.011

Çolak, A. B. (2021). A novel comparative investigation of the effect of the number of neurons on the predictive performance of the artificial neural network: An experimental study on the thermal conductivity of ZrO_2 nanofluid. *International Journal of Energy Research*, *45*(13), 18944–18956. https://doi.org/10.1002/er.6989

Deng, C., & Bailey, R. T. (2020). Assessing causes and identifying solutions for high groundwater levels in a highly managed irrigated region. *Agricultural Water Management*, *240*. https://doi.org/10.1016/j.agwat.2020.106329

D'souza, R. N., Huang, P.-Y., & Yeh, F.-C. (2020). Structural Analysis and Optimization of Convolutional Neural Networks with a Small Sample Size. *Scientific Reports*, *10*(1), 834. https://doi.org/10.1038/s41598-020-57866-2

Erazo, E., Carbonell, J., & Isaacs, C. (2015). Factors inducing spatial variability in the productivity of sugarcane crops. *International Sugar Journal*, *117*(1403).

Fritsch, S., Guenther, F., Wright, M., Suling, M., & Mueller, S. (2019). *Package "neuralnet": Training of Neural Networks* (pp. 1–15).

Gallardo-Romero, D. J., Apolo-Apolo, O. E., Martínez-Guanter, J., & Pérez-Ruiz, M. (2023). Multilayer Data and Artificial Intelligence for the Delineation of Homogeneous Management Zones in Maize Cultivation. *Remote Sensing*, *15*(12), 3131. https://doi.org/10.3390/rs15123131

Gavioli, A., De Souza, E. G., Bazzi, C. L., Schenatto, K., & Betzek, N. M. (2019). Identification of management zones in precision agriculture: An evaluation of alternative cluster analysis methods. *Biosystems Engineering*, *181*, 86–102. https://doi.org/10.1016/j.biosystemseng.2019.02.019

Guastaferro, F., Castrignanò, A., De Benedetto, D., Sollitto, D., Troccoli, A., & Cafarelli, B. (2010). A comparison of different algorithms for the delineation of management zones. *Precision Agriculture*, *11*(6), 600–620. https://doi.org/10.1007/s11119-010-9183-4

Hengl, T. (2009). *A Practical guide to Geostatistical Mapping*. Office for Official Publications of the European Communities. https://doi.org/10.1016/0277-9390(86)90082-8

Jeczmionek, E., & Kowalski, P. A. (2022). Input reduction of convolutional neural networks with global sensitivity analysis as a data-centric approach. *Neurocomputing*, *506*, 196–205. https://doi.org/10.1016/j.neucom.2022.07.027

Lin, H., Wheeler, D., Bell, J., & Wilding, L. (2005). Assessment of soil spatial variability at multiple scales. *Ecological Modeling*, *182*(3–4), 271–290. https://doi.org/10.1016/j.ecolmodel.2004.04.006

Lu, C.-T., Chen, D., & Kou, Y. (2003). Algorithms for spatial outlier detection. *Third IEEE International Conference on Data Mining*, 597–600. https://doi.org/10.1109/ICDM.2003.1250986

Nawar, S., Corstanje, R., Halcro, G., Mulla, D., & Mouazen, A. M. (2017). Delineation of Soil Management Zones for Variable-Rate Fertilization. In *Advances in Agronomy* (Vol. 143, pp. 175–245). Elsevier. https://doi.org/10.1016/bs.agron.2017.01.003

Ohana-Levi, N., Bahat, I., Peeters, A., Shtein, A., Netzer, Y., Cohen, Y., & Ben-Gal, A. (2019). A weighted multivariate spatial clustering model to determine irrigation management zones. *Computers and Electronics in Agriculture*, *162*, 719–731. https://doi.org/10.1016/j.compag.2019.05.012

Ortega, R. A., & Santibáñez, O. A. (2007). Determination of management zones in corn (*Zea mays* L.) based on soil fertility. *Computers and Electronics in Agriculture*, *58*(1), 49–59. https://doi.org/10.1016/j.compag.2006.12.011

Peukert, S., Bol, R., Roberts, W., Macleod, C. J. A., Murray, P. J., Dixon, E. R., & Brazier, R. E. (2012). Understanding spatial variability of soil properties: A key step in establishing field- to farm-scale agro-ecosystem experiments. *Rapid Communications in Mass Spectrometry*, *26*(20), 2413–2421. https://doi.org/10.1002/rcm.6336

Robert, P. C. (2002). Precision agriculture: A challenge for crop nutrition management. *Plant and Soil*, *247*(1), 143–149. https://doi.org/10.1023/A:1021171514148

Salah, W., & Hannan, A. (2020). Data Processing Using Artificial Neural Networks. In D. G. Harkut (Ed.), *Dynamic Data Assimilation—Beating the Uncertainties*. IntechOpen. https://doi.org/10.5772/intechopen.91935

Santos, R. T., & Saraiva, A. M. (2015). A Reference Process for Management Zones Delineation in Precision Agriculture. *IEEE Latin America Transactions*, *13*(3), 727–738. https://doi.org/10.1109/TLA.2015.7069098

Satizábal, H. F., & Pérez-Uribe, A. (2007). Relevance Metrics to Reduce Input Dimensions in Artificial Neural Networks. In J. M. de Sá, L. A. Alexandre, W. Duch, & D. Mandic (Eds.), *Lecture Notes in Computer Science, vol 4668, Artificial Neural Networks – ICANN 2007* (pp. 39–48). Springer. https://doi.org/10.1007/978-3-540-74690-4_5

Shaikh, T., Rasool, T., & Rasheed Lone, F. (2022). Toward leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, *198*, 107119. https://doi.org/10.1016/j.compag.2022.107119

Shamir, O. (2016). Distribution-specific hardness of learning neural networks. *Journal of Machine Learning Research*, *19*, 32:1-32:29.

Von Hebel, C., Reynaert, S., Pauly, K., Janssens, P., Piccard, I., Vanderborght, J., Van Der Kruk, J., Vereecken, H., & Garré, S. (2021). Toward high-resolution agronomic soil information and management zones delineated by ground-based electromagnetic induction and aerial drone data. *Vadose Zone Journal*, *20*(4), e20099. https://doi.org/10.1002/vzj2.20099

Yuan, Y., Deng, Y., Zhang, Y., & Qu, A. (2020). Deep learning from a statistical perspective. *Stat*, *9*(1), e294. https://doi.org/10.1002/sta4.294

Zhang, Y., Zhen, Q., Cui, Y., Zhang, P., & Zhang, X. (2020). Use of montmorillonite-enriched siltstone for improving water condition and plant growth in sandy soil. *Ecological Engineering*, *145*, 105740.

https://doi.org/10.1016/j.ecoleng.2020.105740

Zurada, J. M., Malinowski, A., & Usui, S. (1997). Perturbation method for deleting redundant inputs of perceptron networks. *Neurocomputing*, *14*, 177–193.