Proximal and remote sensor data fusion for in-depth salinization mapping in the Brazilian semiarid via machine learning

TAVARES, Sílvio R. L.[1]; VASQUES, Gustavo M.[1]; OLIVEIRA, Ronaldo P.[1]; DANTAS, Marlon M. [2]; RODRIGUES, Hugo M.[3]
[1] Embrapa Solos, silvio.tavares@embrapa.br, gustavo.vasques@embrapa.br, ronaldo.oliveira@embrapa.br, luis.hernani@embrapa.br; [2] Instituto Federal do Rio Grande do Norte, dantasmm1@gmail.com; [3] Universidade Federal Rural do Rio de Janeiro, rodrigues.machado.hugo@gmail.com

Thematic Session: Advances in soil sensing

Abstract
Mapping the salinization in irrigated cropland is a challenging practice. As an alternative, data from proximal and remote sensors have been implemented together via datafusion and machine learning algorithms. The present work was carried out on a farm with 11 ha and used data from the proximal sensor EM38-MK2 associated with radar C-band data obtained by the Sentinel1 satellite. The salinization classes were created from electrical conductivity data measured at 35 points using a 50 x 50 m sampling grid and at three depths: 0 – 10, 10 – 30, and 30 – 50 cm using conventional laboratory approach. The accuracy values of the class prediction models presented values between 0.66 and 0.74 and Kappa values between 0.43 and 0.59 using Random Forest. The salinization decreased in layers 0 - 10 and 10 - 30 cm due to implementing a surface drainage system but the depth 30 - 50 cm had the highest occurrence of Salic classes, with a potentially harmful effect on the roots.

Introduction

The salinized regions in high-temperature areas occur via water evaporation from the soil, transpiration for vegetables and the carry-over of salts that settle on the surface. Some agricultural techniques have been proposed to mitigate these effects, such as using water with lower electrical conductivity values and applying drainage systems to remove salts. However, methods to monitor the occurrence of salinization after the application of mitigating activities are challenging, as identifying the occurrence of soil salinization requires a high number of soil samples for laboratory analysis through the analysis of electrical conductivity in saturated paste this being a financially costly method that does not allow covering large areas.
Proximal sensors have been proposed as an alternative to monitoring areas where salinization occurs. Apparent electrical conductivity and apparent magnetic susceptibility data have been reported as potential attributes that allow to identify and map salinization advance or retreat since these sensor attributes are closely related to clay content, moisture, cation exchange capacity, and pH (LOPES; MONTENEGRO, 2019). As an aid to mapping the occurrence of salinization in irrigated plantation areas, it is also possible to combine data from proximal sensors with other data sources, such as radar data obtained by satellites (HUANG; PROCHAZKA; TRIANTAFILIS, 2016) and which are freely available for use.
Therefore, the objective of the work was to spatialize the occurrence of soil salinization from the predictive mapping of the salinization classes reflected by the electrical conductivity measured in the laboratory as a function of proximal and remote sensor data via machine learning algorithm Random Forest.

Methodology

The study was conducted on a family farm at Baixo Açu irrigated perimeter and has approximately 11 ha. It is in northeastern Brazil in the region of Alto do Rodrigues – RN (Figure 2). The region's climate is Aw via Köppen-Geiger and has a rainfall regime with an average annual occurrence of 400 mm and an average annual maximum temperature of 34°C. Due to its semiarid conditions, the area is subject to natural soil salinization processes that have been increased by irrigated crop production.
The EM38-MK2 (Geonics Limited, Mississauga, Canada) was used for continuous aEC and aMS readings (N=5,168 points) on zig-zag footstep tracks in "1 m" (aEC and aMS 1 m) and "0. 5 m" (aEC and aMS 0.5 m) coil separation mode on vertical orientation. These data were spatially characterized using semivariogram adjustments and then interpolated by ordinary kriging with 10 m resolution. One thousand five hundred fifty-one points or 30% of the original dataset were intended to validate the four maps produced (Figure 2).
The two vertical-vertical and vertical-horizontal polarizations present in the C-band of the Sentinel-1 satellite were selected (Figure 1. h. and i.).
For soil salinity from laboratory analysis, soil core samples were collected in a 35 points uniform grid (50 x 50 m; Figure 1), at 0 – 10, 10 – 30, and 30 – 50 cm depth, and the samples were analyzed using the method of the electrical conductivity measured in the saturated paste (EC_{lab}) as described in Embrapa's methods manual (TEIXEIRA et al., 2017). In addition, the pH data at the 35 points were also measured, and then these were spatialized using the inverse square distance method.
The salinity data for the three depths were classified according to their degree of salinization using the limits defined for the characteristics "Not saline" (EC_{lab} < 4 dS m^{-1}), "Saline" (4 dS m^{-1} < EC_{lab} < 7 dS m^{-1}), and "Salic" (7 dS m^{-1} < EC_{lab}) using the limits defined in the Brazilian Soil Classification System (SANTOS et al., 2018).
The set of covariates comprises four maps of proximal sensors stacked with the two radar polarized band maps and the three pH maps, totaling nine predictor covariates. EC_{lab} data were modeled and mapped from the model's fit for each salinity class by depth using the Random Forest classifier present in the caret package in the R software and will be evaluated for accuracy using the kappa index and leaving one out cross-validation.

Results and discussion

The aEC and aMS maps showed higher values in the central locations with a tendency to grow to the east, agreeing with the direction of drainage (Figure 1; a., b.). The pH maps show high values (> 7) for the entire study area, reinforcing the presence of salts in the soil (Figure 1; e., f., g.). The radar C-band also showed higher values in the west and northwest regions of the study area, agreeing with the drainage orientation (Figure 1; h., i.).
The aEC and aMS maps showed external validation errors less than 80 mS/m and 0.2 ppt, respectively. The pH covariates for all depths were the most important in all salinization Random Forest models (Table 1). While the 1 m aMS was the second most important for the 10 – 30 and 30 – 50 cm salinization models considering the data from

proximal sensors. The 0 – 10 cm salinization model did not show high significance between remote and proximal sensor data. Accuracy values for all models were around 0.7, while kappa values were close to 0.5 (Table 1).
The depth salinity map (Figure 3) shows a higher concentration of the "Not saline" class in the center and northeast of the maps, in agreement with the behavior of the aEC maps shown in Figure 2. The occurrence of salinization shows a decrease as it approaches the surface due to the existence of drainage channels built to drain the water used, represented in figure 3 in blue lines. The 30 - 50 cm map has a higher occurrence of the "Salic" class, demonstrating that it is a layer with toxic effects for some crops where the drainage activity may not have been enough to cannot attenuate the effects of salinization.

Conclusions

The data from proximal electromagnetic sensors combined with radar data obtained by remote sensors allowed to spatialize the phenomenon of salinization in an irrigated crop area with good accuracy via adjustments to prediction models using the Random Forest algorithm.
The pH data measured in the laboratory was fundamental for the construction of predictive models of salinization.
Data from remote and proximal sensors proved to be essential tools for monitoring and mapping the effects of salinization on the soil.

References

HUANG, J.; PROCHAZKA, M. J.; TRIANTAFILIS, J. Irrigation salinity hazard assessment and risk mapping in the lower Macintyre Valley, Australia. **Science of the Total Environment**, v. 551–552, p. 460–473, 2016.
LOPES, I.; MONTENEGRO, A. A. D. A. Spatialization of electrical conductivity and physical hydraulic parameters of soils under different uses in an alluvial valley. **Revista Caatinga**, v. 32, n. 1, p. 222–233, 2019.
SANTOS, H. G. DOS; JACOMINE, P. K. T.; ANJOS, L. H. C. DOS; OLIVEIRA, V. Á. DE; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A. DE; FILHO, J. C. DE A.; OLIVEIRA, J. B. DE; CUNHA, T. J. F. **Sistema brasileiro de classificação de solos**. 5. ed. Brasília, Brazil: Embrapa, 2018.
TEIXEIRA, P. C.; DONAGEMMA, G. K.; FONTANA, A.; TEIXEIRA, W. G. **Manual de métodos de análise de solo**. 3. ed. rev ed. Brasília, DF: Embrapa, 2017.

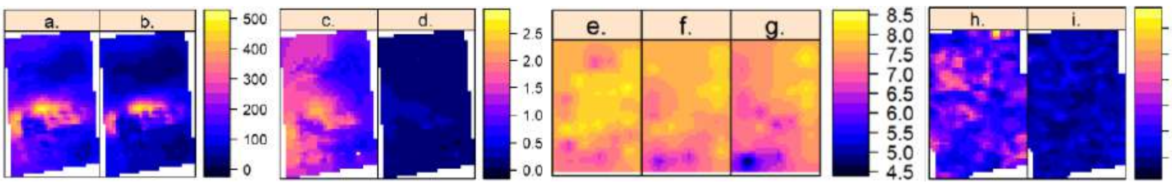

Figure 1: a. aEC 1 m; b. aEC 0.5 m; c. aMS 1 m; d. aMS 0.5 m; e. pH 0-10 cm; f. pH 10-30 cm; g. pH 30-50 cm; h. C-band vertical-vertical Sentinel-1; C-band vertical-horizontal Sentinel-1.

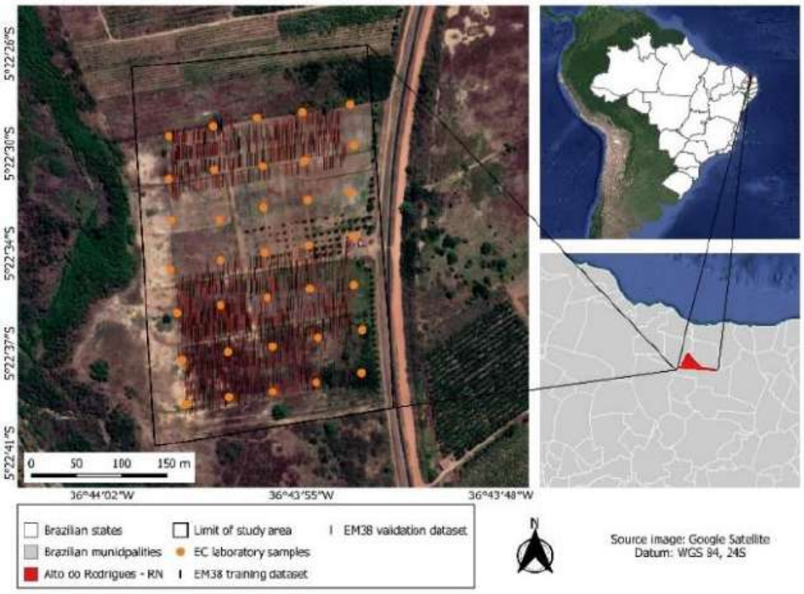

Figure 2: Location map of the study area and the sampling design.

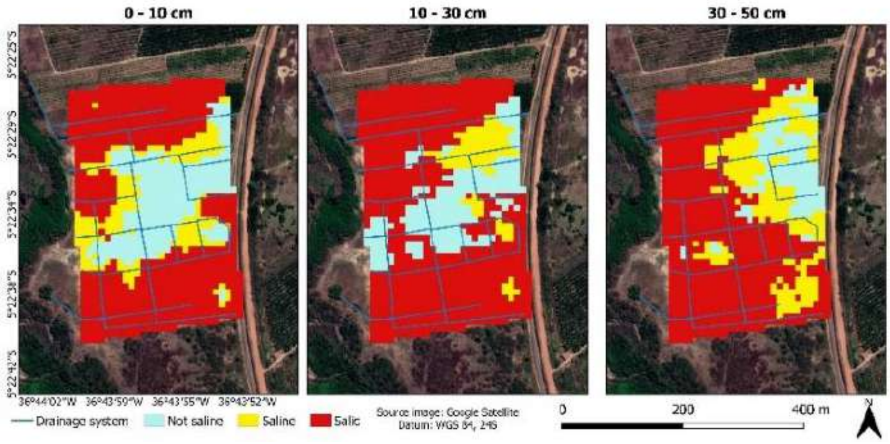

Figure 3: Spatialized occurrence of soil salinization in depth. A) 0 – 10 cm in depth; B) 10 – 30 cm in depth; C) 30 – 50 cm deep.

Table 1: Ranking of the importance of covariates in the Random Forest model and the accuracy and kappa values for each layer.

Salinity 0 - 10 cm	importance	Salinity 10 - 30 cm	importance	Salinity 30 - 50 cm	importance
pH 0 – 10 cm	100	pH 0 – 10 cm	100	pH 30 – 50 cm	100
B2	10.943	pH 10 – 30 cm	62.58	pH 0 – 10 cm	57.38
pH 10 – 30 cm	9.17	aMS 1 m	48.3	aMS 1 m	52.05
aMS 0.5 m	5.27	aEC 0.5 m	44.64	B1	27.35
aMS 1 m	4.998	pH 30 – 50 cm	43.51	aEC 0.5 m	25.41
aEC 0.5 m	3.835	aEC 1 m	32.86	pH 10 – 30 cm	19.91
pH 30 – 50 cm	3.485	B2	30.38	aEC 1 m	13.13
aEC 1 m	3.028	aMS 0.5 m	20.16	aMS 0.5 m	12.02
B1	0	B1	0	B2	0

	Salinity 0 - 10 cm	Salinity 10 - 30 cm	Salinity 30 - 50 cm
Accuracy	0.74	0.71	0.66
Kappa	0.59	0.52	0.43