https://doi.org/10.17221/101/2019-SWR

Delimitation of low topsoil moisture content areas in a vineyard using remote sensing imagery (Sentinel-1 and Sentinel-2) in a Mediterranean-climate region

Maria Paula Mendes, Magda Matias, Rui Carrilho Gomes, Ana Paula Falcão*

CERIS, Civil Engineering Research and Innovation for Sustainability, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

**Corresponding author: mpaulamendes@tecnico.ulisboa.pt*

Citation: Mendes M.P., Matias M., Gomes R.C., Falcão A.P. (2021): Delimitation of low topsoil moisture content areas in a vineyard using remote sensing imagery (Sentinel-1 and Sentinel-2) in a Mediterranean-climate region. Soil & Water Res., 16: 85–94.

Abstract: Irrigation can be responsible for salt accumulation in the root zone of grapevines when late autumn and winter precipitation is not enough to leach salts from the soil upper horizons, turning the soil unsuitable for grape production. The aim of this work is to present a novel methodology to outline areas, within a drip-irrigated vineyard, with a low soil moisture content (SMC) during, and after, an 11-month agricultural drought. Soil moisture (SM) field measurements were performed in two plots at the vineyard, followed by a geostatistical method (indicator kriging) to estimate the SM class probabilities according to a threshold value, enlarging the training set for the classification algorithms. The logistic regression (LR) and Random Forest (RF) methods used the features of the Sentinel-1 and Sentinel-2 images and terrain parameters to classify the SMC probabilities at the vineyard. Both methods classified the highest SMC probabilities above 14% that is located close to the stream at the lower altitudes. The RF method performed very well in classifying the topsoil zones with a lower SMC during the autumn-winter period. This delineation allows the prevention of the occurrence of areas affected by salinisation, indicating which areas will need irrigation management strategies to control the salinity, especially under climate change, and the expected increase in droughts.

Keywords: logistic regression; radar; random forest; soil moisture content; soil salinisation; vineyard

Droughts can increase the processes of salt mobilisation and accumulation within the upper soil horizons (FAO 2018). Over the last several years, Portugal showed a reduction in the actual evapotranspiration and an increase in the soil water deficit (Kurnik et al. 2015). Drip irrigation increases the water use efficiency, but can be responsible for salt accumulation in vineyard soils, due to the low water application rates (Aragüés et al. 2014). Vines can decrease their yield and, eventually, the soil salinity can cause vine mortality (FAO 2018). The use of imaging radar such as the Sentinel-1 (S1) satellite constellation can be a cost-effective solution to monitor the soil moisture content (SMC) due to the open access of the data with a high spatial and temporal resolution. The radar signal scattered by the surface is dependent on the ground surface roughness, as well as the dielectric characteristics of the first centimetres of the soil (Bousbih et al. 2017). The use of multispectral imagery bands of Sentinel-2 (S2), such as RED or Near-Infrared (NIR) bands, enables one to register the reflected energy by the vegetation and surface, allowing, in an expedited way, one to monitor the agriculture drought (Amani et al. 2016). Multi-source

Supported by Transversal Projects inter CERIS Groups, FCT, Project AQUIREMOTE: "Use of remote sensing data as a support for the EU agri-environmental policies for achieving a good status". M. Matias was funded by AQUIREMOTE.

https://doi.org/10.17221/101/2019-SWR

satellite data are advantageous in terms of classification, because optical sensors are sensitive to the reflective and spectral characteristics of ground targets whereas synthetic aperture radar (SAR) sensors are sensitive to their structural, textural, and dielectric characteristics. The penetration capacity of the signal depends on the wavelength, the greater the length, the greater the ability to penetrate the vegetation (El Hajj et al. 2019).

Artificial neural networks have been widely used to retrieve surface SMCs (Santi et al. 2018), but they normally act like a black box, thus not presenting the relationship between the variables (Bishop 1995). Supervised classification algorithms (SCA), such as logistic regression (LR) and Random Forest (RF), are more transparent about their results, since a feature's importance can be computed, improving the model interpretability (Rodriguez-Galiano et al. 2018). LR and RF are used for the classification (high or low) of the surface SMC (top 5 cm) considering the different grapevines growing stages and weather conditions. The aim was to develop a methodology that can delineate the probability zones within a drip-irrigated vineyard that have low SMCs during late autumn and winter.

MATERIAL AND METHODS

Catapereiro's vineyard

The case study is part of a state-owned farm and is located at Catapereiro (38°48'44.00''N, 8°53'2.03''W, 20 m a.s.l.) in the south-western part of the Portuguese mainland, in the Tagus River Basin (Figure 1). This region was identified as Mediterranean Temperate and Sub-oceanic, considering the effect of the water availability and the thermal regime of the soil (Hartwich et al. 2005). This area is filled by recent alluvial deposits, mainly composed of sandy or sandy loam soils (Mendes et al. 2016), and is located in an area threatened by salinisation (Adhikari 2008). Catapereiro is in the discharge area of the Tagus alluvium aquifer. Two plots were delimited with an average area of 3 ha, named as plot A and plot C, at Catapereiro's Vineyard (area = 48 ha; Figure 1).

Climate variability

In the study area, most of the precipitation falls between October to March/April. The coldest month is January with a mean temperature of 9 °C and July is the hottest month with a mean temperature of 22 °C (Mendes et al. 2016). The Standardized Precipitation-

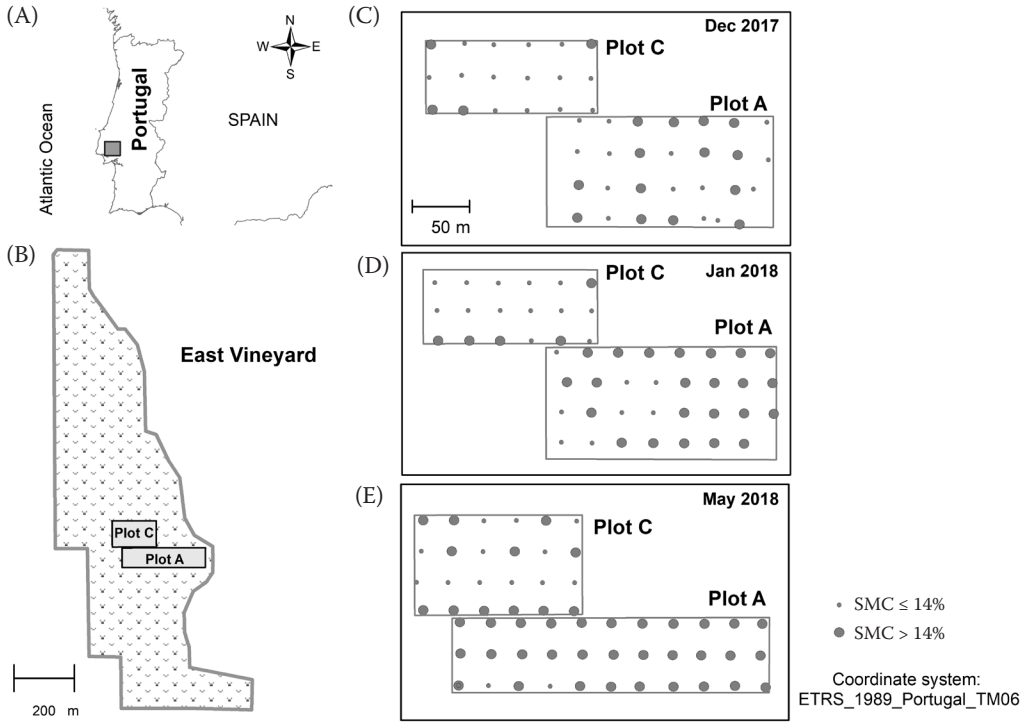

Figure 1. Location of the study area in mainland Portugal (A); setting the pilot plots within the Catapereiro's Vineyard (B); soil moisture sampling plans at the two sites for Dec-2017, Jan-2018 and May-2018 (C), (D) and (E), respectively (the dots indicate the measurement locations and they were classified according to the two classes of soil moisture content (SMC))

https://doi.org/10.17221/101/2019-SWR

Evapotranspiration Index (SPEI) (Vicente-Serrano et al. 2010) was estimated to assess the agricultural drought and to establish the baseline SMC values related to the drier soil conditions. The SPEI considers the monthly difference between the precipitation (P) and the potential evapotranspiration (Vicente-Serrano et al. 2010). A six-month scale was chosen since we were assessing the agricultural drought. According to the SPEI 6 months results (Figure 2), 2017 was a drought year (with 11 months classed as a drought), where December 2017 was classified as a normal month (SPEI ≤ 0) with low mean temperatures and precipitation amount. January 2018 had average conditions, but May 2018 was classified as extremely wet (SPEI ≥ 2).

Logistic regression and Random Forest

SCA are applied to a training set (features or explanatory variables) and a target feature defined as a class label, creating a classifier that is used to assign class labels to a new set (Kotsiantis 2007). LR is used in this study since it is the most widely applied classifier, especially outside the machine learning community (Yang & Loog 2018). LR considers a dummy target feature (in this study, encoding values of SMC ≤ 14% equal to 1, and zero otherwise) and a set of attributes, i.e., explanatory variables. A logit transformation:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \quad (1)$$

is applied to the logistic regression model:

$$\pi(x) = \frac{e^{\beta_0} + e^{\beta_1 x}}{1 + e^{\beta_0} + e^{\beta_1 x}} \quad (2)$$

becoming:

$$g(x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_q x_q \quad (3)$$

which estimates the probability of the class membership (Hosmer & Lemeshow 2000). The parameters in the LR model were estimated by the maximum likelihood.

RF is a supervised machine learning algorithm, where a large collection of classification trees is ensembled by bagging (Ghasemian & Akhoondzadeh 2018). The trees' growth occurs from a subset of evidential features that have been randomly chosen by bagging (normally 2/3 of the total overall set of the input evidential features). The samples that were not selected for the training of a tree are included as part of another subset called the out-of-bag (oob), representing, 1/3 of the training sample (Rodriguez-Galiano et al. 2018). The proportion of misclassifications (%) of the overall out-of-bag elements is called the oob error. These oob elements allow the RF to compute an unbiased estimation of the generalisation error without using an external data subset (Breiman 2001). RF only needs two parameters to tune it: the number of classification trees (ntree) and the number of predictive variables (mtry), which are used in each node to make decision trees grow (Liaw 2018). The RF model will be chosen based on the best compromise between the number of ntree and the generalisation error.

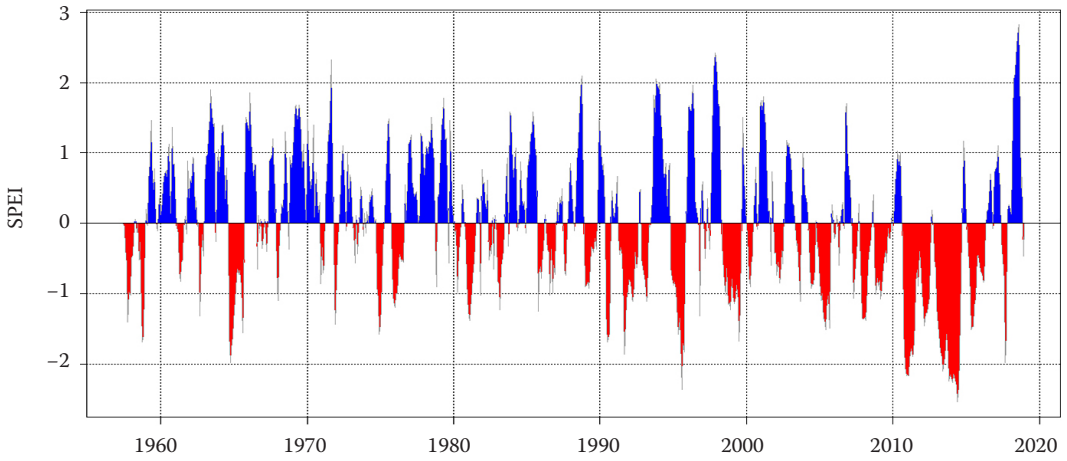

Figure 2. Values of the 6-month Standardized Precipitation-Evapotranspiration Index (SPEI), between 1958 and 2018; the red areas (SPEI < 0) are for the dry months and blue areas (SPEI > 0) for the wet months

https://doi.org/10.17221/101/2019-SWR

Performance evaluation. The validation of LR and RF can be presented as a binary confusion matrix with two classes (positive and negative). Observations classified correctly into the positive/negative class are called true positives/negatives. Observations classified incorrectly into the negative/positive class are false negatives/positives. From the frequency of the true and false positive/negative observations, several performance measurements may be built, such as (Tharwat 2020): (*i*) accuracy, i.e., the likelihood of success in identifying the right class of an instance; (*ii*) sensitivity, i.e., the probability of correctly recognising a positive class instance; (*iii*) precision, i.e., the probability that a predicted positive class instance is a true positive; (*iv*) specificity, i.e., the likelihood in recognising a negative class instance well; (*v*) false positive rate, i.e., the probability of failure in recognising a negative class instance and; (*vi*) the *F*-measure, i.e., the harmonic mean of the precision and accuracy:

$$F-\text{measure}=\frac{2\times\text{accuracy}\times\text{precision}}{\text{accuracy}+\text{precision}} \quad (4)$$

where its best value is 1 and worst is 0.

The *F*-measure will be used as the overall performance evaluation measure, since the error rate is not an appropriate evaluation criterion when there is a class-imbalance (Liu et al. 2009).

Features selection. The Feature selection (FS) is a process of selecting a subset of original input explanatory variables, supported by an optimisation criterion (Blum & Langley 1997). The Wald test (Hosmer & Lemeshow 2000) was used to test the statistical significance of each coefficient (β, Equation (3)) in the model, i.e., to evaluate the feature's importance. Features with smaller Wald test values were removed and a new LR model was computed with the remain features. If an improvement in the performance of the LR model was observed, then this new model will be the one that is used to classify the SMC. RF is an embedded method where the assessment of the FS considers the Gini importance and the Mean Decrease Accuracy (Rodriguez-Galiano et al. 2018).

Field data and remote sensing images

Sentinel-1 and Sentinel-2 data. During the monitoring period, 14 scenes, 11 from Sentinel-1 A and B, Level 1, interferometric wide and Single Look Complex and three scenes from the Sentinel-2 Multispectral Instrument, processed to level 2A, were downloaded from the ESA (European Space Agency) Hub and processed for this study (Table 1). Sentinel-1 carries a 12 m-long advanced SAR, working in the C-band. The vertical transmit and receive (VV) and the vertical transmit and horizontal receive (VH) polarisations and the local incidence angles range from

Table 1. Characteristics of the remote sensing variables for the monitoring period

Image	Date	Sensor	Reference	Date acquisition	Pass direction	Product type	Spat res. (m)	Polarization
1	Dec 2017	S-1 A	S1AA041217	04/12/17	ascending	SLC	~13*	VV
2		S-1 B	S1BA51217	05/12/18	ascending	SLC	~13*	VV
3		S-2	S2061217 NIR	06/12/17	descending	MSI1C	10	
4		S-1 A	S1AD091217	09/12/17	descending	SLC	~13*	VV
5	Jan 2018	S-2	S2250118	25/01/18	descending	MSI1C	10	
6		S-1 A	S1AD260118	26/01/18	descending	SLC	~13*	VV
7		S-1 B	S1BD270118	27/01/18	descending	SLC	~13*	VV
8		S-1 B	S1BA270118	27/01/18	ascending	SLC	~13*	VV
9		S-1 A	S1AA280118	28/01/18	ascending	SLC	~13*	VV
10	May 2018	S-1 A	S1AA080518	08/05/18	ascending	SLC	~13*	VV
11		S-1 B	S1BD080518	08/05/18	descending	SLC	~13*	VV
12		S-1 A	S1AD090518	09/05/18	descending	SLC	~13*	VV
13		S-1 B	S1BA090518	09/05/18	ascending	SLC	~13*	VV
14		S-2	S2100518:	10/05/18	descending	MSI2A	10	

Spat res – spatial resolution; NIR – near-infrared band reflectance; DEM – digital elevation model; *with a pixel resolution of 2.3 × 14.1 m (range × azimuth), the IW SLC products are re-sampled to approximately 13 m after processing; SLC – Single Look Complex; MSI – Multispectral Instrument; VV – Vertical transmit and receive

https://doi.org/10.17221/101/2019-SWR

27° to 48°. In order to harmonise the Sentinel-1 and Sentinel-2 image geometries, ESA's SNAP Sentinel-1 toolbox was used. A pre-processing procedure was conducted including updating the orbit metadata, the multilooking process, the radiometric calibration (i.e., backscatter intensity) and the terrain correction (i.e., orthorectification), resulting in geo-coded backscatter intensity images. The backscatter intensity images were then converted into normalised backscattering coefficient (σ 0) values in dB (i.e., the standard unit for the SAR backscattering representation). All the datasets (S1, S2 and *in situ* measurements) were georeferenced and converted to the same coordinate system in order to ensure its correct spatial location. This procedure aims to minimise the effect of having diverse spatial data sources and although a residual error still exists, the rate of noise from such imperfections is negligible when compared to the modelled information.

According to the proposed methodology, all the datasets were used as independent inputs and the selection of the values for the training and test the algorithm was made based on a predefined grid in the same coordinate system as the data.

The values of S1 and S2, for the training test, were used as the independent inputs in the classification process. Only the backscatter measurements from the VV polarisation were used, based on a previous statistical analysis, and supported by publications of other authors, such as Bousbih et al. (2017); Gherboudj et al. (2011).

A univariate statistics summary of the remotely sensed imagery is given in Table 2.

Field and ancillary data. The SMCs were measured *in situ* at a 5 cm depth, using a Theta Probe type ML2x (Delta T Devices, UK). Following the procedure of Roth et al. (1992), the volumetric water content in m^3/m^3 versus the dielectric constant was determined in the laboratory using several samples collected at both pilot sites. The soil texture characterisation was performed in the Geotechnics Laboratory of the Civil Engineering Department by the dry sieving method. The soil samples were classified as SW – well-graded sand with low fine content (~< 5%).

Three field campaigns were carried out: on the 5th of December 2017 (Dec-17), the 26th of January 2018 (Jan-18) and the 5th of May 2018 (May-18) (Figure 3). The grapevines have a 2.3 m space between

Table 2. Descriptive statistics of the features used to define the ranges of the classes

Date	Features	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Dec-17	DEM	14.79	15.79	16.48	16.98	17.79
	S1A041217	0.02	0.03	0.03	0.04	0.07
	S1BA051217	0.03	0.03	0.04	0.05	0.12
	S1AD091217	0.04	0.05	0.05	0.07	0.28
	S2061217-RED	657.00	808.25	909.00	1183.00	2253.00
	S2061217-NIR	2142.00	2532.00	2786.00	3353.00	3899.00
Jan-18	DEM	14.91	16.04	16.75	17.16	17.67
	S1BA270118	0.03	0.04	0.05	0.05	0.07
	S1BD270118	0.03	0.04	0.04	0.05	0.45
	S1AA280118	0.03	0.05	0.06	0.07	0.12
	S1AD260118	0.04	0.05	0.05	0.06	0.09
	S2250118-RED	582.00	657.25	753.50	925.75	1347.00
	S2250118-NIR	2568.00	3155.00	3419.00	3632.00	4063.00
May-18	DEM	14.79	16.29	16.87	17.16	17.79
	S1BA090518	0.014	0.020	0.023	0.028	0.064
	S1BD080518	0.025	0.041	0.052	0.068	0.147
	S1AA0805184	0.013	0.018	0.023	0.028	0.088
	S2100518-RED	583.00	972.50	1347.50	1798.50	2589.00
	S2100518-NIR	3023.00	3436.00	3677.00	3858.00	4169.00

NIR – near-infrared band reflectance; DEM – digital elevation model; S1A and S1B features are the reference of each image, described and detailed in Table 1

https://doi.org/10.17221/101/2019-SWR

the rows, have a height of 1.80 m and are irrigated with groundwater. The vine rows are planted in a horizontal manner (E-W) not following the contour lines (Figure S1 in the Electronic Supplementary Material (ESM)). A grid of sampling points every 25 m (which correspond to n = 46, 48 and 57 measurements points for Dec-2017, Jan-2018, and May-2018, respectively) was used to monitor the SMCs in the two plots (Figure 1C, D, E). At least five measurements were taken at each measurement point and the mean value of the SMC was assigned. These SMC punctual measurements were converted into a raster format by Indicator Kriging (IK). IK was used to estimate the class probabilities of the SMC according to the SMC threshold value. This threshold value will be calculated based on the median value of the SMC measured in the field. By using IK, we enlarge our training set to around 200 points per campaign. A digital elevation model (DEM), with a resolution of 6 × 6 m, of the vineyard was generated from 216 points georeferenced with differential Trimble GPS equipment. The highest altitude is around 18.9 m, the lowest is 10.4 m and the mean altitude is 16 m. This is a flat area with gentle slopes (slope ≈ 1.1%), where the altitude decreases from the SSW-NNE direction towards the stream (Figure 3).

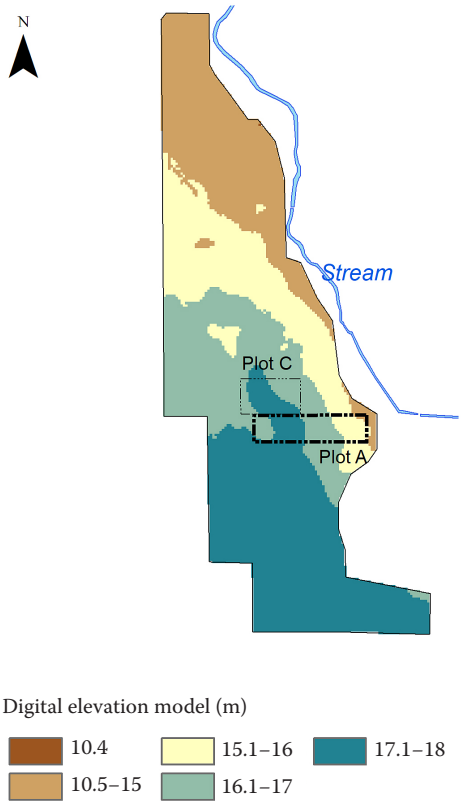

Figure 3. Digital elevation model

Figure 4 shows the box-plots of the SMC measurements in plots A and C for Dec-17, Jan-18 and May-18. A positive trend in the SMC values was observed, where plot A always had higher values compared to plot C. Plot A presented a slightly higher upper quartile (around 18%) in Dec-18, than those obtained in plot C in Jan-18 and May-18. Plot C always had an SMC inferior to 19% during the monitoring period, and the first quartile was always less than 13%. Based on this analysis, an SMC of 14% was chosen as the threshold value that will be used in the IK, since corresponds to the lowest median observed in the plot A. The probability values mapped by the IK were reclassified as a binary variable, where probability values higher than 50% were set as 1 or otherwise 0. These target features were used to train the LR and RF algorithms.

RESULTS AND DISCUSSION

Training and performance of LR and RF. For the logistic regression and Random Forest induction, remote sensing variables and a DEM were used with the target binary variable, SMC, as described in the previous section.

In Dec-2017, during the grapevines' dormancy, the LR showed the best performance (Table 3, F-measure = 0.72), where three features (DEM, S2061217-NIR, and S1B S2061217-; Table 3) were selected. In Jan-2018, the LR still had a reasonable performance (F-measure = 0.59). Only the data collected in May-18 exhibited a poor performance.

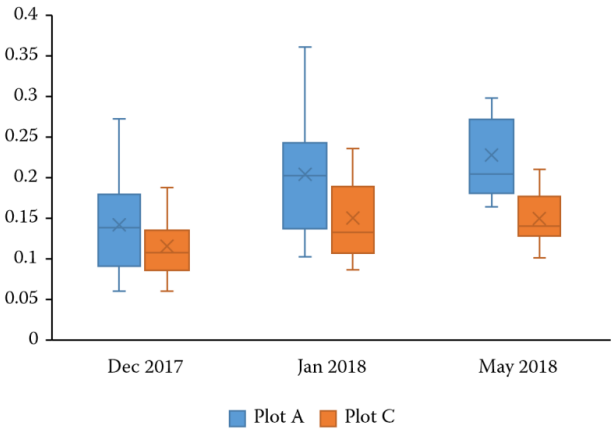

Figure 4. Box-plots of the soil moisture content of the training sites measured *in situ* for the three monitoring dates

https://doi.org/10.17221/101/2019-SWR

Table 3. Output from the logistic regression on the soil moisture contents (SMCs) with values < 14%

Date	Total No. of features	Features selected (Wald number)	Confusion matrix (percentage correct)			*F*-measure
			SMC ≤ 14%	SMC > 14%	overall	
Dec-17	6	DEM (5.61); S2061217-NIR (30.86); S1BA051217 (3.86)	68.0	88.6	81.2	0.72
Jan-18	7	DEM (25.50); S1AA280118 (3.13); S1D260118 (7.90); S1BD270118 (0.02); S1BA270118 (2.10); S2250118-RED (3.35)	56.3	78.6	70.2	0.59
May-18	7	DEM (6.90); S2100518-NIR (2.13); S1B A090518 (0.17); S1BD080518 (3.24); S1AD090518 (1.79)	8.3	100.0	82.3	0.15

NIR – near-infrared band reflectance; DEM – digital elevation model; S1A and S1B features are the reference of each image, described and detailed in Table 1

Likewise, the RF method showed the best performance measures (Table 4) during Dec-17. The DEM was the feature that best described the areas with high (i.e., probabilities lower than 50% of an SMC < 14%) and low SMCs (i.e. probabilities higher than 50% of an SMC < 14%). In fact, the DEM was the most important feature for the three field campaigns (Table 4) as expected, since it is related to the infiltration capacity. After a severe drought (Figure 2), in Dec-17, the grapevines still had some leaves, which explains the selection of S2061217-NIR and S2061217-RED by the RF for the SMC classification (Table 4). Concerning the S1 images, the best feature is given by the image acquired on the same day of

Table 4. Random Forest output of the soil moisture content (SMC) classification according to the threshold value of SMC < 14%

Date	Total No. of features	ntree	Features selected [Mean decrease Gini (G) & mean decrease Accuracy (A)]	Confusion matrix (percentage correct)			oob error (%)	*F*-measure
				SMC ≤ 14%	SMC > 14%	overall		
Dec-17	6	1400	DEM (G = 23.3; A = 49.7); S2061217-RED (G = 20.1; A = 38.0); S2061217-NIR (G = 28.1, A = 53.9); S1BA051207 (G = 8.0; A = 35.4); S1AD091217 (G = 9.0; A = 34.4); S1AA041217 (G = 5.7; A = 25.9)	77.3	93.2	85.3	12.6	0.82
Jan-18	7	400	DEM (G = 32.0; A = 35.3); S1B270118 (G = 20.4; A = 15.7); S1AD260118 (G = 19.2; A = 15.6); S2250118-NIR (G = 16.0; A = 16.0)	69.0	87.2	78.1	19.7	0.73
May-18	7	1600	DEM (G = 14.2; A = 15.0); S2100518-RED (G = 14.0; A = 12.2); S2100518-NIR (G = 12.1; A = 9.6); S1BA090518 (G = 2.9, A = 4.7); S1BD080518 (G = 5.4; A = 9.6); S1AD090518 (G = 3.9; A = 8.1); S1AA080518 (G = 4.0; A = 11.7)	25.0	95.3	60.2	12.6	0.35

ntree – number of classification trees of the Random Forest model; oob error – out-of-bag error; DEM – digital elevation model; NIR – near-infrared band reflectance

https://doi.org/10.17221/101/2019-SWR

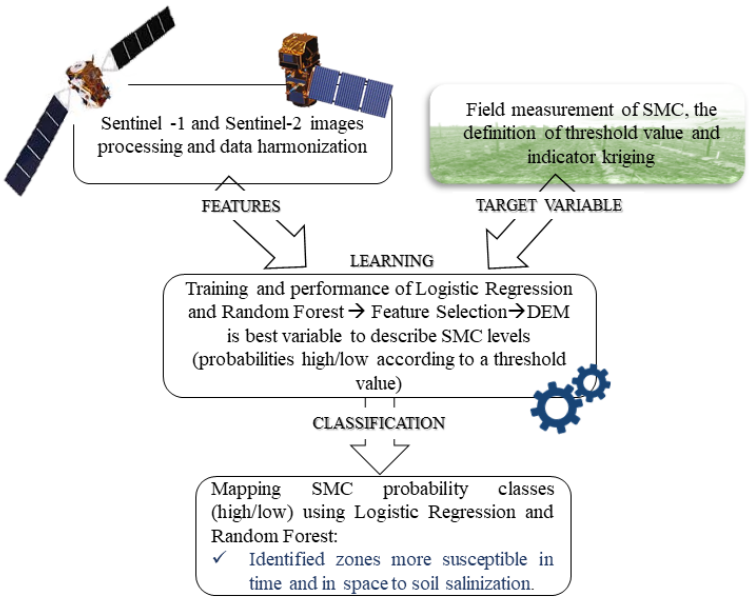

Figure 5. Workflow diagram of the methodology used in this study to map the soil moisture content (SMC) classes

the fieldwork, highlighting the importance of a good coordination between the satellite imagery acquisition and monitoring period. In Jan-2018, the RF still had a good performance (Table 4) with the DEM, two SAR images S1BA270118 (ascending) and S1AD260118 (descending) and S2250118-NIR being selected (Table 4). For the time of this campaign, the vines were completely leafless, two S1 images were chosen as the best features to classify the SMC and justifying the lower importance of the optical features. Looking at Table 2, an increase in the median value of the NIR band reflectance can be observed, which may be due to the decrease in the overall soil moisture (Amani et al. 2016). The RF had the worst performance (*F*-measure = 0.35) in May-2018 (grapevine flowering period), where all the SMC values of plot A were above the SMC of 14%, showing a low oob error of 12.6%, which reflects the class-imbalance. In this campaign, the vines were already covered by leaves, somehow explaining the selection of the optical images. Furthermore, due to reasons related to field maintenance works, it was not possible to collect the samples on the same day

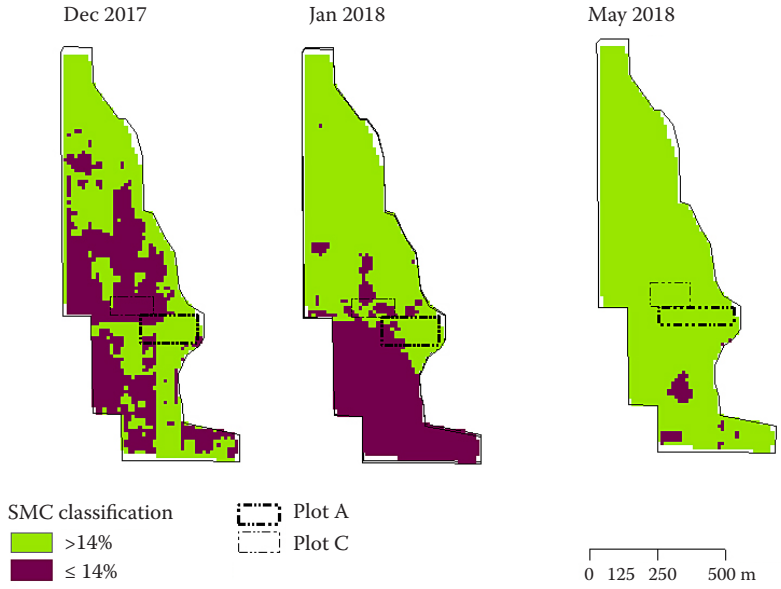

Figure 6. Results of the classification of the soil moisture content (SMC) using the Logistic Regression method for the three monitoring dates (Dec-17, Jan-18, and May-18)

https://doi.org/10.17221/101/2019-SWR

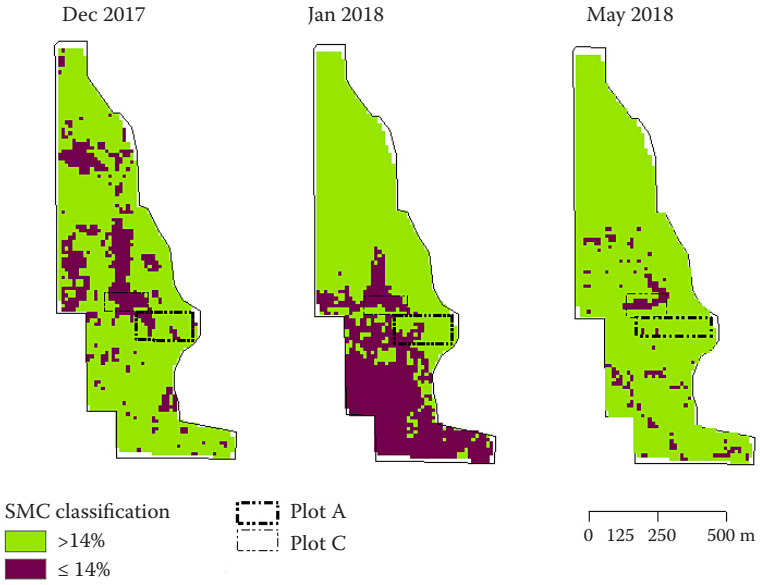

Figure 7. Results of the classification of the soil moisture content (SMC) using the Random Forest method for the three monitoring dates (Dec-17, Jan-18, and May-18)

as the satellite image acquisition, which may have compromised the results.

Mapping soil moisture content classes (below and above 14%). In this section, the mapping of the SMC classes (high or low) is presented. Figure 5 shows the workflow diagram of the methodology adopted to perform this mapping.

In Dec-2017, the LR classified 45% of the vineyard with SMC occurrence probabilities of an SMC below 14% (Figure 6), increasing the dry area to the south (around 62%) in Jan-18 and decreasing in May-18 (2.6%; Figure 7). Figure 7 shows that the areas of an SMC ≤ 14% obtained by RF were less extensive than those delineated by LR. In fact, in Dec-17, the area delimited by RF as having an SMC ≤ 14% was 45% less than the one obtained by LR and, in Jan-18, 58% smaller than the dry area classified by LR.

Figure 8 presents the SMC classification of the vineyard by RF according to the occurrence frequency of an SMC < 14% during the monitoring period. The map shows that the central part of the vineyard can be more susceptible to soil salt accumulation, as well as some small areas in the south. Although a flat area, these lower SMC zones were delineated passing along the contour lines, with higher altitudes (southwest area).

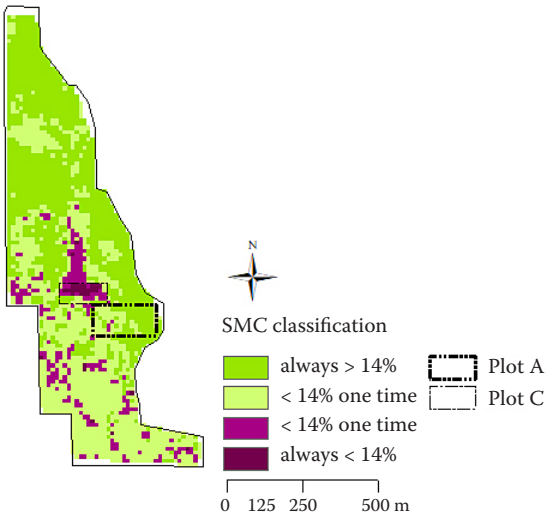

Figure 8. Synthesis map derived from the Random Forest method representing the occurrence of the soil moisture content (SMC) below or equal to 14% during the three monitoring dates (Dec-17, Jan-18, and May-18)

CONCLUSION

A cost-effective approach, comprehending remotely sensed imagery and field measurements, was tested to delineate the zones in a vineyard with a low SMC throughout the winter and early spring, demonstrating that: (*i*) The RF method outperformed the widely used LR one; (*ii*) After the drought period, the RF method had its best performance; (*iii*) The central part of the vineyard can be more susceptible to soil salt accumulation, as well as some small areas in the south with higher altitudes; (*iv*) The temporal and spatial behaviour of the probabilities of an SMC below the threshold value was established on a vineyard scale, allowing one to systematically cross-check the zones where the SMC can be very low and, therefore, more prone to soil salinisation. This work points out the importance of a multi-sensor approach (different wavelengths), since each one has a different

https://doi.org/10.17221/101/2019-SWR

sensitivity to variables in the field, bringing all their integration benefits to this assessment.

Acknowledgement: The authors would like to thank Companhia das Lezírias, S.A. (Benavente, Portugal) for the development of this study, namely Eng. Rui Alves (Coordinator of the Department of Forests and Natural Resources) and winemaker Eng. Bernardo Cabral.

REFERENCES

Adhikari K. (2008): Updated map of salt affected soils in the European Union. In: Montanarella L., Rusco E., Tóth G. (eds.): Threats to Soil Quality in Europe. Luxembourgh, JRC European Commision.

Amani M., Parsian S., MirMazloumi S.M., Aieneh O. (2016): Two new soil moisture indices based on the NIR-red triangle space of Landsat-8 data. International Journal of Applied Earth Observation and Geoinformation, 50: 176–186.

Aragüés R., Medina E.T., Clavería I., Martínez-Cob A., Faci J. (2014): Regulated deficit irrigation, soil salinization and soil sodification in a table grape vineyard drip-irrigated with moderately saline waters. Agricultural Water Management, 134: 84–93.

Bishop C.M. (1995): Neural Networks for Pattern Recognition. New York, Oxford University Press, Inc.

Blum A.L., Langley P. (1997): Selection of relevant features and examples in machine learning. Artificial Intelligence, 97: 245–271.

Bousbih S., Zribi M., Lili-Chabaane Z., Baghdadi N., El Hajj M., Gao Q., Mougenot B. (2017): Potential of sentinel-1 radar data for the assessment of soil and cereal cover parameters. Sensors (Basel), 17: 2617.

Breiman L. (2001): Random forests. Machine Learning, 45: 5–32.

El Hajj M., Baghdadi N., Bazzi H., Zribi M. (2019): Penetration analysis of SAR signals in the C and L bands for wheat, maize, and grasslands. Remote Sensing, 11: 31.

FAO (2018): Handbook for Saline Soil Management. Food and Agriculture Organization of the United Nations and Lomonosov Moscow State University.

Ghasemian N., Akhoondzadeh M. (2018): Introducing two Random Forest based methods for cloud detection in remote sensing images. Advances in Space Research, 62: 288–303.

Gherboudj I., Magagi R., Berg A.A., Toth B. (2011): Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. Remote Sensing of Environment, 115: 33–43.

Hartwich R., Baritz R., Fuchs M., Krug D., Thiele S. (2005): Die Bodenregionenkarte der Europäischen Union und ihrer Nachbarstaaten im Maßstab 1 : 5 000 000 (Version 2.0). Hannover.

Hosmer D.W., Lemeshow S. (2000): Applied Logistic Regression. 2nd Ed. New Jersey, John Wiley & Sons, Ltd.

Kotsiantis S.B. (2007): Supervised machine learning: A review of classification techniques. In: Maglogiannis I., Karpousiz K. (eds.): Proc. 2007 Conf. Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies. Amsterdam, IOS Press: 3–24.

Kurnik B., Kajfež-Bogataj L., Horion S. (2015): An assessment of actual evapotranspiration and soil water deficit in agricultural regions in Europe. International Journal of Climatology, 35: 2451–2471.

Liaw A. (2018): Breiman and Cutler's Random Forests for Classification and Regression. Documentation for Package 'RandomForest' 29. R package. Available at https://cran.r-project.org/web/packages/randomForest/randomForest.pdf.

Liu X.Y., Wu J., Zhou Z.H. (2009): Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man and Cybernetics – Part B (Cybernetics), 39: 539–550.

Mendes M.P., Ribeiro L., David T.S., Costa A. (2016): How dependent are cork oak (*Quercus suber* L.) woodlands on groundwater? A case study in southwestern Portugal. Forest Ecology and Management, 378: 122–130.

Rodriguez-Galiano V.F., Luque-Espinar J.A., Chica-Olmo M., Mendes M.P. (2018): Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. Science of the Total Environment, 624: 661–672.

Roth C.H., Malicki M.A., Plagge R. (1992): Empirical evaluation of the relationship between soil dielectric constant and volumetric water content as the basis for calibrating soil moisture measurements by TDR. European Journal of Soil Science, 43: 1–13.

Santi E., Paloscia S., Pettinato S., Brocca L., Ciabatta L., Entekhabi D. (2018): Integration of microwave data from SMAP and AMSR2 for soil moisture monitoring in Italy. Remote Sensing of Environment, 212: 21–30.

Tharwat A. (2020): Classification assessment methods. Applied Computing and Informatics, 17: 168–192.

Vicente-Serrano S.M., Beguería S., López-Moreno J.I. (2010): A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. Journal of Climate, 23: 1696–1718.

Yang Y., Loog M. (2018): A benchmark and comparison of active learning for logistic regression. Pattern Recognition, 83: 401–415.

Received: August 16, 2019
Accepted: November 16, 2020
Published online: December 17, 2020