Sentinel-2 images to assess soil surface characteristics over a rainfed Mediterranean cropping system

Cécile Gomez, Maman Sani Aboubacar, Dino Ienco, Denis Feurer, Zakia Jenhaoui, Attia Rafla, Maguelonne Teisseire, Jean-Stéphane Bailly

Sentinel-2 images to assess soil surface characteristics over a rainfed Mediterranean cropping system

Cécile Gomez[1]*, Maman Sani Aboubacar[1,2], Dino Ienco[3], Denis Feurer[1], Zakia Jenhaoui[4], Attia Rafla[5], Maguelonne Teisseire[3], Jean-Stéphane Bailly[1,6]

[1] LISAH, Univ. Montpellier, INRAE, IRD, Institut Agro Montpellier, Montpellier, France.
Corresponding author. cecile.gomez@ird.fr
[2] LIRMM, University of Montpellier, CNRS, Montpellier, France
[3] TETIS, Univ. Montpellier, IRSTEA, Montpellier, France
[4] IRD, Tunis, Tunisia
[5] DG-ACTA, Ministry of Agriculture, Tunis, Tunisia
[6] AgroParisTech, Paris, France

Abstract: Soil surface characteristics (SSCs) are of high importance for water infiltration processes in crop fields. As SSCs present strong spatiotemporal variability influenced by climatic conditions and agricultural practices, their monitor has already been explored by using UAV images and multispectral remote sensing. However, each technique has encountered difficulties characterizing this spatiotemporal variability. The objective of this work was to explore the potential of Sentinel-2 images to assess three SSCs - the green vegetation fraction, dry vegetation fraction and physical soil surface structure - at several dates. This work explored two approaches for classifying these three SSCs from five Sentinel-2 images acquired from August to November 2016. In the “single-date” approach, a Random Forest Classifier (RFC) model was trained to classify one SSC_j from a dataset extracted from one Sentinel-2 image i (model noted $RF\_sd_{i,SSCj}$). In the “multi-date” approach, a RFC model was trained to classify one SSC_j from a dataset extracted from the five Sentinel-2 images (noted RF_md_{SSCj}). The classification analysis showed that i) the $RF\_sd_{i,SSCj}$ and RF_md_{SSCj} models provided accurate performances (overall accuracy > 0.79) regardless of the studied SSC_j and the tested Sentinel-2 image, ii) the $RF\_sd_{i,SSCj}$ model did not allow the classification of SSC classes that were not observed on the studied date, and iii) the RF_md_{SSCj} model allowed the classification of all SSC classes observed in the five Sentinel-2 images. This indicated that several Sentinel-2 images can favourably be used to increase knowledge of spatiotemporal representation of SSCs by extending results of infrequent, localized and cumbersome field work.

Keywords: Soil surface characteristics; green and dry vegetation; physical soil surface structure; Sentinel-2; classification.

1. Introduction

Soil infiltration is one of the most important earth surface processes controlling the water budget equation. It controls the water cycles among surface-water and allows the soil to temporarily store water, making water available for uptake by plants and soil organisms. Soil infiltration may substantially affect a series of ecological processes including water supply for plant growth and groundwater recharge (Ludwig et al., 2005), solute transport to deep soil and groundwater (Jarvis, 2007), and the development of surface runoff and soil erosion (De Roo et al., 1992).

Soil infiltration characteristics, commonly represented by macroscale parameters such as soil hydraulic conductivity, sorptivity and infiltration rate can be measured directly on field (e.g., Mubarak et al., 2010). Nevertheless in-situ infiltration parameters could be difficult to measure precisely as some environmental factors, such as temperature, humidity and initial soil water content, could change during the time-consuming infiltration measurements (Mubarak et al., 2010). Moreover direct measurements of infiltration characteristics are time consuming, expensive and often involve large spatial and temporal variability (Mishra et al., 2003). Soil infiltration characteristics can be also indirectly estimated using soil surface characteristics (SSCs; i.e. surface crust development, roughness, vegetation cover, texture…) as inputs of pedotransfer functions (Ghorbani-Dashtaki et al., 2016), such the ones developed by Børgesen et al. (2008), Rashidi et al. (2014) or Patle et al. (2019).

The construction of these pedotransfer functions needs to know the strong relations between SSCs and hydrological processes, but researches reached a consistent conclusion that the links between SSCs and hydrological processes are site-specific (e.g., Bormann and Klaassen, 2008). Yimer et al., 2008 showed that principal factors causing the decline in infiltration capacity in the Bale Mountains National Park in Ethiopia are the changes in topsoil structure caused by surface soil compaction because of tillage and animal trampling coupled with a smaller soil organic carbon content. Joshi and Tambe (2010) showed that infiltration rate in Western India vary from subtle to noteworthy depending on slope angle, grass coverage, crop residue and gravels. Neris et al. (2012) showed that infiltration rate in volcanic island of Tenerife (Canary Islands, Spain) is highly dependent to soil aggregation and structural stability. Leonard and Andrieux (1998) showed that the major SSCs that drive infiltration processes in Mediterranean areas are green and dry vegetation coverage, topsoil structure, surface stone content, and soil texture (i.e., relative contents of particles of various sizes, such as sand, silt and clay).

Except soil texture which can be characterized by high spatial variability but low temporal variability and may therefore be considered as a permanent property, other SSCs that impact soil infiltration (namely, green and dry vegetation coverage, topsoil structure, surface stone content) are changing in time and space depending on climatic conditions (Chahinian et al., 2005) and agricultural practices (e.g., tillage, seeding, plant growing, maturity and harvesting) (Van 1993; Martin et al., 2004; Bormann and Klaassen, 2008). So, the characterization of soil infiltration processes requires monitoring of the SSCs in both space and time at the plot resolution.

The need for spatial SSCs characterization could be addressed by the use of visible, near-infrared and short-wave infrared (VNIR/SWIR) remote sensing data, as this technology provides synoptic coverage at a single date. VNIR/SWIR multispectral imagery has been used for mapping SSCs with different degrees of success levels. The green vegetation fraction is usually successfully mapped by using the normalized difference vegetation index (NDVI) (e.g., Zhang et al, 2006), which can be calculated using the red and near-infrared bands measured by the multispectral VNIR/SWIR sensors (e.g., Carlson and Rizile, 1997).

The dry vegetation fraction has a unique absorption feature near 2100 nm associated with cellulose and lignin (Daughtry 2001), but most of the multispectral VNIR/SWIR sensors do not allow the use of this specific absorption feature. The Normalized Difference Tillage Index (NDTI) was demonstrated to be the best of the Landsat-based tillage indices for estimating residue cover, exploiting the difference in reflectance between the two Landsat shortwave infra-red (SWIR) bands centered near 1600 nm and 2300 nm (Deventer et al., 1997). The ASTER bands have been used with success to compute advanced multispectral residue indices such as the Shortwave Infrared Normalized Difference Residue Index (SINDRI) (e.g., Serbin et al., 2019). Finally, the SINDRI was demonstrated to provide better accuracy than the Lignin Cellulose Absorption Index (LCA) for estimating residue cover, exploiting the WordView data (Hively et al., 2018). The soil texture influences both the spectral intensity and absorption band depth at 2200 nm (e.g., Clark et al., 1990; Gomez et al., 2012) and can be mapped using a linear regression built based on the entire spectra (Vaudour et al., 2019) or on a spectral index using SWIR bands (Shabou et al., 2015). The topsoil structure may influence the general shape of the spectrum as crust, cracked clay and roughness may influence the surface colour, brightness and surface structure (e.g., Matthias et al., 2000; Ben-Dor et al., 1999); but, from our knowledge, the topsoil structure has not been studied with VNIR/SWIR multispectral data. Finally, VNIR/SWIR multispectral imagery has been successfully used to map the typology of the hydrological SSC classes according to a predefined typology based on the infiltration rates instead of mapping the single SSC attributes with multispectral images acquired by unmanned aerial vehicles (e.g., Corbane et al., 2008).

The need for SSCs monitoring could be addressed by the use of remote sensing image time series. The remote sensing image time series are mostly used as a chronicle of data in which the temporal patterns of the spectral response are considered as inputs to characterize elements of the land surface (e.g., crop type and land use management) (e.g., Bellón et al., 2017; Wang et al., 2019; Vuolo et al., 2018). Following this approach, Belgiu and Csillik (2018) took into account the temporal sequences in a time series of Sentinel-2 data (the order of the Sentinel-2 data acquisition dates) to extract the temporal phenological patterns and then classify the croplands. Their classifier inputs were the NDVI time series, which were considered temporal phenological patterns, and the outputs were cropland maps for each study area and for the entire selected period. Another example is the study of Rapinel et al. (2019), which attempted to map floodplain grassland plant communities using a time series of Sentinel-2 data (without considering the order of the Sentinel-2 data acquisition dates) and a random forest method. Another approach consists in using remote sensing image time series to detect changes between two dates or during a period (e.g., Navarro et al, 2017), for example based on differences in spectral indices (e.g., NDVI) between images. Following this approach, Sicre et al. (2016) successfully used a time series of FORMOSAT and SPOT data for summer crop detection based on a decision tree using thresholds on NDVI values. Finally, another approach consists in using remote sensing image time series as a succession of single-date remote sensing images, where each remote sensing data acquisition is treated separately. The characterization of earth surface elements (e.g., green vegetation stages, Vuolo et al. 2018) may be done at each date with each single-date remote sensing data. From our knowledge about the SSCs that impact the hydrological processes, only the green vegetation fraction has been studied both in time and space using time series VNIR/SWIR multispectral data, such as the Chinese GF-1 data used by Jia et al. (2016) and the LANDSAT data used by Jia et al. (2017).

The objective of this work was to explore the potential of the Sentinel-2 images to assess three SSCs - the green vegetation fraction, dry vegetation fraction and physical soil surface structure - at several dates. The study area and data are presented in section 2. The methodology was described in section 3 based on random forest supervised classification trained with field data. Finally, the results are presented in section 4 and discussed in section 5.

2. Materials and Methods

2.1 Study area

The study area is the Kamech catchment (2.63 km^2) located on the Cap Bon peninsula in north-eastern Tunisia (Figure 1) with a semi-arid climate. The Kamech catchment belongs to the long-term environmental research observatory OMERE (Mediterranean Observatory of Water and Rural Environment), which aims to investigate the anthropogenic impacts on water and sediment budgets at the catchment scale (Molénat et al., 2018). The Kamech catchment is characterized by rolling hills with a maximum drop of 110 m. The substrate within the catchment formed from Miocene marine sediments, and is mainly composed of alternations of slightly calcareous laminated mudstone and thin hard sandstone layers. The soils were developed both over and from the Miocene deposits. The main soil types include Calcil or Chromic Vertisols (52.5%), Clayic Calcisols (22%), Vertic Regosols (17%), Leptosols (5%) and Colluvic Cambisols (3.5%), according to the FAO classification (WRB, 2015). These soils are characterized by a narrow and low range of soil organic matter content (from 0.3 to 2 %), a moderate range of soil calcium carbonate (from 0.2 to 19.9 %) and a large range of clay content (from 12 to 67 %) (Molenat et al., 2018). This area is mainly rural (> 95%) and is devoted to cereals cultivation in addition to legumes and fodder for animals. Cultivation practices throughout the Kamech catchment are representative of traditional agriculture in the relief zone of Cap Bon peninsula.

[Figure 1]

Within the study area, crop emergence occurs between October and December, depending on farmers and meteorological conditions. Agricultural practices between the harvest and the growth of the new crop include several steps: 1) the harvest lets surface dry vegetation (litter and/or crop stubble) (July to September), 2) a first surface ploughing is conducted after the first rain (often in October over the Kamech catchment), 3) a second deeper ploughing is conducted, approximately 15 days after the first ploughing, 4) the crops are seeded and then 5) the crop is grown.

Finally, with inter-annual precipitation of 600 mm a most intense precipitations occur between September and December (more than 350 mm) and lower precipitations occur during the remaining months, with a very dry summer.

2.2 Field data

2.2.1. Field boundaries and land use map

The field boundaries and a land use map for Kamech were produced in 2016 through field work with a handheld GPS (Figure 1c) (Jenhaoui et al., 2008). The observed land uses were annual crops, natural vegetation, olive and fruit tree plantations, lakes, urban areas and roads. The total number of crops over the study area is 384. These 384 fields are characterized by a minimum, maximum and mean plot area of approximately 0.03, 1.4 and 0.59 ha, respectively (Table 1).

[Table 1]

2.2.2. Soil Surface Characteristic (SSC) observations

Starting in 2003, the SSCs were routinely observed in 34 plots of the Kamech catchment dedicated to annual crops (red, blue and green fields on Figure 1c). The SSCs observations follow a protocol initially proposed by (Andrieux et al., 2001), then adapted to the Kamech catchment in Tunisia (Molénat et al., 2018). The protocol was initially based on works developed by Leornard and Andrieux (1998) dedicated on the Roujan catchment (91 ha), also belonging to the long-term environmental research observatory OMERE (Molénat et al., 2018), which is located in Southern France about 60 km West of Montpellier, in a Mediterranean context mostly devoted to vineyard culture. The 34 observed fields were selected based on hydrological purposes, and the selected SSCs and their associated ranges were chosen in regard to their effect on hydrological processes such as infiltration rate and runoff generation (Leornard and Andrieux, 1998; Pare et al., 2011).

The SSCs were routinely observed every 2 weeks on average during the September-July period. As these SSCs field observations are dedicated to hydrological studies, e.g. runoff and infiltration (Leonard et Andrieux, 1998; Pare et al., 2011), observations timing is adapted to meteorological conditions and farmers practices. No observations are conducted in August because all crops are harvested and neither management practices nor rains occur during this month. The field observations dates result from a trade-off between field accessibility after rainfall events and known agricultural practices, including harvest (occurring in July), ploughing (occurring from the first rains around October), and seeding (occurring after soil ploughing, around November) until crop growth (occurring from mid-November). From beginning of August to end of December 2016, six SSCs field observations were done by the same operator and five were used in this work (Table 2).

Each SSC was described at field scale by the same operator by visual inspection, where a field is an area of land used for one specific crop per cultivated season (Figure 1c). Only one class was written down per field per SSC, regardless of the field size. When a single field is composed by several classes of SSC, the operator writes down the majority class of this SSC. First, the operator has to observe elements characterizing the soil, such as ploughing, physical soil surface structure and roughness. Second, the operator has to observe elements characterizing the soil coverage such as vegetation fraction and coarse elements cover.

Three major SSCs were studied in this work: 1) the green vegetation fraction, 2) the dry vegetation fraction and 3) the physical soil surface structure. The green vegetation fraction was observed within the following six classes: green vegetation fraction of 0% (which means total absence of green vegetation), 0% < green vegetation fraction ≤ 5%, 5% < green vegetation fraction ≤ 25%, 25% < green vegetation fraction ≤ 50%, 50% < green vegetation fraction ≤ 75% and 75% < green vegetation fraction ≤ 100%. The dry vegetation fraction was observed within the following six classes: dry vegetation fraction of 0% (which means total absence of dry vegetation), 0% < dry vegetation fraction ≤ 5%, 5% < dry vegetation fraction ≤ 25%, 25% < dry vegetation fraction ≤ 50%, 50% < dry vegetation fraction ≤ 75% and 75% < dry vegetation fraction ≤ 100%. The physical soil surface structure was observed within six classes: dry soil surface without crust, aggregate or clod, mainly observed after a recent tillage (noted F0); surface with fine and continuous crust, mainly observed after some moderate rainfall without water flow (noted F1); surface with crust, mainly observed after a heavy rain (may happen around October-November) or after a long period of dry climate (may happen from August to end of September) (noted F2); saturated soil (called saturated); and two intermediate classes which express transient states (between classes F0 and F1, noted F0/F1 and between classes F1 and F2, noted F1/F2).

Among the 34 observed fields, twenty-three fields belong to a sub-catchment highly observed (network of five hydrometric stations equipped with flumes) because of its high runoff process and erosion, which is located in the Western side of the Kamech catchment (red polygons, Figure 1c). Seven fields belong to the Eastern side of the Kamech catchment (blue polygons, Figure 1c). The remaining four fields are located on the top North of the Kamech catchment (green polygons, Figure 1c). These 34 observed fields were characterized by a large diversity of shapes (Figure 1c) and a low diversity of sizes (from 0.07 to 1.27 ha, with a mean of 0.51 ha, Table 1). The SSCs were described by the same operator along the crop season.

[Table 2]

2.3 Remote Sensing data

The ESA's Sentinel-2A satellite was launched on the 23th of June 2015. The satellite orbits at an altitude of 786 km and has a swath width of 290 km. In 2016, it acquired multispectral data with a revisit of 10 days in 13 bands covering the visible, NIR and SWIR spectral domain with spatial resolutions ranging from 10 to 60 m. The three bands acquired at 60 m spatial resolution (coastal at 443 nm, water vapour at 945 nm and cirrus at 1380 nm) were only used to perform atmospheric corrections and cloud detection. For each date, the Level 2A Sentinel-2 data were corrected from atmospheric effects using the MACCS (Multi-sensor Atmospheric Correction and Cloud Screening) algorithm (Hagolle et al., 2015; Baetens et al., 2019), taking into account adjacency effects and illumination variations due to topography. MACCS was specifically designed to process time series of optical images at high resolution, acquired under quasi constant viewing angles. Output data from MAACS algorithm were orthoimage Bottom-of-Atmosphere corrected reflectance images and were obtained from the French space agency website (CNES, theia.cnes.fr). The six spectral bands initially acquired with 20 m spatial resolution were resampled to 10 m. We used the function "disaggregate" provided in the raster package (Robert, 2019) in R version 3.2.1 (R Development Core Team, 2015). So the values in the resampled bands are the same as in the larger original cells. Finally, the natural vegetation, olive and fruit tree plantations, lakes, urban areas and roads were masked over each Sentinel-2 data using the land use map (section 2.2.1) to keep only the 384 fields dedicated to annual crops. After this mask process, the 384 fields which have to be classified represent 199 698 Sentinel-2 pixels over each Sentinel-2 image. A total of 1264 pixels are associated to observed SSCs, based on the survey over the 34 cultivated plots of the Kamech catchment. As the 34 observed fields were characterized by a mean, minimum and maximal size of 0.51 ha, 0.07 ha and 1.27 ha (Table 1), respectively, from around 5 to 120 pixels were considered per field. So depending on the field size, the SSC observations done at each date of field observation (Table 2) were representative of 5 to 120 pixels.

The dates of Sentinel-2 images (Table 2) were chosen to fit the period of agricultural practices realized after the harvest (July) and the summer season and until crop growth (December). From beginning of August to end of December 2016, fifteen Sentinel-2 were acquired over our study area. Among these fifteen acquisitions, five images had less than 10% cloud over the Kamech catchment and were kept.

The Sentinel-2 image acquired on the 4th of August 2016 was considered adequate to align the field observations acquired on the 1st of September 2016 (Table 2) because neither agricultural practices nor rainfall happened in August. Additionally, the Sentinel-2 image acquired on the 3rd of October 2016 was considered adequate to align with the field observations acquired on the 28th of September 2016 (Table 2). The other Sentinel-2 images were acquired with a maximum delay of 1 day relative to the field observations (Table 2).

3. Methods

This work explored two approaches to classify the targeted SSCs. Differently from discrimination that attempts to separate distinct sets of objects, classification attempts to allocate new objects to predefined groups (labels). A classification model (machine learning approach) is firstly calibrated on a training set that involves examples already labelled with class information and, successively it is deployed to perform classification of new unlabelled data. To summarize, the main objective of a classification task is to categorize unlabelled data in a predefined set of known classes. This paper explored two approaches to classify the classes of targeted SSCs:

i) In the "single-date" approach, a Random Forest Classifier (RFC) model is trained to classify one SSC_j, based on pixels extracted from one Sentinel-2 image acquired at t_i (Figure 2A, steps 1 and 2). Once trained, the RFC model was then applied to this Sentinel-

2 image acquired at t_i (Figure 2A, step 6). Following this "single-date" approach, a RFC model was built for classifying each SSC_j and trained from each Sentinel-2 image. As three SSCs have to be classified at the five Sentinel-2 dates, fifteen RFC models were trained in the "single-date" approach. These RFC models would be noted $RF\_sd_{i,SSCj}$ where i is the Sentinel-2 date (Table 2) and SSC_j is the SSC predicted by the model.

ii) In the "multi-date" approach, a RFC model is trained to classify one SSC_j, based on pixels extracted from the five Sentinel-2 images (Figure 2B, steps 1 and 2). Once trained, the RFC model was then applied to the five Sentinel-2 images (Figure 2B, step 6). Following this "multi -date" approach, a RFC model was built for classifying each SSC_j and trained from the five Sentinel-2 images. As three SSCs have to be classified, three RFC models were trained in the "multi-date" approach. These RFC models would be noted RF_md_{SSCj} where SSC_j is the SSC predicted by the model. In this approach, the five images were used to train the models without considering the chronological order of images.

The aim of using the "multi-date" approach compared to the "single-date" approach is to increase the training data, in term of both number of predictors and number of labelled pixels, compared to the "single-date" approach.

The classification models were developed in R version 3.2.1 (R Development Core Team, 2015) using the caret package (Kuhn et al., 2016).

[Figure 2]

3.1 Random Forest

The random forest (RF) takes part of the ensemble machine learning techniques. The random forest was developed by Breiman (2001) and demonstrated as the best classifier among 179 classifiers arising from 17 families tested by Fernández-Delgado (2014). Its effectiveness in remote sensing has been demonstrated due to its robustness (e.g., Ok et al., 2012). The RF produces a large number of classification trees that contribute via a voting system to classify data (Kuhn et al., 2016). As part of the ensemble machine learning techniques, RF has higher accuracy than single classifiers as a group of classifiers performs more accurately than any single classifier (Ok et al., 2012) and RF is considered as efficient and effective even with non-normally distributed training data set (Rodriguez-Galiano et al., 2012) which is the case of our datasets. Belgiu and Dragu (2016) proposed a review of the limitations and advantages of the algorithm.

Two parameters need to be tuned: the number of trees (ntree parameter), which are created by randomly selecting samples from the calibration samples, and the number of variables used to split each tree node (mtry parameter). As reported by Belgiu and Dragu (2016), most studies are performed using an ntree value of 500 because the errors are stable before this number of classification trees is achieved. So we selected an ntree value of 500 to fit with the outputs of Belgiu and Dragu (2016). Belgiu and Dragu (2016) also reported that the mtry parameter is usually set to the square root of the number of input variables. So we tested 8 values of mtry ranging between 2 and 10, and the optimal value was defined for the best overall accuracy obtained when performing a 10-fold cross-validation on the calibration dataset.

3.2 Calibration and Validation dataset

In the "single-date" approach, the full dataset is composed of the 1264 pixels extracted from one Sentinel-2 image acquired at t_i (Figure 2A, step 2). The $RF\_sd_{i,SSCj}$ models were trained on a subset of 70% of this full dataset (i.e. 884 pixels), while the remaining 30% (i.e. 380 pixels) was used to test the performance of the model (Figure 2A, steps 3 and 5). The split between training and test datasets was done following a stratified random sampling. Thanks to this stratified random sampling, the calibration and test datasets are characterized by a similar distribution of the targeted SSC_j.

In the "multi-date" approach, the full dataset is composed of 6320 pixels extracted from the five Sentinel-2 images (1264 pixels extracted per Sentinel-2 image) (Figure 2B, step 2). The RF_md_{SSCj} model were trained on a subset of 70% of this full dataset (i.e. 4420 pixels), while the remaining 30% (i.e. 1900 pixels) was used to test the performance of the models (Figure 2B, steps 2 and 3). The split between training and test datasets was done following a stratified random sampling, providing a similar distribution of the targeted SSC_j in the calibration and test datasets.

3.3 Accuracy assessment

The overall accuracy and kappa coefficients, calculated on test data, were used to measure the performance of the RF classifications (Cohen, 1960). Overall accuracy is commonly measured as the percentage of pixels correctly classified in the validation dataset. The kappa coefficient compares the observed accuracy with the expected accuracy resulting from randomness. The kappa statistics are used to assess the proportion of the results that is due to pure randomness, especially when classes with few individuals occur in the classification process. A kappa coefficient of 1 indicates perfect classification, and a kappa coefficient of 0 corresponds to a random classification (Congalton, 1991). Based on Congalton and Green (1999), kappa values greater than 0.80 represent strong agreement between the classification results and ground truth data, kappa values between 0.4 and 0.8 represent moderate agreement, and kappa values below 0.4 represent poor agreement. The 95% confidence intervals (95 % CI) of the overall accuracy were also calculated.

Accuracies of individual class were calculated in a similar way than overall accuracy. The producer's accuracy was used to indicate the probability of a reference pixel being correctly classified (Story and Congalton, 1986). The producer's accuracy for class *A* was calculated as the ratio between the number of pixels correctly classified in class *A* and the total number of reference pixels (ground true) for that class *A*. And user's accuracy was used to indicate the probability that a pixel classified on the map represents the class on the field (Story and Congalton, 1986). The user's accuracy for class *A* was calculated as the ratio between the number of pixels correctly classified in class *A* and the total number of pixels classified in class *A*.

The overall, producer's and user's accuracies, 95 % CI and Kappa-coefficient were calculated with Caret R package (Breiman, 2001) by using *confusionMatrix* function.

3.4 Classification mapping

In the "single-date" approach, after validating a $RF\_sd_{i,SSCj}$ model on the corresponding test dataset for predicting the SSC_j (Figure 2A, step 5), the $RF\_sd_{i,SSCj}$ model was applied to the entire Sentinel-2 image acquired at date t_i (Figure 2A, step 6) providing one classification map of the targeted SSC_j for the date t_i.

In the "multi-date" approach, after validating a RF_md_{SSCj} model on the test dataset for predicting the SSC_j (Figure 2B, step 5), the RF_md_{SSCj} model was applied to the five Sentinel-2 images (Figure 2B, step 6) providing five classification maps of the targeted SSC_j (each classification map corresponding to one date of Sentinel-2 acquisition).

3.5 Classes aggregation from pixel to field scale

For each classification map, the classes affected to pixels were secondly aggregated at the field scale using field boundaries (Figure 1c). The class labelling process for a field intersecting a collection of pixels was fixed as the most frequent pixel class.

As it is expected to get similar class of each SSC at field scale, and as a highest frequent class of pixels within a field may just result from pure randomness, an indicator of the non-randomness of the most frequent pixel class was computed at field scale. The selected indicator for a given field *j* was the probability value (*P-value*) resulting from a chi-squared test, where *P-value*$_j$ denotes the probability of the *chi2* variable under pure random process (*H0*) with *dl* degree of freedom for the given field *j.* The *chi2* variable with one degree of freedom for any field *j* composed of *n* pixels is computed as follows:

$$Chi2_j = \frac{\left(\hat{F_j}^2 - F\right)^2}{F^2} \qquad (1)$$

where $\hat{F_j}$ is the frequency of the most frequent class for field *j* and F is the theoretical one resulting from a binomial random law, knowing the overall proportion of this class against others at the entire image scale. For a given field *j*, the *P-value*$_j$ lower than 0.05 indicates that the observed higher frequency in a field is significant and does not result from randomness.

4. Results

4.1 Preliminary analysis of observed SSCs

The distribution of the observed classes did not follow a normal distribution, regardless of the date and the SSC (Figure 3). The green vegetation fraction mainly varied because of the tillage and secondarily the meteorological conditions that drove vegetation growth. Only one class was observed in August (0%, Figure 3a) as crops were harvested between June and July, and the dry and hot weather during this period prevented any grass growth. From September, the number of observed green vegetation fraction classes increased over time to reach 6 classes in December (Figure 3a) after seeding and crop emergence. During the selected period, only four classes were represented for the dry vegetation fractions: 5% < vegetation fraction ≤ 25%, 25% < vegetation fraction ≤ 50%, 50% < vegetation fraction ≤

75% and 75% < vegetation fraction ≤ 100% (Figure 3b). The dry vegetation fraction varies due to vegetation decomposition and the management practices, such as tillage. The four classes of dry vegetation were never observed on the same date (Figure 3b). From August to December, the number of dry vegetation fraction classes decreased over time to reach only one class in December (0%, Figure 3b) after tillage, seedling and crop emergence. The physical soil surface structure varied mainly due to rainfall and secondarily as a result of management practices. The six classes of the physical soil surface structure were never observed on the same date (Figure 3c). Most of the fields were characterized by a transient state of F1/F2 whatever the date. And numerous saturated fields were observed in 21th of November and 2^{sd} of December.

Green and dry vegetation fraction were inversely correlated as in August, the absence of green vegetation is associated to a high proportion of dry vegetation (Figure 3a and b), and then more the green vegetation fraction increase, more the dry vegetation fraction decreases. Whereas the soil surface structure was not correlated to the vegetation fraction.

[Figure 3]

4.2 Classification models performances

Among the fifteen $RF\_sd_{i,SSCj}$ models initially planned to be built, the $RF\_sd_{1,green}$ and $RF\_sd_{5,dry}$ models have not enough classes to be trained (only one class was observed, Figure 3a and 3b) so these two models were not built. The remaining thirteen $RF\_sd_{i,SSCj}$ models were trained from their dedicated training dataset and tested on their dedicated test datasets (Figure 2A, steps 2, 3 and 4). Nevertheless, when the training and test datasets are very unbalanced, the models performances must be considered carefully, as for $RF\_sd_{2,dry}$, $RF\_sd_{3,dry}$, $RF\_sd_{4,dry}$ and $RF\_sd_{3,struc}$ (Figure 3b and c; in grey and italics in Table 3). Considering the almost-balanced training and test datasets, only nine $RF\_sd_{i,SSCj}$ models can be explored. The $RF\_sd_{2,green}$ and $RF\_sd_{1,struct}$ provided the highest performances with an overall accuracy and kappa of 0.93 and 0.82, respectively, whereas the $RF\_sd_{3,green}$ provided the lowest overall accuracy and kappa of 0.84 and 0.76, respectively (Table 3). These $RF\_sd_{i,SSCj}$ models provided high user's accuracies, such as the $RF\_sd_{3,green}$ model ranging from 77.0 % to 86.2 % (Table 4). The $RF\_sd_{i,SSCj}$ models provided also high producer's accuracy, such as the $RF\_sd_{3,green}$ model ranging from 67.6 % to 85.7 % (Table 4).

The three RF_md_{SSCj} models were trained from training datasets and tested on their dedicated test datasets (Figure 2B, steps 2, 3 and 4). The classification performances obtained from RF_md_{green} and RF_md_{dry} on test datasets extracted from S2 images acquired on 4th of August and 2sd of December, respectively, have to be considered carefully as only one class was represented on these test datasets (0% and 5-25%, respectively, Figure 3a and b). These three RF_md_{SSCj} models provided high user's accuracies, such as the RF_md_{green} model ranging from 74.2 % to 92.6 % (Table 4). The RF_md_{green} models provided also high producer's accuracy, such as the RF_md_{green} model ranging from 78 % to 84.5 % (Table 4).

Finally, the performances of the $RF\_sd_{i,green}$ models were slightly superior to those of the RF_md_{green} model (Table 3). As well the performances of the $RF\_sd_{i,struc}$ models were slightly superior to those of the RF_md_{struc} model (Table 3). Additionally, no difference in the performance behaviour of the $RF\_sd_{i,dry}$ models and the RF_md_{dry} model were underlined for the classification of the dry vegetation fraction (Table 3).

[Table 3]

[Table 4]

4.3 Classification maps

Once the RF models were calibrated following both approaches, they were applied to their corresponding Sentinel-2 images. The resulting classifications were aggregated at the field scale using the field boundaries map (Figure 1c), and the majority class was maintained to label the field.

Only the classes used in the training database can be predicted by the $RF\_sd_{i,SSCj}$ models (Figure 2A). For instance, because only three classes of the green vegetation fraction were observed throughout the 34 fields on the 3rd of November 2016 (Figure 3a), the classification map of the green vegetation fraction over Kamech using the $RF\_sd_{3,green}$ model contains only three classes (Figure 4A1). As well, as only two classes of dry vegetation fraction were observed over the 34 fields on the 3rd of November 2016 (Figure

3b), the classification map of the dry vegetation fraction over Kamech using the *RF_sd*$_{3,dry}$ model contains only these two classes (Figure 4B1).

With the use of the “multi-date" approach, which calibrates a unique classification model per SSC from the five Sentinel-2 images and field observations (Figure 2b), all the classes can be predicted. Hence, whereas only three classes of green vegetation fraction were observed over the 34 fields on the 3rd of November 2016 (Figure 3a), the classification map obtained from the *RF_md*$_{green}$ model shows five classes (Figure 4A2). As well, whereas only two classes of dry vegetation fraction were observed throughout the 34 fields on the 3rd of November 2016 (Figure 3b), the classification map obtained from the *RF_md*$_{dry}$ model contained three classes (Figure 4B2). Moreover, whereas all classes could be predicted as they were represented in the calibration dataset, the classification maps for each date that were obtained using the “multi-date" approach do not contain all the classes (Figures 4A2, B2 and C2).

[Figure 4]

4.4 Significance of classifications

The frequency of the majority class within a field may reflect the variability of the classifications at the field scale and thus may give information on the classifications uncertainty at this scale given that the fields are expected to show limited internal variability.

The significance of the majority class was studied based on the chi-squared test, which determined whether there was a significant difference between the *i)* expected frequency of the class due to the random and *ii)* observed frequency of the majority class in each field. The $P\text{-}value_j$ was calculated for each field j, at each date and for each SSC (Figures 5 and 6). The field j associated to $P\text{-}value_j$ lower than 0.05 indicates that the observed higher frequency in this field j is significant and does not result from randomness.

A majority of fields are associated to a low variability of classifications (*P-value* < 0.05) (example in Figures 5a and b, Figure 6). The median of the *P-values* was near 0, and the third quartile was lower than 0.2 regardless of the SSC, date and approach (Figure 6). The p-values obtained for green vegetation were higher than the p-values obtained for other SSCs, except on the 4th of August 2016 and 3rd of October 2016 with the "multi-date" approach (Figure 6b). Finally, regardless of the SSC, approach or date, no spatial pattern appeared in the p-value mapping (Figures 5a and b), as fields associated to high variability of classifications (*P-value* > 0.05) are not the same from one approach to the other (example in Figures 5a and b).

[Figure 5]

[Figure 6]

4.5 Classification comparisons between both approaches

The maps obtained by both approaches for the same SSC_j and date t_i may present some classification differences. These classification differences between the both approaches for the same SSC and date were calculated at the field scale as the percentage of fields classified differently from one approach to the other one for each SSC_j and each date t_i (Table 5). The most important difference in classification between both approaches was obtained for the green vegetation fraction classification on the 22nd of November (Table 5). A less important difference in the classifications between both approaches was obtained for the dry vegetation fraction classification, still on the 22nd of November (Table 5).

Large differences in classification between both approaches and for the three SSCs were observed on the 3rd of October. This could be explained by the interval of 5 days between the field observation date (28th of September 2016) and Sentinel-2 acquisition date (3rd of October 2016). Some agricultural practices may have happened during these 5 days and changed the SSCs, which may have caused flawed associations between the image and the ground information, which may have caused misclassification. Moreover, as this image was slightly cloudy (less than 5% and outside of our study area), these misclassifications may also have been related to flawed atmospheric corrections.

Finally, no correlation was observed between the number of observed classes on field at $t_{\_i}$ and the percent difference of the classification between both approaches (Table 5).

[Table 5]

5. Discussion

Models performances analysis

From the overall accuracy and kappa values, our results showed that both the $RF\_sd_{i,SSCj}$ and RF_md_{SSCj} models provided correct classifications for the three SSCs (Table 3). The good performances obtained for dry vegetation fraction classification are in agreement with the ones obtained with LANDSAT data by Van Deventer et al. (1997) and the ones obtained with ASTER data by Serbin et al. (2009). The good performances obtained for green vegetation fraction classification are in agreement with the ones obtained with Sentiel-2 data by Wang et al. (2018) and the ones obtained with LANDSAT data by Jia et al. (2017).

Nevertheless, as our validation and calibration sets were not completely independent, our overall accuracy and kappa values may have been over-estimated as the pixels in the validation dataset belonged to the same fields as the pixels in the calibration dataset. To be absolutely independent, the validation dataset should be composed of pixels from other fields than those used to calibrate the classification model. However, this perfect independence can be ensured only when the number of observed fields is large enough to be divided into calibration and validation fields, which is rarely the case as field observations are time consuming and costly, especially in case of time series.

Advantages and limitations of both approaches

The “single-date” approach consists in training a RFC model $RF\_sd_{i,SSCj}$ from a calibration database extracted from one Sentinel-2 image t_i, to be applied to a test database extracted from the same image t_i and then to be applied to the entire image t_i (Figure 2A). The “multi-date” approach consists in training a RFC model RF_md_{SSCj} from a calibration database extracted from our five Sentinel-2 images, to be applied to a test database extracted from the five Sentinel-2 images and then to be applied to each image (Figure 2B).

Compared with the “single-date” approach, the use of five Sentinel-2 images in the “multi-date” approach for classifying SSCs allowed to increase the calibration dataset in term of both number of calibration samples and number of observed classes. So the “multi-date” approach allowed predicting a class *A* of an SSC at a date *i*, even if this class *A* was not observed by the operator on this date *i*.

However, the use of these five Sentinel-2 images in RF_md_{SSCj} models provided slightly lower performances compared with the $RF\_sd_{i,SSCj}$ models (Table 3). As the calibration datasets used in RF_md_{SSCj} models were based on five Sentinel-2 images (Figure 2B), the calibration datasets may contain some slight reflectance heterogeneity due to differences in acquisition dates of Sentinel-2 images and so in atmosphere conditions and corrections which may impact the RF_md_{SSCj} models. Slight reflectance differences have been observed between Sentinel-2 spectra acquired on same targets but corrected by different atmospheric methods (Martins et al., 2017; Sola et al., 2018). As well, it can be guess that slight reflectance differences may be observed between Sentinel-2 spectra acquired on same targets, corrected by same atmospheric method, but acquired on different dates.

Finally, whereas most publications have studied dynamic multispectral signals for one final classification (such as Lenney et al. (1996) or Bagan et al. (2005), who used multi-temporal NDVI from LANDSAT and MODIS data, respectively, for land cover classification), whatever our approach applied on the time series of Sentinel-2 images, both the spatial and temporal information of the SSCs were obtained.

Other approach for future

Another approach in future experiments may use some Sentinel-2 images for calibration and an independent Sentinel-2 image for testing, all images acquired over the same study area. This approach would allow the temporal extension of SSCs classifications to other dates. Nevertheless, the potential of this approach might be affected due to soil characteristics differences (e.g., differences of soil humidity) or atmospheric effects differences between calibration and test images. This approach would require i) a calibration from Sentinel-2 dataset images associated with field observations that include all classes of the SSCs and ii) focusing on how to manage such surface directional effects radiometric and seasonal shifts in the classification process. From our knowledge, this approach was never tested, whatever the target (SSCs, land use, etc.).

Classification uncertainties

Calibration of the classification models required the collection of ground truth data and remote sensing images to be as close together as possible, as the SSCs are highly variable both in space and time, depending from punctual anthropic actions. Without this close acquisition, there is uncertainty in the match between ground truth data and spectral information, which may negatively impact the classification results. Indeed, when field-observed data are collected before remote sensing images, some agricultural practices (e.g., ploughing, weeding and seeding) that occur between the data collections may change the reflectance signal, causing the field-observed data to not correspond with the recorded signal. In addition, when remote sensing images are collected before field-observed data, some agricultural practices that occur between the data collections may be recorded in the field observations but not in reflectance signals. A good field expertise is necessary, as it may help to estimate an acceptable interval between field observations and remote sensing data acquisition.

In our case, the uncertainties in the classification obtained on the 4^{th} of August were estimated as null, as no agricultural practices happened between the Sentinel-2 acquisition on the 4^{th} of August and the field observations on the 1^{st} of September. Inversely, the uncertainties in the classification obtained on the 3^{rd} of October may be present, as agricultural practices may have occurred between the Sentinel-2 acquisition on the 3^{rd} of October and the field observations on the 28^{th} of September.

As the SSCs presented strong spatial and temporal variability, each class of SSCs was not represented in the same manner at each observation date (Figure 3). This unbalanced distribution of classes may have produced high uncertainties in the classification results. For example, only two classes of dry vegetation were observed on the 3^{rd} of November, and among both classes, the class 5-25% was overrepresented. Therefore, the classifications obtained by the “single-date” approach with this highly unbalanced distribution of classes must be exploited very carefully.

6. Conclusions

The spatiotemporal monitoring of SSCs is still one of the major challenges for soil infiltration processes modelling, as it is a costly and time-consuming procedure. The successful recent deployment of the Sentinel-2 satellites created a unique opportunity to address the need for the characterization of the earth surface elements both in space and time, including the soil surface characteristics. This study suggested that the proposed approaches applied on a time series of Sentinel-2 images provided spatiotemporal information on three SSCs linked to soil infiltration processes: the green vegetation fraction, dry vegetation fraction and physical soil surface structure. Futures works may focus on combining these SSC maps obtained at each date by the time series remote sensing data, to produce maps of infiltrability classes using pedotransfer functions or typology of the hydrological SSC classes as suggested by Andrieux et al. (2001). Another future study could test a direct mapping of the infiltration classes, following Corbane et al. (2008), who demonstrated that several hydrological SSC classes could be distinguished on the basis of spectral and spatial information collected with aerial RGB photographs over Mediterranean vineyard areas. Finally, although the multispectral remote sensing data acquisition is still increasing and although the data are free and shared thanks to the ESA Copernicus programme, one remaining issue may arise from the limitations in the field data, still necessary for calibrating the classification models. Thus, concurrently with this remote sensing data acquisition and sharing, a special effort could be made on field data acquisition and sharing.

Acknowledgments

This research was granted by the TOSCA- CNES project "AMUSE - Analyse MUlti-temporelle de données SENTINEL 2 pour le monitoring de caractéristiques observables de la surface du sol, en lien avec l'infiltrabilité" (2018–2020). The OMERE Observatory (Molénat et al., 2018, https://www.obs-omere.org/fr/sites/kamech) which is part of the OZCAR network (Gaillardet et al., 2018, http://www.ozcar-ri.org/ozcar/), is acknowledged.

References

Andrieux, P., Hatier, A., Asseline, J., de Noni, G., Voltz, M., 2001. Predicting infiltration rates by classifying soil surface features in a Mediterranean wine-growing area. Oral communication. *International Symposium "The Significance of Soil Surface* Characteristics in Soil Erosion. COST 623 "Soil Erosion and Global Change" workshop, Strasbourg.

Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sensing*, 11, 433. https://doi.org/10.3390/rs11040433

Bagan, H., Wang, Q., Watanabe, M., Yang, Y., Ma, J. 2005. Land cover classification from MODIS EVI times-series data using SOM neural network. *Int J Remote Sens,* 26:4999–5012. https://doi.org/10.1080/01431160500206650

Belgiu, M, Drăgu, L., 2016. Random forest in remote sensing: a review of applications and future directions. ISPRS J. Photogramm. *Remote Sens.* 114. 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Belgiu, M., Csillik, O, 2018. Sentinel-2 cropland mapping using pixel-based and objectbased time-weighted dynamic time warping analysis. *Remote Sens. Environ*. 204. 509–523. https://doi.org/10.1016/j.rse.2017.10.005

Bellón, B., Bégué, A., Lo Seen, D., de Almeida, C.A., Simões, M., A 2017. Remote Sensing Approach for Regional-Scale Mapping of Agricultural Land-Use Systems Based on NDVI Time Series. *Remote Sens.* 9, 600. https://doi.org/10.3390/rs9060600

Ben-Dor, E., Irons, J.R., Epema, G.F., 1999. Soil reflectance. In: Rencz, N., Ed., Remote Sensing for the Earth Sciences, Manual of Remote Sensing. John Wiley & Sons, New York, pp. 111–188.

Børgesen, C.D., Iversen B.V., Jacobsen O.H., Schaap M.G., 2008. Pedotransfer functions estimating soil hydraulic properties using different soil parameters. *Hydrological Processes*, 22 (11), 1630–1639. DOI:10.1002/hyp.6731

Bormann, H., Klaassen, K., 2008. Seasonal and land use dependent variability of soil hydraulic and soil hydrological properties of two Northern German soils. *Geoderma* 145, 295–302. DOI:10.1016/j.geoderma.2008.03.017

Breiman, L., 2001. Random forests. *Mach. Learn.* 45. 5–32. 10.1023/A:1010933404324

Brutsaert, W. Hydrology: an introduction. Cambridge University Press, 2005. DOI:10.1017/CBO9780511808470

Carlson, T.N., Rizile, D.A., 1997. On the Relation between NDVI. Fractional Vegetation Cover and Leaf Area Index. *Remote Sens. Environ.* 62241-252. https://doi.org/10.1016/S0034-4257(97)00104-1

Chahinian, N., Moussa, R., Andrieux, P., Voltz, M., 2005. Comparison of infiltration models to simulate flood events at the field scale. *Journal of Hydrology*, 306 (1-4), 191-214. DOI:10.1016/j.jhydrol.2004.09.009

Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., 1990. High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research*, 95(B8), 12:653 – 12:680. https://doi.org/10.1029/JB095iB08p12653

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46. 10.1177/001316446002000104

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46. 10.1016/0034-4257(91)90048-B

Congalton, R.G., Green, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, *Boca Raton, FL, USA: CRC/Lewis Press*. DOI:10.1201/9780429052729

Corbane, C., Raclot, D., Jacob, F., Albergel, J., Andrieux, P., 2008. Remote sensing of soil surface characteristics from a multiscale classification approach, *Catena*, 75, 308–318. 10.1016/j.catena.2008.07.009

Daughtry, C.S.T., 2001. Discriminating crop residues from soil by shortwave infrared reflectance. *Agronomy Journal* 93, 125–131. 10.2134/agronj2001.931125x

De Roo, A.P.J., Hazelhoff, L., Heuvelink, G.B.M., 1992. Estimating the effects of spatial variability of spatial variability of infiltration on the output of a distributed runoff and soil erosion model using Monte Carlo methods. *Hydrol. Process*. 6, 127–143. 10.1002/hyp.3360060202

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorin, D. 2014 Do we Need Hundreds of Classifiers to Solve Real World classification Problems? Journal of Machine Learning Research, 15, 3133-3181.

Gaillardet, J., Braud, I., Hankard, F., Anquetin, S., Bour, O., Dorfliger, N., de Dreuzy, J.R., Galle, S., Galy, C., Gogo, S., et al., 2018. OZCAR: The French Network of Critical Zone Observatories. *Vadose Zone Journal*, 17, 180067. 10.2136/vzj2018.04.0067

Ghorbani-Dashtaki, S., Homaee M., and Loiskandl W. 2016 Towards using pedotransfer functions for estimating infiltration parameters, *Hydrological Sciences Journal*, 61:8, 1477-1488. 10.1080/02626667.2015.1031763

Gomez, C., Lagacherie, P., Bacha, S., 2012. Using an VNIR/SWIR hyperspectral image to map topsoil properties over bare soil surfaces in the Cap Bon region (Tunisia). In:

Minasny, B., Malone, B.P., McBratney, A.B. (Eds.), Digital Soil Assessments and Beyond. Springer, pp. 387–392. 10.1201/b12728-76

Jarvis, N.J., 2007. A review of non-equilibrium water flow and solute transport in soil macropores: principles, controlling factors and consequences for water quality. *Eur. J. Soil Sci.* 58, 523–546. 10.1111/j.1365-2389.2007.00915.x

Jenhaoui, J., Raclot, D., Lamachère, J.M., 2008. Le parcellaire et l'occupation du sol entre 2004 et 2008 sur le bassin versant de Kamech (Cap Bon, Tunisie). Tunis: *UMR LISAH, IRD* (15 pp. + annexes).

Jia, K., Liang, S., Gu, X., Baret, F., Wei, X., Wang, X., Yao, Y., Yang, L., Li, Y., 2016. Fractional vegetation cover estimation algorithm for Chinese GF-1 wide field view data, *Remote Sensing of Environment* 2016, 177. 184–191. 10.1016/j.rse.2016.02.019

Jia, K., Li, Y., Liang, S., Wei, X., Yao, Y., 2017. Combining Estimation of Green Vegetation Fraction in an Arid Region from Landsat 7 ETM+ Data. *Remote Sens*. 9. 112. 10.3390/rs9111121

Joshi, V. U., & Tambe, D. T. (2010). Estimation of infiltration rate, run-offand sediment yield under simulated rainfall experiments in upper Pravara Basin, India: Effect of slope angle and grass-cover. *Journal of Earth SystemScience*,119(6), 763–773. 10.1007/s12040-010-0055-0

Hagolle, O., Huc, M., Villa Pascual, D., Dedieu, G., 2015. A Multi-Temporal and Multi-Spectral Method to Estimate Aerosol Optical Thickness over Land, for the Atmospheric Correction of FormoSat-2, LandSat, VENμS and Sentinel-2 Images, *Remote Sensing* (3), 2668-2691. 10.3390/rs70302668

Hively, W.D.; Lamb, B.T.; Daughtry, C.S.T.; Shermeyer, J.; McCarty, G.W.; Quemada, M. Mapping Crop Residue and Tillage Intensity Using WorldView-3 Satellite Shortwave Infrared Residue Indices. *Remote Sens.* 2018, 10, 1657. 10.3390/rs10101657

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., 2016. R Core Team, et al. Caret: Classification and *Regression Training*, R Core Team: Vienna, Austria.

Leonard, J., Andrieux, P., 1998. Infiltration characteristics of soils in Mediterranean vineyards in southern France. *Catena*, 32, 209–223. 10.1016/S0341-8162(98)00049-6

Lenney, MP, Woodcock, C.E., Collins, J.B., Hamdi, H., 1996. The status of agricultural lands in Egypt: the use of multitemporal NDVI features derived from Landsat TM. *Remote Sens Environ*, 56:8–20. 10.1016/0034-4257(95)00152-2

Ludwig, J.A., Wilcox, B.P., Breshears, D.D., Tongway, D.J., Imeson, A.C., 2005. Vegetation patches and runoff-erosion as interacting ecohydrological processes in semiarid landscapes. *Ecology* 86, 288–297. 10.1890/03-0569

Martin, P., Joannon, A., Souchère, V., Papy, F., 2004. Management of soil surface characteristics for soil and water conservation, case of a silty loam region : the Pays de Caux. Soil Surface Characteristics: dynamics and impact on soil erosion. *Earth Surface Processes and Landforms*, 29: 1105-1115. 10.1002/esp.1105

Martins, V.S., Barbosa C.C.F., De Carvalho L.A.S., Jorge D.S.F., Lobo F.D.L., Novo E.M.L.d.M. 2017. Assessment of Atmospheric Correction Methods for Sentinel-2 MSI Images Applied to Amazon Floodplain Lakes. *Remote Sensing*. 9(4):322. 10.3390/rs9040322

Matthias, A.D., Fimbres, A., Sano, E.E0, Post, D.F., Accily, L., Batchily, A.K., Ferreira, L.G., 2000. Surface roughness effects on soil albedo. *SSSAJ*, 63(3):1035–1041. https://doi.org/10.2136/sssaj2000.6431035x

Mekki, I., Bailly, J. S., Jacob, F., Chebbi, H., Ajmi, T., Blanca, Y., Zairi, A., Biarnès, A. 2018. Impact of farmland fragmentation on rainfed crop allocation in Mediterranean landscapes: A case study of the Lebna watershed in Cap Bon. Tunisia, *Land Use Policy*, vol. 75, pp.772 – 783. DOI:10.1016/j.landusepol.2018.04.004

Mishra, S.K., Tyagi, J.V., and Singh, P., 2003. Comparison of infiltration models. *Hydrological Processes*, 17 (13), 2629–2652. doi:10.1002/hyp.1257.

Molénat. J., Raclot. D., Zitouna. R., Andrieux. P., Coulouma. G., Feurer. D., Grünberger. O., Lamachère. J., Bailly. J.-S., Belotti. J., Ben Mechlia. N., Ben Younès Louati. M., Biarnès. A., Blanca. Y., Carrière. D., Chaabane. H., Dagès. C., Debabria. A., Dubreuil. A., Fabre. J.-C., Fages. D., Floure. C., Garnier. F., Geniez. C., Gomez. C., Hamdi. R., Huttel. O., Jacob. F., Jenhaoui. Z., Lagacherie. P., Le Bissonnais. Y., Louati. R., Louchart. X., Mekki. I., Moussa. R., Negro. S., Pépin. Y., Prévot. L., Samouelian. A., Seidel. J., Trotoux. G., Troiano. S., Vinatier. F., Zante. P., Zrelli. J., Albergel. J., Voltz. M., 2018. OMERE: A Long-Term Observatory of Soil and Water Resources. in Interaction with Agricultural and Land Management in Mediterranean Hilly Catchments. *Vadose Zone Journal,* 17(1). 10.2136/vzj2018.04.0086

Mubarak, I., Angulo-Jaramillo, R., Mailhol, J.C., Ruelle, P., Khaledian, M., Vauclin, M., 2010. Spatial analysis of soil surface hydraulic properites: is infiltration method dependent? *Agric. Water Manage*. 97, 1517–1526. 10.1016/j.agwat.2010.05.005

Navarro, G., Caballero, I., Silva, G., Parra, P.-C., Vázquez, Á., Caldeira, R. 2017. Evaluation of forest fire on Madeira Island using sentinel-2A MSI imagery. *Int. J. Appl. Earth Obs. Geoinf.* 58. 97–106. 10.1016/j.jag.2017.02.003

Neris, J., Jiménez, C., Fuentes, J., Morillas, G., Tejedor, M., 2012. Vegetation and land-use effects on soil properties and water infiltration of Andisols in Tenerife (Canary Islands, Spain). *Catena* 98, 55–62. 10.1016/j.catena.2012.06.006

Ok, A.O., Akar, O., Gungor, O., 2012. Evaluation of random forest method for agricultural crop classification, *European Journal of Remote Sensing*, 45:1,421-432. 10.5721/EuJRS20124535

Patle, G.T., Sikar, T.T., Rawat, K.S., Singh, S.K., 2019. Estimation of infiltration rate from soil properties using regression model for cultivated land, *Geology, Ecology, and Landscapes*, 3:1, 1-13. https://doi.org/10.1080/24749508.2018.1481633

Pare, N., Andrieux, P., Louchart, X., Biarnes, A., Voltz M., 2011. Predicting the spatio-temporal dynamic of soil surface characteristics after tillage. *Soil & Tillage Research*, 114, 135–145. 10.1016/j.still.2011.04.003

Rapinel, S., Mony, C., Lecoq, L., Clément, B., Thomas, A., Hubert-Moy, L., 2019. Evaluation of Sentinel-2 time series for mapping floodplain grassland plant communities. *Remote Sensing of Environment*. 223. Pages 115-129. 10.1016/j.rse.2019.01.018

Rashidi, M., Ahmadbeyki, A., Hajiaghaei, A., 2014. Prediction of soil infiltration rate based on some physical properties of soil. *American-Eurasian Journal of Agricultural and Environmental Science*,14(12), 1359–1367. DOI: 10.5829/idosi.aejaes.2014.14.12.12461

Robert, J.H., 2019. raster: Geographic Data Analysis and Modeling. R package version 2.9-5. https://CRAN.R-project.org/package=raster

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing,* 67, 93-104. 10.1016/j.isprsjprs.2011.11.002

Serbin G., Hunt Jr., E.R., Daughtry, C.S.T., McCarty, G.W., Doraiswamy, P.C., 2009. An Improved ASTER Index for Remote Sensing of Crop Residue. *Remote Sensing* 2009, 1, 971-991. 10.3390/rs1040971

Shabou. M., Mougenot. B., Chabaane. Z.L., Walter. C., Boulet. G., Aissa. N.B., Zribi. M., 2015. Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi-Arid Lands. *Remote Sensing*. 7. pp. 6059-6078. 10.3390/rs70506059

Sicre, C.M., Inglada, J., Fieuzal, R., Baup, F., Valero, S., Cros, J., Huc, M., Demarez, V., 2016. Early Detection of Summer Crops Using High Spatial Resolution Optical Image Time Series. *Remote Sens.* 8, 591. https://doi.org/10.3390/rs8070591

Sola, I., García-Martín, A., Sandonís-Pozo, L., Álvarez-Mozos, J., Pérez-Cabello, F., González-Audícana, M., Montorio Llovería, R., 2018. Assessment of atmospheric correction methods for Sentinel-2 images in Mediterranean landscapes. Int J Appl Earth Obs Geoinformation 73, 63–76. 10.1016/j.jag.2018.05.020

Story, M., Congalton, R.G., 1986. Accuracy assessment: A user's perspective, *Photogrammetric Engineering and Remote Sensing*, 52, pp. 397-399.

Yimer, F., Messing, I., Ledin, S., Abdelkadir, A., 2008. Effects of different land use types on infiltration capacity in a catchment in the highlands of Ethiopia. *Soil Use Manage.* 24, 344–349. 10.1111/j.1475-2743.2008.00182.x

Yue J., Tian, Q., 2020. Estimating fractional cover of crop, crop residue, and soil in cropland using broadband remote sensing data and machine learning, *Int J Appl Earth Obs Geoinformation* 89, 102089. https://doi.org/10.1016/j.jag.2020.102089

Van Es, H.M., 1993. Evaluation of temporal, spatial, and tillage-induced variability for parameterization of soil infiltration. *Geoderma*, 60, 187–199. 10.1016/0016-7061(93)90026-H

Van Deventer, A.P., Ward, A.D., Gowda, P.H., Lyon, J.G., 1997. Using Thematic Mapper data to identify contrasting soil plains and tillage practices. *Photogrammetric Engineering & Remote Sensing* 63, 87–93.

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sensing of Environment* 2019, 223. 21–33. 10.1016/j.rse.2019.01.006

Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.T., 2018. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf*. 72. 122–130. 10.1016/j.jag.2018.06.007

Wang, B.; Jia, K.; Liang, S.; Xie, X.; Wei, X.; Zhao, X.; Yao, Y.; Zhang, X., 2018. Assessment of Sentinel-2 MSI Spectral Band Reflectances for Estimating Fractional Vegetation Cover. *Remote Sensing*.10, 1927.https://doi.org/10.3390/rs10121927

Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sensing of Environment*. 222. Pages 303-317. 10.1016/j.rse.2018.12.026

WRB, IUSS. 2015. Working Group World reference base for soil resources 2014 (update 2015), World Soil Resources Reports No. 106, FAO, Rome.

Zhang, X., Yan, G., Li, Q., Li, Z.-L., Wan, H., Guo, Z., 2006. Evaluating the fraction of vegetation cover based on NDVI spatial scale correction model. *International Journal of Remote Sensing*. 27:24. 5359-5372. 10.1080/01431160600658107

Figure1

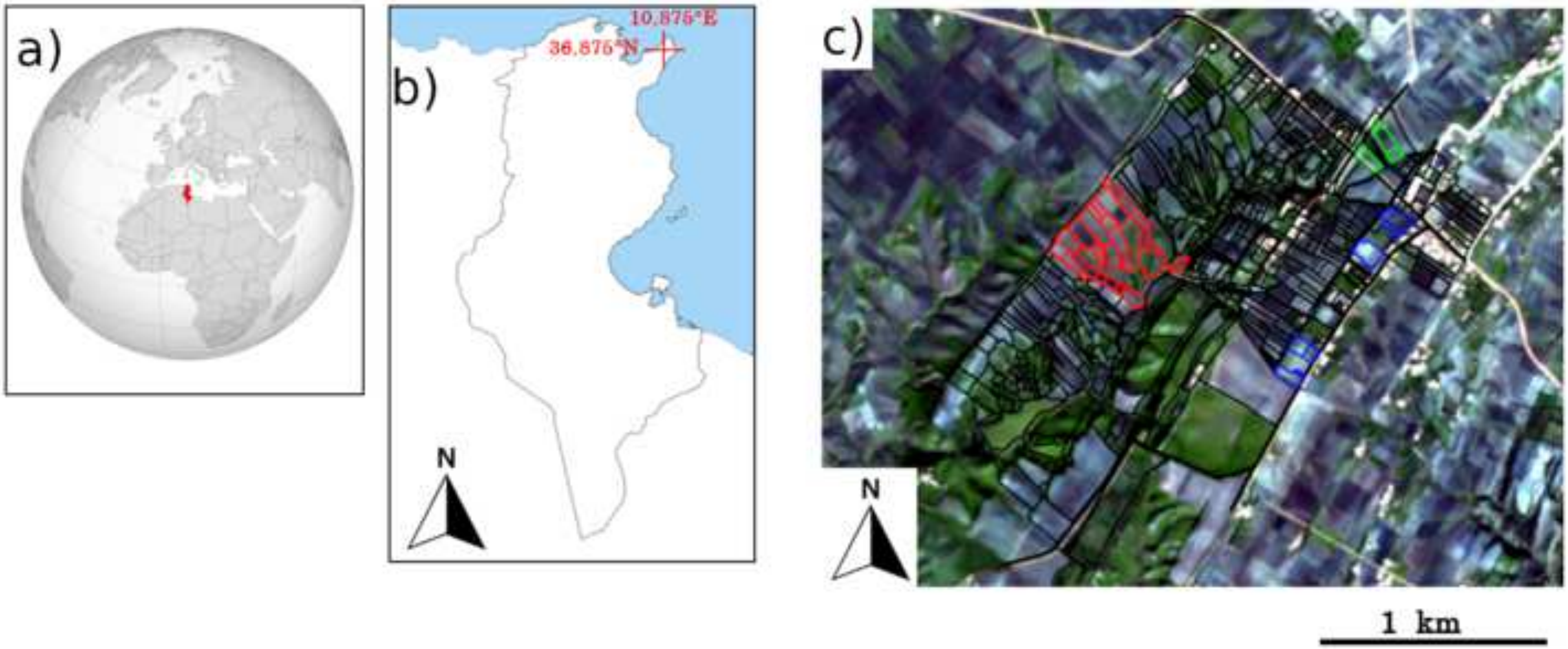
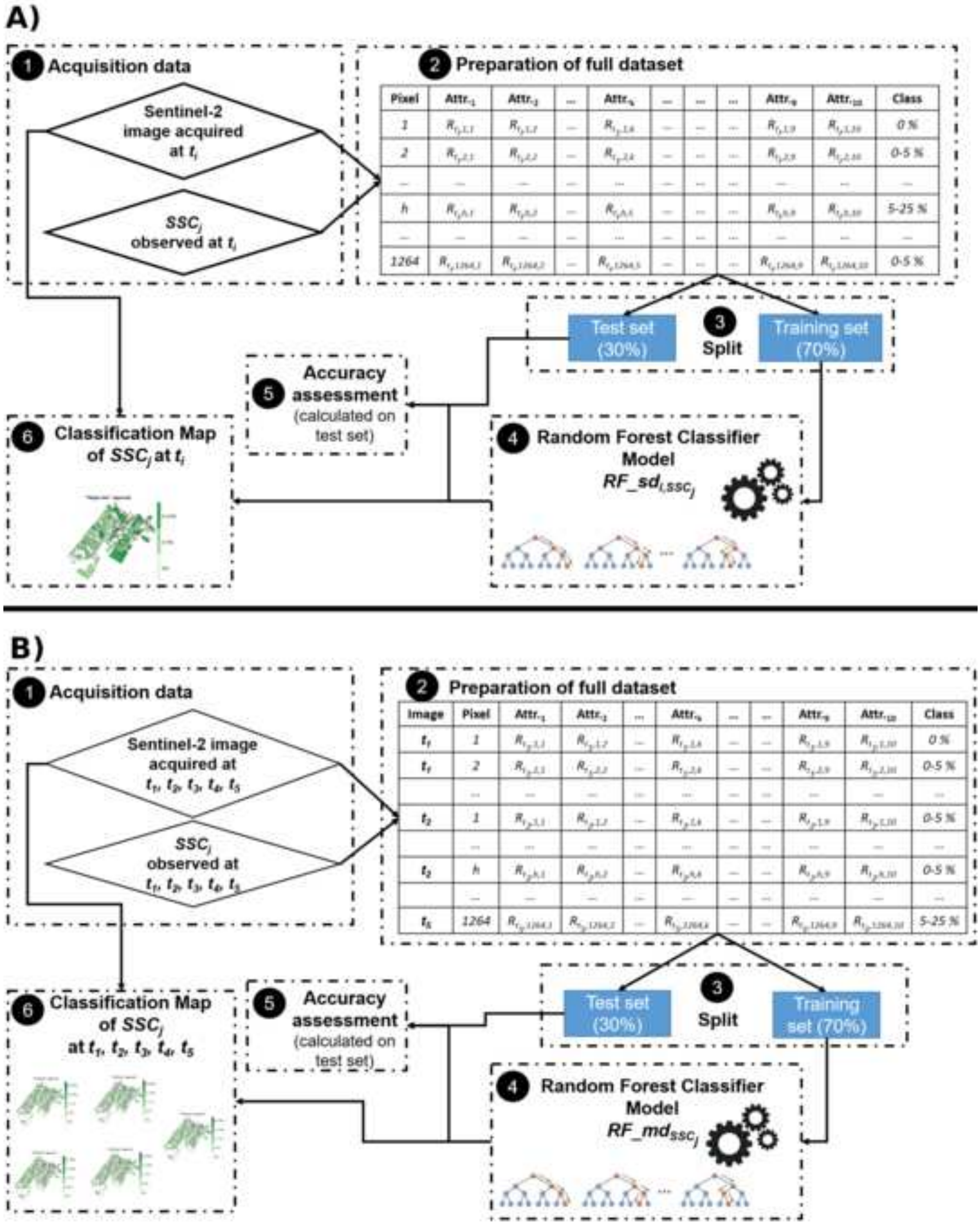


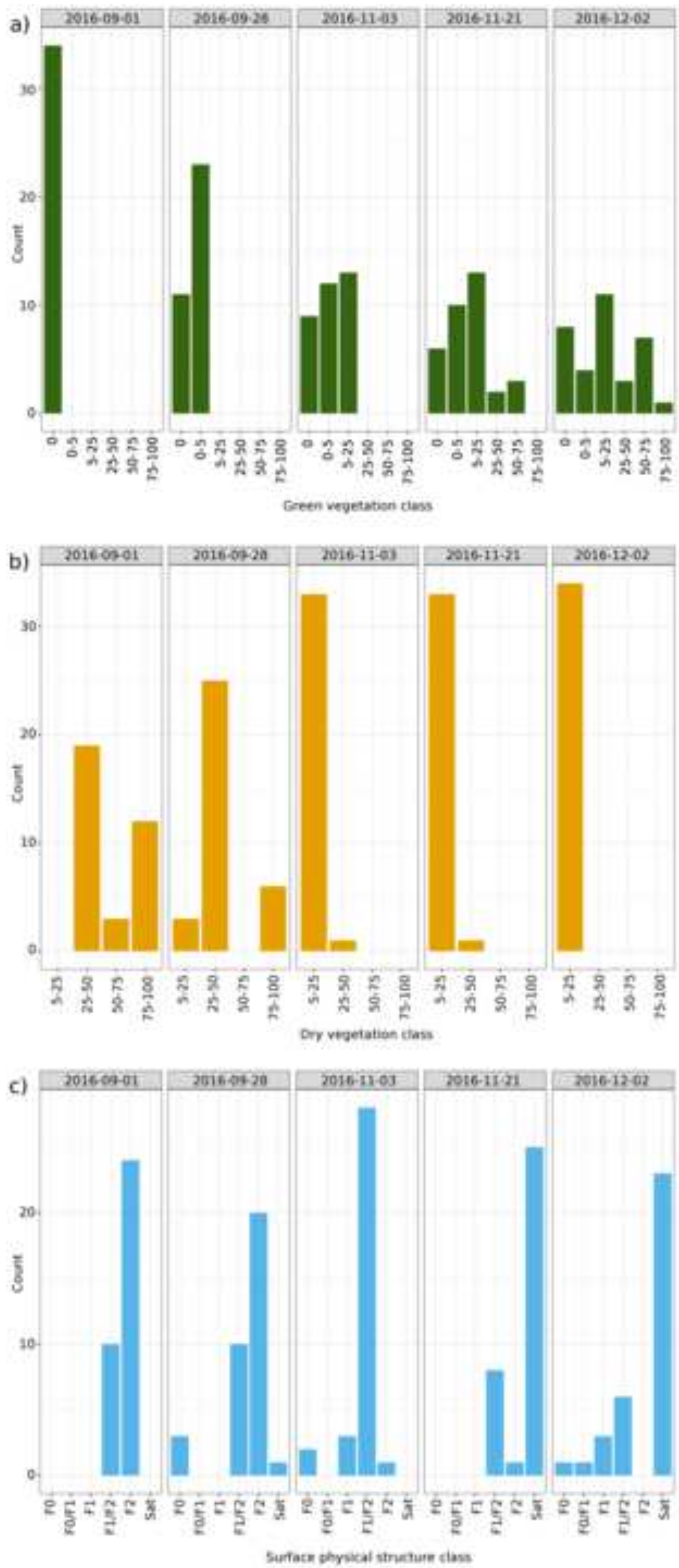
Figure2

-



A)
1 Acquisition data
Sentinel-2 image acquired at t_i
SSC_j observed at t_i
2 Preparation of full dataset
Pixel
Class
0 %
0-5 %
5-25 %
1264
3 Split
Test set (30%)
Training set (70%)
5 Accuracy assessment (calculated on test set)
6 Classification Map of SSC_j at t_i
4 Random Forest Classifier Model
RF_sd_{i,SSC_j}
B)
1 Acquisition data
Sentinel-2 image acquired at t_1, t_2, t_3, t_4, t_5
SSC_j observed at t_1, t_2, t_3, t_4, t_5
2 Preparation of full dataset
Image
Pixel
Class
0 %
0-5 %
5-25 %
1264
3 Split
Test set (30%)
Training set (70%)
5 Accuracy assessment (calculated on test set)
6 Classification Map of SSC_j at t_1, t_2, t_3, t_4, t_5
4 Random Forest Classifier Model
$RF\_md_{SSC_j}$

Figure3



a)
2016-09-01
2016-09-28
2016-11-03
2016-11-21
2016-12-02
Count
0
10
20
30
0
0-5
5-25
25-50
50-75
75-100
Green vegetation class
b)
2016-09-01
2016-09-28
2016-11-03
2016-11-21
2016-12-02
Count
0
10
20
30
5-25
25-50
50-75
75-100
Dry vegetation class
c)
2016-09-01
2016-09-28
2016-11-03
2016-11-21
2016-12-02
Count
0
10
20
F0
F0/F1
F1
F1/F2
F2
Sat
Surface physical structure class

Figure4

Arriˆ¤re-plan

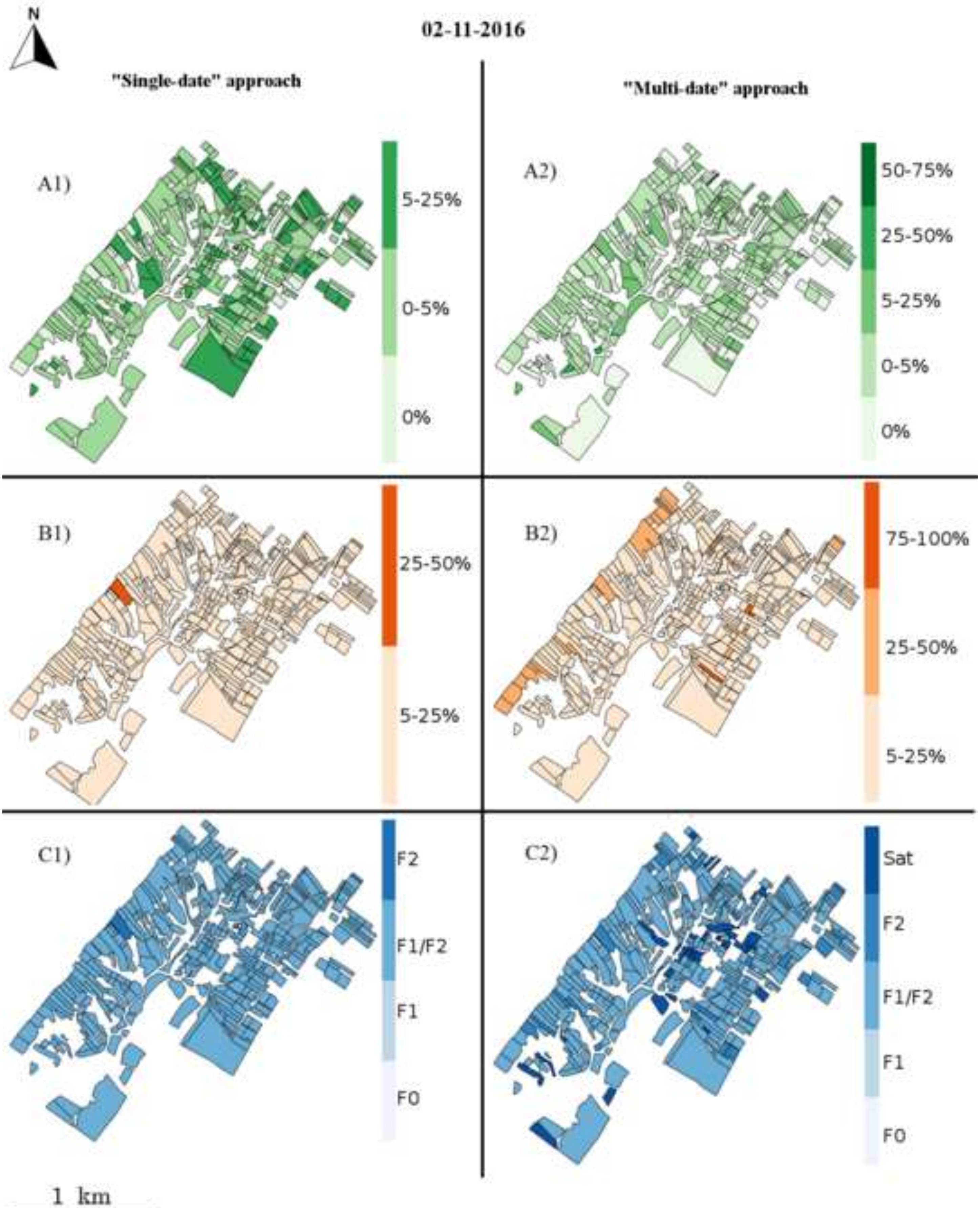


Figure5

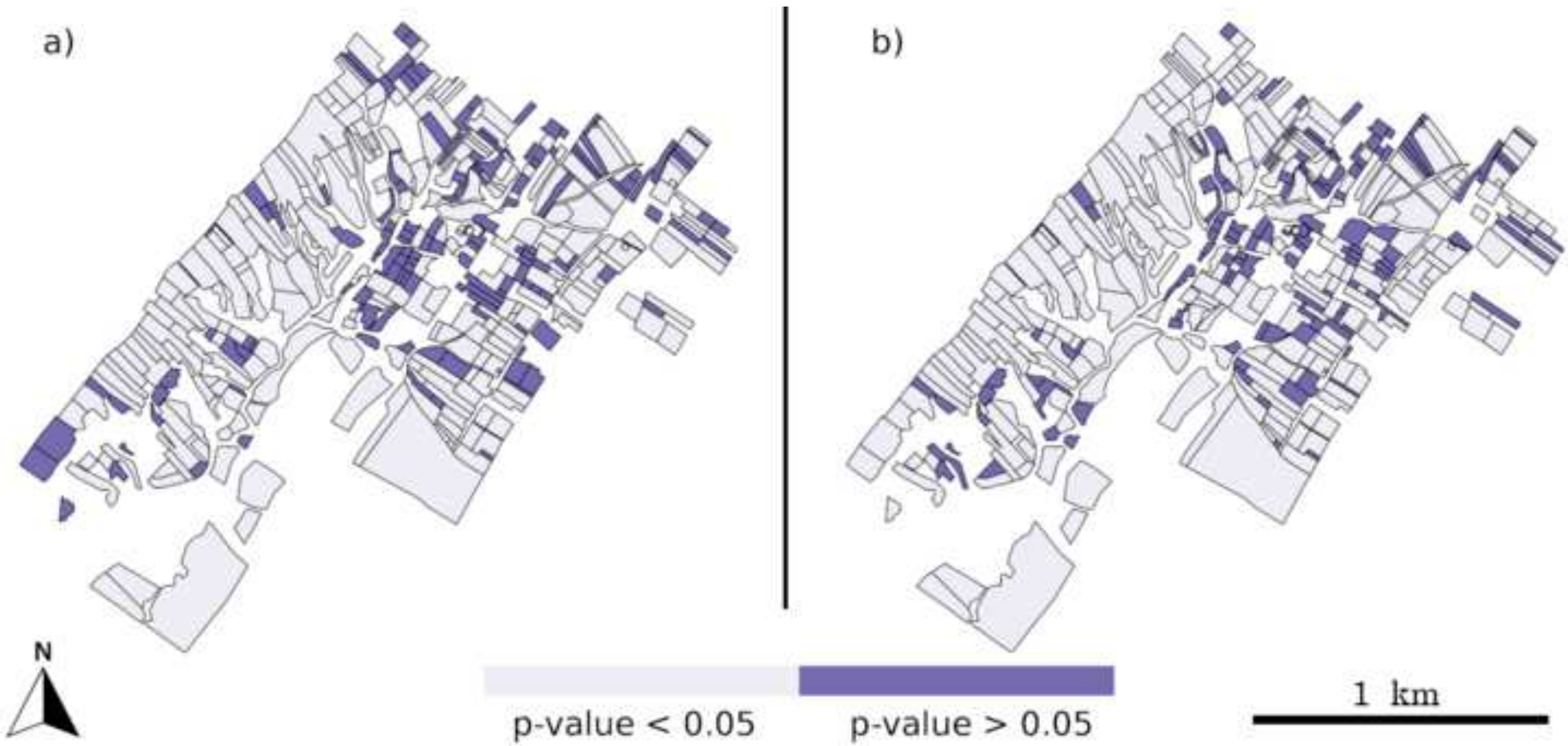


Figure5_R3_300dpi.tiff

Figure6

Figure6_R3_300dpi.png

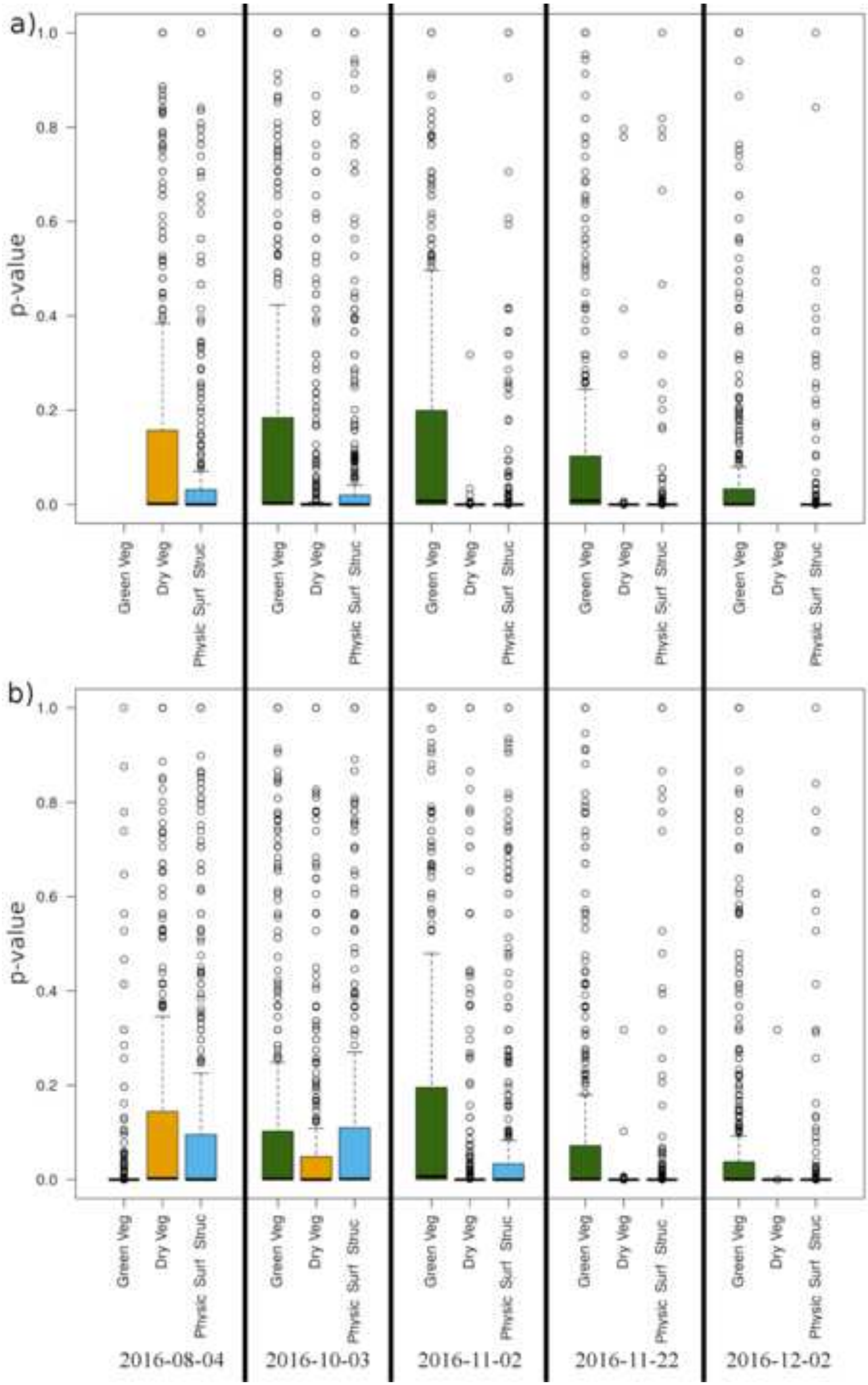


Figure 1: a) Location of Tunisia in Africa, b) Location of the Kamech catchment on Cap Bon peninsula and c) Field boundaries in the Kamech catchment, plotted over a Sentinel-2 image acquired on the 4th of August 2016 (the 34 observed fields are indicated in red, blue and green depending on their location).

Figure 2: Workflow of the SSC classification using the A) “single-date” and B) “multi-date” approach, where $R_{ti,h,k}$ is the reflectance value acquired over the pixel *h* (*h* varying from 1 to 1264) at the spectral band *k* (*k* varying from 1 to 10) for the Sentinel-2 image acquired at t_i (*i* varying from 1 to 5).

Figure 3: Distribution of the **a)** green vegetation, **b)** dry vegetation, and **c)** physical soil surface structure classes observed in the field over 34 agricultural plots on five dates (Y-M-D).

Figure 4: Majority class at the field scale, obtained with the Sentinel-2 image acquired on the 2nd of November 2016 using A1) the $RF\_sd_{3,green}$ model, A2) the RF_md_{green} model, B1) the $RF\_sd_{3,dry}$ model and B2) the RF_md_{dry} model C1) the $RF\_sd_{3,struc}$ model and C2) the RF_md_{struc} model.

Figure 5: *P-values* of the chi-squared test obtained from the Sentinel-2 image acquired on the 2nd of November 2016 using the a) $RF\_sd_{3,green}$ and b) RF_md_{green} model. In clear purple: P-values < 0.005 referring to a significant frequency. In dark purple: P-values > 0.005 referring to frequency close to randomness.

Figure 6: *P-values* of the chi-squared test for green vegetation fraction (green), dry vegetation fraction (orange) and physical soil surface structure (blue), obtained by a) the “single-date” approach and b) the “multi-date” approach.

Table 1: Statistics of area (in ha) on fields.

	All fields over Kamech	34 fields with observed SSCs
Min	0.03	0.07
Max	14	12.27
Mean	0.59	0.51
Standard Deviation	1.04	0.32

Table 2: Acquisition dates of the five Sentinel-2 images and associated field observations dates where *i* is the date number. **There was neither cultural operation nor significant rainfall during this time period.*

Date Number *i*	**Date of Sentinel 2 images acquisition (Y-M-D)**	**Date of field observations (Y-M-D)**	**Number of days between images acquisition and field observation**
1	2016-08-04	2016-09-01	28*
2	2016-10-03	2016-09-28	5
3	2016-11-02	2016-11-03	1
4	2016-11-22	2016-11-21	1
5	2016-12-02	2016-12-02	0

Table 3: Overall accuracy, 95% confidence intervals and kappa calculated from the test datasets for each model (*$RF\_sd_{i,SSC}$ and RF_md_{SSC}*). When no *$RF\_sd_{i,SSCj}$* has been built because of insufficient number of class (i.e., *$RF\_sd_{1,green}$* and *$RF\_sd_{5,dry}$*), the cells were darken. The values calculated from test datasets composed by two unbalanced observed classes are indicated in grey and italics. The performances of *RF_md_{SSC}* models calculated from test datasets composed by only one observed class are indicated in italics.

Date of Sentinel 2 image acquisition (Y-M-D)		**2016-08-04**	**2016-10-03**	**2016-11-02**	**2016-11-22**	**2016-12-02**
Date Number *i*		**1**	**2**	**3**	**4**	**5**
$RF\_sd_{i,green}$	**Overall accuracy**		0.93	0.84	0.85	0.88
	95% CI		[0.9 - 0.94]	[0.82 - 0.87]	[0.82 - 0.88]	[0.84-0.91]
	Kappa		0.82	0.76	0.79	0.84
$RF\_sd_{i,dry}$	**Overall accuracy**	0.89	*0.89*	*0.99*	*0.98*	
	95% CI	[0.87 - 0.92]	*[0.87 - 0.92]*	*[0.99 – 1]*	*[0.97 – 1]*	
	Kappa	0.8	*0.7*	*0.93*	*0.9*	
$RF\_sd_{i,struc}$	**Overall accuracy**	0.93	0.88	*0.95*	0.92	0.89
	95% CI	[0.91 – 0.95]	[0.85 – 0.9]	*[0.92 -0.97]*	[0.89 – 0.95]	[0.85 – 0.91]
	Kappa	0.82	0.78	*0.83*	0.8	0.76
RF_md_{green}	**Overall accuracy**	*1*	0.91	0.79	0.8	0.81
	95% CI	*[0.98 – 1]*	[0.88 – 0.93]	[0.75 – 0.83]	[0.76 – 0.84]	[0.77 – 0.85]
	Kappa	*0*	0.77	0.69	0.73	0.74
RF_md_{dry}	**Overall accuracy**	0.88	*0.9*	*0.94*	*0.99*	*1*
	95% CI	[0.84 – 0.91]	*[0.87 – 0.93]*	*[0.91 – 0.95]*	*[0.97 – 0.99]*	*[0.99 – 1]*
	Kappa	0.78	*0.76*	*0.69*	*0.95*	*0*
RF_md_{struc}	**Overall accuracy**	0.93	0.85	*0.91*	0.91	0.86
	95% CI	[0.89 – 0.94]	[0.8 – 0.87]	*[0.87 – 0.93]*	[0.87 – 0.93]	[0.82 – 0.89]
	Kappa	0.82	0.7	*0.74*	0.79	0.68

Table 4: User's and producer's accuracy obtained on test datasets using the *RF_sd*$_{3,green}$ model (i.e., built from the Sentinel-2 image acquired on the 2nd of November 2016) and the *RF_md*$_{green}$ model.

		Classes on the 2nd of November 2016 (*i=3*)		
		0%	**0 – 5 %**	**5 – 25 %**
***RF_sd*$_{3,green}$**	**user's accuracy (%)**	86.2	77.0	80.7
	producer's accuracy (%)	67.6	85.7	83.4
***RF_md*$_{green}$**	**user's accuracy (%)**	74.2	78.4	92.6
	producer's accuracy (%)	80.2	84.5	78

Table 5: Percentage of classification differences, calculated at the field scale, between maps obtained from $RF\_sd_{i,green}$ and RF_md_{green}, from $RF\_sd_{i,dry}$ and RF_md_{dry} and from $RF\_sd_{i,struc}$ and RF_md_{struc}. When no $RF\_sd_{i,SSCj}$ has been built because of insufficient number of class (i.e., $RF\_sd_{1,green}$ and $RF\_sd_{5,dry}$), the comparison was impossible so the cell was darken.

Date of images acquisition t_{im_i} *(Y-M-D)*		**2016-08-04**	**2016-10-03**	**2016-11-02**	**2016-11-22**	**2016-12-02**
Date Number *i*		**1**	**2**	**3**	**4**	**5**
Green vegetation fraction	**Number of observed classes**	1	2	3	5	6
	% of mapping differences obtained between $RF\_sd_{i,green}$ and RF_md_{green}		19.5	41.9	42.7	11.7
Dry vegetation fraction	**Number of observed classes**	3	3	2	2	1
	% of mapping differences obtained between $RF\_sd_{i,dry}$ and RF_md_{dry}	18.0	32.0	9.4	0.3	
Physical soil surface structure	**Number of observed classes**	2	4	4	3	5
	% of mapping differences obtained between $RF\_sd_{i,struc}$ and RF_md_{struc}	13.3	39.3	18.5	3.6	2.1