Predicting soil electro-conductivity using Sentinel-1 images

Eduardo Medeiros[1]
efarofia@uevora.pt
Sajib Ahmed[1,3]
sajib@uevora.pt
Teresa Gonçalves[1,3]
tcg@uevora.pt
Luís Rato[1,2,3]
lmr@uevora.pt

[1] Departamento de Informática, Universidade de Évora, Portugal
[2] CIMA, Universidade de Évora, Portugal
[3] EaRSLab, Universidade de Évora, Portugal

Abstract

The quality and yield of a soil can be measured by using a wide range of soil indicators. One such indicator is soil's electro-conductivity which is an excellent indicator of the presence of soil nutrients. This work aims to create a machine learning model to predict the soil's electro-conductivity (EC) using radar images from the satellite Sentinel-1. Using EC readings from 14 corn field parcels and Sentinel-1 readings over the course of one agriculture year, several regression models were generated. These models were designed using information from the full agriculture year or only 3 months, both or only one of the VV and VH polarisations. The results show that when using a full year data VV and VH polarisations are able to generate models with similar performance (R^2 of 0.888 for VH and 0.884 for VV) but when using only 3 months data, only April to June trimester using both polarisations are able to reach similar a performance (R^2 of 0.867); moreover VH polarisation seems to carry out more descriptive information when compared with VV (specially when using only 3 months data). Finally, performance results seem to be independent of the yearly radar data time-window.

Keywords: Soil electro-conductivity, Remote sensing, Sentinel-1, Regression, K-nearest neighbours

1 Introduction

Precision farming incorporates a series of strategies and tools that allow farmers to optimise and increase soil quality and productivity putting in place a set of targeted key interventions. These interventions are selected based on collected information of minerals, nutrients, water, soil texture, drainage conditions, salinity, and other soil characteristics over farmland [3]. Soil electro-conductivity (EC) is one of simplest, and least expensive soil measurements available to precision farming [5].

Recently, soil properties are being obtained using remote sensing techniques [2]. Sentinel-1 [4] is a synthetic aperture radar instrument (SAR) satellite that consists of a set of two satellites, Sentinel-1A and Sentinel-1B, which share the same orbital plane with a 12-day revisiting period. This set of satellites provides images in two different polarisations: VV (vertical transmit, vertical receive) and VH (vertical transmit, horizontal receive).

In a previous work [1], Sentinel-1 information was used to build models able to classify soils as sandy, free and clayish (by discretizing EC values) achieving F1-scores between 54.44% and 75.6% for clayish and free soils, respectively, over a test set of 13001 points. The current work, instead, aims at predicting the EC value itself. Besides studying if polarisations have different discriminative power, it also studies which months are more informative and if the sentinel data from different years generates models with similar performances.

The rest of the paper is organized as follows: Section 2 introduces the data sets and algorithms used, describes the experiments performed and the experimental setup; Section 3 presents and discusses the results obtained; finally, Section 4 concludes the paper and presents future work.

2 Materials and Methods

2.1 Data sets

The on-site EC values were obtained between March 28 and May 3, 2016, on a set of 14 parcels of corn fields located in Alentejo with a 10-meter interval resulting in a total of 65003 points. Measured values ranged between 0.226 and 240.592; Figure 1 presents the distribution of the values.

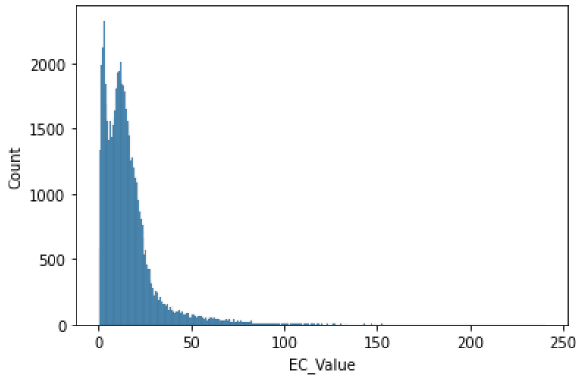

Figure 1: Histogram of EC values on the data set.

Radar data was collected from two time windows each corresponding to full agricultural years: (a) from October 6, 2018 to September 25, 2019 (the most recent year available when data was collected) and (b) from October 3, 2015 to September 28, 2016 (the year corresponding to the EC readings). The data extracted for the 2018-2019 time window corresponds to orbit 147 from both satellites because for Sentinel-1B there was only data for that orbit (although information from more orbits existed for Sentinel-1A). For 2015-2016 time window data was collected from 3 available orbits: 43, 50 and 145.

Similarly to the previous study [1], the 2018-2019 data is composed by 120 descriptive attributes corresponding to VV and VH values from both satellites over the considered time window (60 different dates; 30 from each satellite). Since satellite Sentinel-1B was only launched in 2016, the 2015-2016 data is composed by VV and VH values from Sentinel-1A, composed by 27, 26 and 29 different dates for orbits 43, 50 and 145, respectively.

Table 1 presents a characterisation of the radar values for each polarisation and satellite during the 2018-2019 time window. It is easly seen that the range of values for VH polarisation is much smaller than the range for VV polarisation. On the other hand, both satellite present similar ranges.

	Sat A		Sat B		Sat A+B	
	VH	VV	VH	VV	VH	VV
mean	95.81	216.82	88.09	219.49	91.95	218.16
std	24.93	66.60	24.05	67.88	24.80	67.26
min	28.33	62.11	25.11	58.78	25.11	58.78
25%	77.78	173.56	70.89	173.44	74.11	173.56
50%	94.78	207.89	86.78	209.89	90.67	208.89
75%	111.89	248.44	103.00	253.78	107.67	251.11
max	250.44	2009.33	242.22	1988.78	250.44	2009.33

Table 1: Characterisation of the attributes over a one year period.

2.2 Algorithms

Several machine learning algorithms for regression were tested namely Support Vector Machines (SVM), K Nearest-Neighbours (KNN), Ridge and Lasso.

2.3 Experiments

A first set of experiments was done using the 2018-2019 data to study the algorithms and the sets of features that generate the most performing models. SVM was tested with linear kernel and $C = \{0.1, 1, 10\}$, KNN tested with $K = \{1, 5, 9\}$, Ridge with $\alpha = \{0.1, 1, 10\}$ and Lasso with $\alpha = \{1, 0.01, 0.0001\}$ (the rest of the parameters were the default for all algorithms). Models were built using VV and VH values, alone and together, for all available dates (60 attributes per polarisation), by monthly averaging them (12 attributes per polarisation) and by using trimester values (15 attributes per polarisation) instead of the full year.

A second set of experiments, using the 2015-2016 data (using one and both polarisations), was also experimented, aiming to check if the models performed differently using radar information from different years.

2.4 Experimental Setup

For developing the models Python (v3.7.9) with scikit-learn (v0.23.2) were used and a stratified train-test split generated with 75% for training (48752 samples) and 25% for testing (16250 samples). The models were evaluated using the coefficient of determination, R^2, a performance measure that normally ranges between 0 and 1, with 0 corresponding to a constant model that always predict the training test average value and 1 corresponding to a perfect prediction. Experiments were also performed to check if the existence of outliers (namely, very high values of VV) influenced the results; no influence was found.

3 Results

This section presents and discusses the results obtained with 2018-2019 and 2015-2016 time windows data sets.

3.1 2018-2019 data

First, a set of experiments, aiming to find the algorithm and parameters, was performed using both polarisations and all dates (120 attributes) and a 5-folds cross-validation over the training set. Table 2 presents the results over the test set for the best parameters found. As can be seen, KNN performs best by a large margin.

Algorithm	Parameters	R^2
LinearSVM	$C = 1$	0.243
KNN	$K = 5$	**0.883**
Ridge	$\alpha = 1$	0.331
Lasso	$\alpha = 0.01$	0.331

Table 2: R^2 results for the best performing parameters.

Then, a more thorough search over the KNN parameters was conducted (also using a 5-folds cross-validation over the training set) with K={1,3,5,7,9}, weights={*uniform*, *distance*} and Minkowski distance with $p = \{1, 2, 3\}$. The parameters with best results were $K = 3$, $p = 1$ and weights=*distance*.

Finally, using the fine-tuned parameters, the different sets of attributes mentioned in subsection 2.3 were tested. Table 3 presents results using the full year data (all dates and monthly average) and trimester data.

Pol	all	m. avg	Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep
VH	**0.888**	0.777	0.636	0.662	0.777	0.718
VV	0.884	0.716	0.606	0.571	0.714	0.657
both	0.886	**0.860**	0.839	0.848	**0.867**	0.854

Table 3: R^2 results for full year and trimester dates.

Looking at the full year data one can conclude that, when using yearly individual dates, the results obtained are very similar using one (60 attributes) or both polarisations (120 attributes), with VH presenting the best result with $R^2 = 0.888$. This is not true when using monthly values: the best results, by a large margin, are obtained using information from both polarisations (12 attributes for each polarisation), with a result of $R^2 = 0.860$. Nonetheless, when comparing single polarisations, VH continues to carry more discriminant information than VV ($R^2 = 0.777$ vs. $R^2 = 0.716$).

When comparing trimester data one can conclude that Apr-Jun trimester data is the most informative one reaching a value of $R^2 = 0.867$ with both polarisations (30 attributes). On the other end Oct-Dec trimester data is the least informative. Also, when using trimester data, using both polarisations increase the performance substantially (more than 10%, with higher increases for the least performing trimesters).

3.2 2015-2016 data

As previously mentioned, this set of experiments aimed at checking if the models performed differently using radar information from different years; 2015-2016 time window was chosen to include the on-site EC readings. Since Sentinel-1B was only launched in 2016, this data only contains values from Sentinel-1A, but readings of the 3 available orbits were collected. Parameter fine-tuning was also conducted over KNN algorithm, with the best performance obtained with the the same parameters.

In order to compare the results between the two time windows, an additional experiment with 2018-2019 data was conducted using only the Sentinel-1A readings. Table 4 presents the results.

	2015-2016			2018-2019
Pol	Orb 43	Orb 50	Orb 145	Orb 147
VH	0.869	0.869	0.880	0.876
VV	0.865	0.879	0.876	0.873
both	0.880	0.876	0.885	0.886

Table 4: 2015 data set scores for each orbit and polarisation combination.

As can be seen from the table, when using a full year radar information, different orbits generate models with similar performances (with higher values for orbit 145 possibly because it contains more dates); once again VH polarisation outperforms VV and a minor improvement over VH performance is observed when adding VV information. On the other hand, it is easily seen that the results seem to be independent of the time window of the collected radar data.

4 Conclusions and Future Work

This work presents a regression model to determine the soil eletro-condutivity using radar information. The developed models reached a R^2 performance of 88.8% using data from both satellites and a full year time window (2018-2019 data). Nonetheless, similar performances were obtained using information from just one satellite (88.6%). Moreover, VH polarisation seems to carry out more descriptive information when compared with VV, obtaining a similar performance to using both polarisations (when using a full year time window). Also, when using trimester data, Apr-Jun has the highest R^2 (86.7%) while Oct-Dez has the lowest (83.9%). Finally, performance results seem to be independent of the yearly radar data time-window.

As future work, we intend further investigate the use of radar data aiming to produce better models (joining info from different orbits and/or different years) and to build an application that, given a site's set of radar images, is able to generate the corresponding electro-conductivity map.

Funding

This work was supported by NIIAA (Núcleo de Investigação em Inteligência Artificial em Agricultura) project, Alentejo 2020 program (reference ALT20-03-0247-FEDER-036981).

References

[1] Sajib Ahmed, Teresa Gonçalves, Luís Rato, Pedro Salgueiro, J. R. Marques da Silva, and Luis Paixão Filipe Vieira. Classifying soil type using radar satellite images. 2020.

[2] Yufeng Ge, J Alex Thomasson, and Ruixiu Sui. Remote sensing of soil properties in precision agriculture: A review. *Frontiers of Earth Science*, 5(3):229–238, 2011.

[3] Robert Dwight Grisso, Marcus M Alley, David Lee Holshouser, and Wade Everett Thomason. Precision farming tools. soil electrical conductivity. -, 2005.

[4] Paul Snoeij, Evert Attema, Malcolm Davidson, Berthyl Duesmann, Nicolas Floury, Guido Levrini, Björn Rommen, and Betlem Rosich. The sentinel-1 radar mission: Status and performance. In *2009 International Radar Conference" Surveillance for a Safer World"(RADAR 2009)*, pages 1–6. IEEE, 2009.

[5] Iulian-Florin Voicea, Mihai Matache, and Valentin Vladut. Researches regarding the electro-conductivity determination on different soil textures from romania, before sowing. *Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Agriculture*, 66(1), 2009.