

Sentinel data and machine learning algorithms for soil moisture land classification

Salah Benmahmoud¹, Olfa Charfi¹ and Chiraz Masmoudi Charfi²

1. GREEN-TEAM Laboratory (LR17AGR01-INAT), University of Gabes, Gabes National School of Engineers (ENIG), Gabes, Tunisia

2. University of Sfax, Olive Tree Institute- Tunis Station, Ariana, Tunisia

Abstract: Tunisia needs judicious water allocation to mitigate drought. Remote sensing tools, such as Sentinel-1 and Sentinel-2, can be used to estimate soil water status and monitor changes in soil moisture and vegetation cover. The Study areas covers the Kairouan governorate in the center of Tunisia, (35°40'33.29"N, 10°5'30.26"E), which is characterized by a flat relief (plain) covering an area of 6800 Km² and belonging to the 'upper arid bioclimatic stage. The study proposes a formula called "ER" to estimate soil water status from Sentinel-1 data, two color composition images to control changes in soil moisture and vegetation cover, and an unsupervised ISODATA and K-Means classification to monitor the impact of climate change and land use. The data used are nine Sentinel-1 images and twelve Sentinel-2 images downloaded from the Copernicus platform at the date The VV and VH Sentinel-1 GRD Level1 products were selected at different dates based on the daily Precipitation Amount (PA) of the studied region: PA>55mm, 18.5mm < PA < 33.5 mm, PA<10mm and PA =0 mm for Sentinel-2 the dates are 29/01, 23/2, 25/03, 24/04, 16/05, 11/05, 05/06, 13/06, 23/07, 24/08, 18/09 et 11/10 in 2019 for Sentinel-1 the dates of acquisition of images are 24/01, 23/02, 25/03, 12/04, 24/05, 12/06, 27/07, 15/09 et 09/10 in 2019 we are closed were images are cloudness and we chose close dates for the two sentinels and which correspond to the dates of our field visit carried out in June and October 2019. Validation was carried out using two main criteria: (i) the measured mean precipitation deviation over successive months (EP) and (ii) the three reference land cover types (LC) : cereals (LC1), fallow land (LC2) and bare soil (LC3). For this case study, three cases of EP were considered: (1) no precipitation (EP≤0), (2) low precipitation (0<EP≤20 mm), and (3) high precipitation (EP>20 mm). The results showed that, considering the VH polarization, we have ER<0 for LC3 and ER ≈ 0 for the LC1 and LC2 terrestrial cultures when EP≤0; ER>0 for LC3 and ER ≈ 0 for the LC1 and LC2 terrestrial cultures when (0<EP≤20 mm) and ER>0 for all LCs and increases with high rainfall. We validated these ER results using image composition and classified images.

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: Soil moisture; Land crops; Sentinel data; Rainfall; Machine learning algorithms

1. Introduction

The significance of water for the survival of living organisms is an inherent necessity, yet its availability and preservation pose a substantial challenge for humanity. In our current research endeavor, we delve into the study of Soil Moisture (SM), a pivotal parameter in resource management essential for ensuring optimal water productivity and sustainability.

Numerous environmental studies have explored this subject, employing diverse methodologies to monitor SM. Prior investigations have demonstrated the potential of combining radar and optical data to achieve several objectives: (i) recovering SM levels in cereal fields [1], (ii) mapping irrigated and rainfed areas [2], (iii) scrutinizing landscape alterations alongside socio-economic and environmental factors [3], and (iv) establishing

correlations between radar backscatter and SM [4]. More recently, researchers in [5] 1
harnessed neural networks to estimate SM exclusively in unvegetated soils across two 2
research sites in France and Tunisia. In this instance, the neural networks were trained 3
using synthetic data generated through the modified Integral Equation Model (IEM) and 4
validated solely with real unvegetated soil data. 5

This paper presents a holistic approach, involving the synergy of three methods for 6
estimating and monitoring water reserves: (1) the utilization of a formula termed ER to 7
assess the water condition of agricultural land and unvegetated soils, employing Sentinel 8
1 data; (2) the introduction of two novel composite images to monitor moisture levels and 9
changes in crop patterns; and (3) the application of unsupervised classification tech- 10
niques such as ISODATA and K-MEANS to track the impact of climate variations on land 11
usage. 12

2. Dataset 13

2.1. Study Site 14

The research area is situated in the southwestern part of Tunisia, specifically within 15
the Kairouan region, as illustrated as Figure1 and bounded by latitudes 35° to $35^{\circ} 45'$ N 16
and longitudes $9^{\circ} 30'$ to $10^{\circ} 15'$ E. This region is renowned for its arid climate and the 17
excessive depletion of its groundwater resources. 18

2.2. Sentinel's data 19

We obtained the satellite data from the Sentinel-2 and Sentinel-1 GRD Level 1 mis- 20
sions, which were acquired through the <https://scihub.copernicus.eu/dhus/#/home> plat- 21
form. Sentinel-2's MSI sensor offers multi-spectral imagery, while Sentinel-1's synthetic 22
aperture radar antenna provides polarimetric VV and VH data. 23

2.3. GCP Measures 24

Numerous GPS points were recorded at the site, covering various land uses such as 25
orchards, vegetable cultivation, cereal fields, bare soils, and fallow land. Two separate 26
measurement campaigns were conducted, one in June 2019 and another in October 2019, 27
as indicated in Figure 1. We consider three different soil coverage types namely: cereals 28
(LC1), fallow (LC2) and bare soil (LC3). 29

2.4. EP data 30

The validation of the Soil Moisture (SM) estimates generated by our proposed method 31
involved the comparison with the measured average deviations of precipitation over 32
consecutive months, referred to as (EP), which were calculated based on the monthly 33
precipitation data for the year 2019, as illustrated in Figure 2. 34

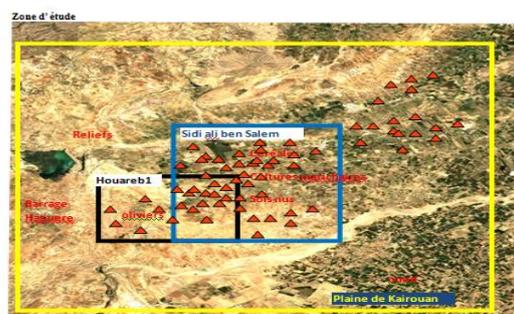


Figure 1. GPS points on Kairouan (Elhouereb) field carried out in June and October 2019. 35

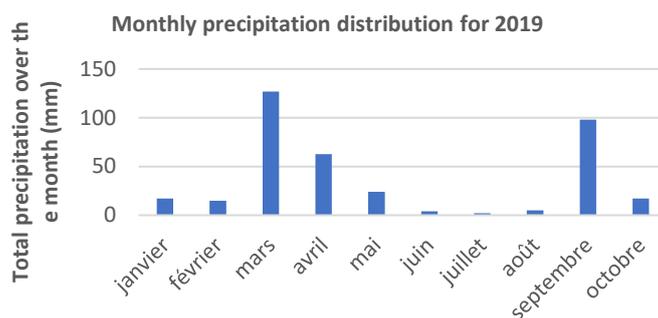


Figure 2. Monthly precipitation distribution for 2019.

3. Methodologies

3.1. The proposed ER formula

We introduce a formula for relative deviation denoted as ER, expressed by equation (1). The ER formula is characterized as a two-dimensional function.

$$ER(k,l) = (\sigma(k,l) - \sigma(k,l+1)) / (\sigma(k,l) + \sigma(k,l+1)) \tag{1}$$

Where ER (k,l) represents the relative deviation of the average backscatter value $\sigma(k,l)$, where k denotes the polarization of the Sentinel 1 image (VV or VH), and l corresponds to the respective month.

3.2. Images composite method

We utilized two distinct colored composition images to monitor changes in humidity and crop patterns:

- I. A colored composition derived from Sentinel2's bands (B8, B4, and B3), primarily employed for land cover identification, particularly vegetation detection.
- II. A colored composition created from Sentinel1's polarizations. We assigned the VH polarization image to the green channel and the VV polarization image to the red channel, primarily for visualizing variations in humidity relief.

The composite image creation process was conducted using the QGIS platform. To enhance the contrast in the colored composition images generated from Sentinel1 data, we applied an enhancement technique that took into account the minimum (equal to 0) and maximum values from the histograms. Additionally, a grayscale threshold of 40 (equal to the maximum value) was applied to both the VH and VV images.

3.3. Machine Learning methods (ML)

The K-MEANS algorithm is employed for clustering data into distinct groups or clusters based on their similarity. This iterative algorithm's objective is to minimize the within-cluster sum of squares, often referred to as inertia.

On the other hand, the ISODATA (Iterative Self-Organizing Data Analysis Technique) methodology is a clustering technique utilized for data analysis and pattern recognition. ISODATA combines elements from the K-means algorithm and the principles of self-organizing systems.

For our analysis, we applied both the K-MEANS and ISODATA algorithms to the set of Sentinel2 spectral bands (B2, B3, B4 and B8). These classification processes were executed using the QGIS platform.

4. Results

4.1. ER model results

We examined the relationship between EP (measured average deviations of precipitation) and ER values across three land use categories: LC1, LC2, and LC3. The findings revealed that when VH polarization was considered:

- I. ER < 0 for LC3, and ER approximately equals 0 for LC1 and LC2 when EP ≤ 0.
- II. ER > 0 for LC3, and ER approximately equals 0 for LC1 and LC2 when (0 < EP ≤ 20 mm).
- III. ER > 0 for all land cover types (LC), with an increase corresponding to higher rainfall, as shown in Figure 3.

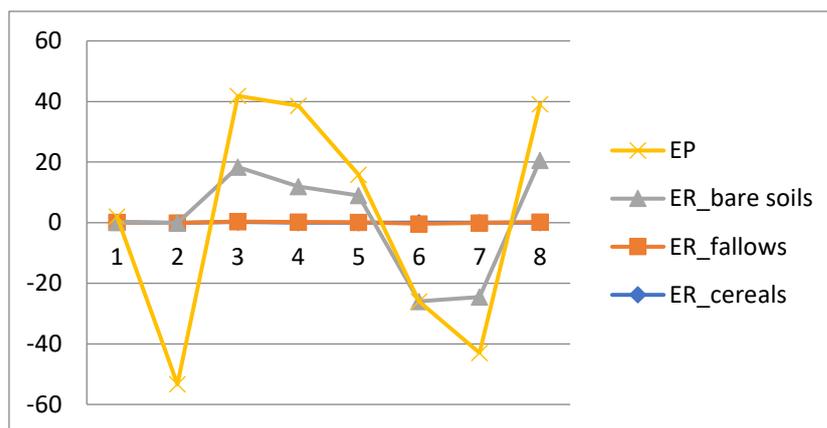


Figure 3. EP and ER_{VH} curve for land covers LC1, LC2 and LC3.

4.2. Image composite results

The Soil Moisture (SM) estimation was conducted through a multi-date analysis of composite images from Sentinel 1 in VH polarization, as presented in Table 1. Meanwhile, the detection of changes in crops was achieved by analyzing multi-date composite images from Sentinel 2, as depicted in Figure 4.

Upon scrutinizing the Sentinel 1 composite images illustrated in Figure 5, it becomes evident that when the polarization backscatter coefficient VV shows an increase from one month to the next, it signifies the occurrence of substantial rainfall. Conversely, if this coefficient remains stable while the VH polarization coefficient decreases, it suggests the absence of rainfall. However, if both VV and VH coefficients decrease, it indicates relatively lower levels of rainfall.

Table 1. Classification of SM based on variation of EP Values and intensities of Sentinel1 VH and VV data backscatter coefficients.

No rain	Little rain	A lot of rain
EP < 0	0 < EP < 20 (mm)	EP > 20 (mm)
VH believes, VV constant	VH decreases, VV constant	VH constant, VV decreases

4.3. Results by ML classifier

The employed classification methods function similarly in recognizing land uses and aiding in the prediction of land use patterns, as depicted in Figure 6 and Figure 7. However, when employing either the ER method or the Sentinel 1 composite image method, it becomes challenging to discern changes in land cover, such as the presence or harvest of cereals or their growth stage.

The growth stages of cereals can be outlined as follows: i) January to March marks the grain's growing season; ii) from March to May, growth tends to stabilize; and iii) from May until the end of June, cereals enter the maturity phase. Through the utilization of

unsupervised classifiers, we were able to identify certain parcels of cereals that had been harvested in June 2019.

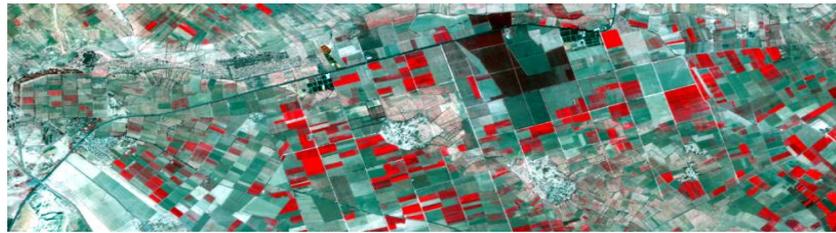
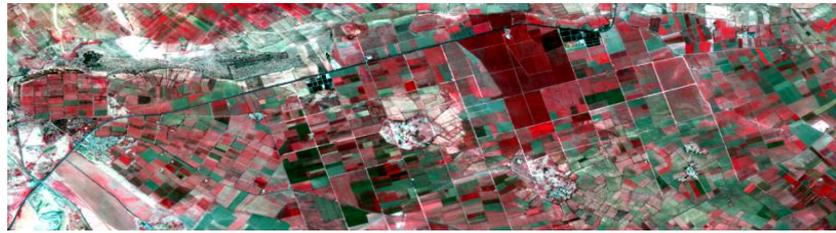


Figure 4. Colored composition of Sentinel2 images for two different dates. Images from up to down: 05.06.2019 and 09.03.2023.

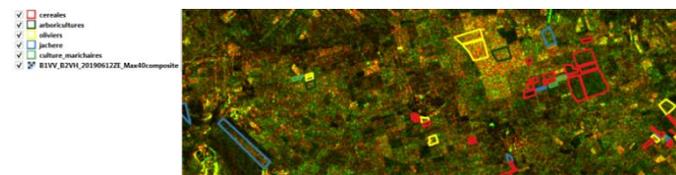
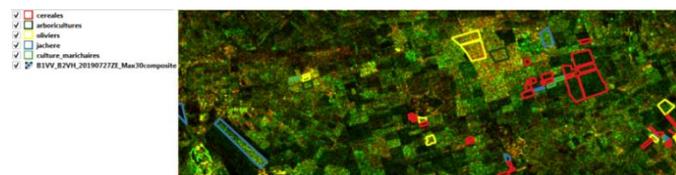
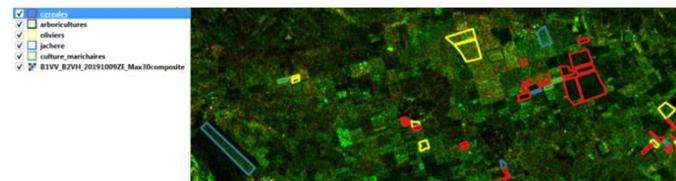
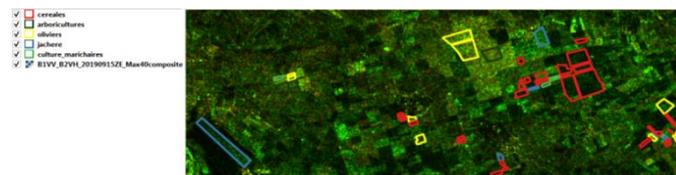


Figure 5. Colored composition of Sentinel1 images for four different dates. Images from up to down: 09.10.2019, 15.09.2019, 27.07.2019 and 12.06.2019.



Figure 6. ISODATA classification of Sentinel2 dataset of 12 June 2019 with 13 clusters. The polygons represent the field truth.

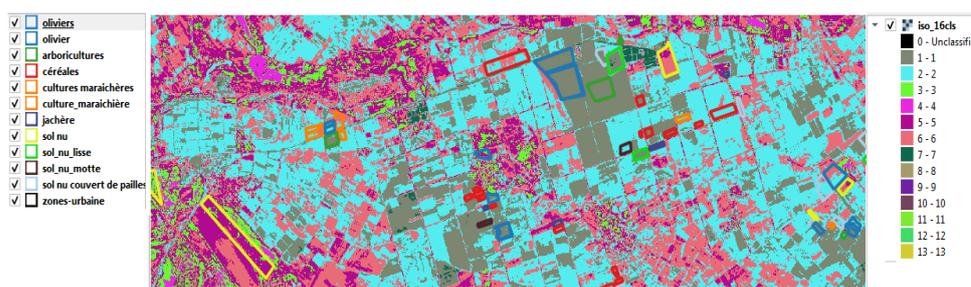


Figure 7. K-MEANS classification of Sentinel2 dataset of 12 June 2019 with 16 clusters. The polygons represent the field truth.

5. Discussion

The results of the ER model provide the best accuracy for soil moisture classification, the Sentinel1 composite images by using VH and VV polarizations allow to classify LC moisture and the Sentinel2 composite images controls the vegetation growth.

Sentinel 1 composite images offer insights into the radar wave backscatter characteristics. These composite images serve as valuable resources for tasks such as soil classification and wetland mapping. The K-MEANS and ISODATA classification methods enable the segmentation of the image into multiple classes, with each class representing a homogenous region of the image based on its spectral properties. This segmentation is particularly useful for soil mapping and crop monitoring purposes.

6. Conclusions

The collaboration among the three methods outlined in this article holds utmost significance. The sequential processing steps should entail: (1) conducting unsupervised classification, (2) computing ER_{VH} , and (3) estimating PE based on Sentinel1 composite images. By adhering to this procedural sequence, we can effectively forecast both soil moisture levels and the growth stages of crops.

References

1. Author 1, A.E.; Author 2, K.Z. Cereal crops soil parameters retrieval using L-Band ALOS-2 and C-Band Sentinel-1 Sensors. *Remote Sens* **2021**,*13*(7), 1393-1418 .
2. Author 1, B.S.; Author 2, Z.M. Sentinel-1 and Sentinel-2 data for characterizing the states of continental surface over a semi-arid region in Tunisia. In: **2020 Mediterranean and Middle-East. Geoscience and Remote Sensing Symposium (M2GARSS)**, pp. 285-288, IEEE Geoscience and Remote Sensing Society 2020.
3. Author 1 M.I.; Author 2, F.I. Oasis extension trajectories in Kebili territory, Southern Tunisia: Drivers of development and actors' discourse. *New Medit, A Mediterranean Journal of economics, agriculture and environment* **2022**, special issue 5, 85-101.
4. Author 1 H.D.; Author 2, V.F. Investigating the correlation between radar backscatter and in situ soil property measurements. *Int J. of App. Ear. Obs. and Geo* **2017**.57, 136-144.
5. Author 1, E.M.; Author 2, B.N. Soil Moisture Retrieval in Bare Agricultural Areas using Sentinel-1 Images. *Preprints.org* **2023**, 2023060661. <https://doi.org/10.20944/preprints202306.0661.v1>.