

# PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://spiedigitallibrary.org/conference-proceedings-of-spie)

## Digital soil mapping using Sentinel-2 imagery supported by ASTER thermal infrared bands

Karyotis, Konstantinos, Tziolas, Nikolaos, Tsakiridis, Nikolaos, Samarinas, Nikiforos, Chatzimisios, Periklis, et al.

Konstantinos Karyotis, Nikolaos Tziolas, Nikolaos Tsakiridis, Nikiforos Samarinas, Periklis Chatzimisios, José A. M. Demattê, George Zalidis, "Digital soil mapping using Sentinel-2 imagery supported by ASTER thermal infrared bands," Proc. SPIE 11524, Eighth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2020), 115240I (26 August 2020); doi: 10.1117/12.2570821

**SPIE.**

Event: Eighth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2020), 2020, Paphos, Cyprus

# Digital soil mapping using Sentinel-2 imagery supported by ASTER thermal infrared bands

Konstantinos Karyotis<sup>\*a,b</sup>, Nikolaos Tziolas<sup>b,c</sup>, Nikolaos Tsakiridis<sup>b,d</sup>, Nikiforos Samarinas<sup>b,e</sup>, Periklis Chatzimisios<sup>a</sup>, José A. M. Demattê<sup>f</sup> and George Zalidis<sup>b,c</sup>

<sup>a</sup>Department of Information and Electronic Engineering, International Hellenic University, Thessaloniki, Greece

<sup>b</sup>Interbalkan Environment Center, i-Bec, Lagkadas, Greece

<sup>c</sup>Laboratory of Remote Sensing, Spectroscopy, and GIS, Faculty of Agriculture, Aristotle University of Thessaloniki, Greece

<sup>d</sup>Automation and Robotics Laboratory, Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece

<sup>e</sup>Laboratory of Rural and Surveying Engineering, Aristotle University of Thessaloniki, Greece

<sup>f</sup>Department of Soil Science, College of Agriculture Luiz de Queiroz, University of São Paulo, Brazil

## ABSTRACT

The importance of monitoring soil properties is constantly increasing among researchers and policy-makers. In this context, it is imperative to identify cost effective and reliable strategies for soil mapping compared to the costlier traditional solutions. A wide range of tools are becoming available that enable better utilization of Earth Observation capabilities to monitor the soil ecosystem. This work is an effort of assessing the potential of Sentinel-2 imagery data for mapping Soil Organic Matter (SOM) contents and investigating the possibilities of its enhancement through ASTER derived information.

The rural area around the lake Zazari, located in the Western Macedonia district of Greece, was chosen as study area. Initially, pixel-wise vegetation indices (NDVI and NBR2) were calculated, utilizing a local version of the CEOS Open Data Cube for masking Sentinel-2 bare soil pixels extending a three-year period (2017–2019). The generated mask was used to extract soil spectral signatures at the image level over selected 100 field samples. The resulting time series was expanded through the conjunction of ASTER Thermal InfraRed bands by matching the exact data acquisition dates of two platforms.

The conclusive part of the work contains the application of regression modelling to effectively assess soil variables. The local Partial Least Square regression algorithm was chosen, due to its characteristics of performing inherently local predictions. Five-fold cross-validation technique was used for reporting the models' accuracy, which was assessed through  $R^2$  coefficient, RPIQ ratio and RMSE. The model estimated SOM values among a synthetic bare soil composite image that was acquired over study area's agricultural fields. Two models were trained and compared; one over Sentinel-2 imagery bands that were used as the predictor variables' set and a second over an expanded predictor variables' set, including ASTER thermal bands. The results signified evidence of accuracy increase of SOM content assessment, through spaceborne imagery analysis.

**Keywords:** Soil spectroscopy, Sentinel-2, ASTER, Local Partial Least Squares, Machine Learning

## 1. INTRODUCTION

Soil is the most important carbon sink after oceans, playing a crucial role in decelerating climate change (Ruddiman, 2003; Paustian, et al., 2016). Its functionality is dependent on a variety of organisms that are part of soil, constituting it an essential component of the planet's biodiversity. Overpopulation strongly affects soil resources sustainability, stimulating awareness for their finiteness (Hartemink & McBratney, 2008) leading to strong concerns about Earths carrying capacity (UNEP, 2013) and soil viability through human exploitation (F.A.O, 2015).

There is an increasing global concern about soil conservation according to (Hartemink & McBratney, 2008) that brings soils back to the global agenda. This rising concern has led to the introduction of a constantly expanding set of soil policies and regulations about soil preservation and restoration around the world regarding the constant soil degradation and decrease of organic carbon in topsoils (Ramesh, et al., 2019). The induced frameworks' main objective is the effective soil management and protection. According to Common Agricultural Policy, measures such as eco-schemes should reward farmers for their performance regarding the environment and climate, by storing and managing soil's carbon and by reducing emissions through improvement nutrient management (European Commission, 2019). The existing global data sets are out of date or restricted to regional areas. As a result, there is an apparent requirement for the development of a new global database, or upgrade of the existing ones, that contains accurate and up to date spatially referenced soil information, addressed to scientists, policy makers and stakeholders in general. In this view, reflectance spectroscopy at the remote sensing scale can play a crucial role in populating the desired databases, while it have been proven to be vigorous approaches for the quantitative determination and modelling of a variety of soil properties, including among others the topsoil mineralogical composition such as soil organic carbon (SOC) concentration, textural composition, iron or carbonate content, etc., and physical attributes (Ben-Dor, et al., 2009; Angelopoulos, et al., 2019).

The typical techniques for the determination of topsoil SOC can be time consuming and cost demanding, thus spectroscopical analysis in the visible and near infrared range (VNIR) and shortwave infrared (SWIR), 400–2500 nm, has proven to be an efficient solution to the aforementioned obstacles. The VNIR spectral signatures that can be extracted in situ or in laboratory enable the instant mining of information regarding key soil properties. A variety of soil properties extending from physical to chemical and biological (such as Soil Organic Carbon, pH, nitrogen among others) can be estimated with low relative-error tolerance through this technique. The VNIR-SWIR analysis trend led to the development of large spectral libraries containing a vast amount of information about soil variables (Tziolas, et al., 2019). The extraction of the desired soil information from spectra-based datasets needs careful treatment, since soil is a multilayered mixture of variables. The successful completion of this procedure enables the application of predictive techniques for an accurate estimation of soil features.

The thermal infrared (TIR) region (8000–11500 nm) may provide useful information regarding the composition and classification of soil particles, since the presence of vibrational frequencies of silicate and carbonate molecules, elicit differences in the spectral emissivity of soils (Sawut, et al., 2014; Rubio, et al., 1997). In this context, the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) that is equipped with a 5 band Thermal InfraRed sensing instrument with spatial resolution of 90m, can support the spaceborne soil spectroscopy.

The key objective of this work is to examine the feasibility of estimating Soil Organic Matter concentration, around the rural area of lake Zazari, using information derived from Sentinel-2 VNIR-SWIR spectral range and ASTER TIR bands, under the hypothesis that TIR spectroscopy can lead to higher accuracy estimations of combined together .

## **2. MATERIALS AND METHODS**

### **1.1 2.1 Soil sampling for local spectral library development**

The rural area around the lake Zazari, that is located in the north-western part of the Central Macedonia district of Greece, was selected as the study area. Lake Zazari is part of a wider lake complex constituted by lakes Vegoritida, Petron and Chimadites. The neighboring area is composed mainly by fertile agricultural fields mostly cultivated with corn and potatoes. Soil samples were collected through two concrete field campaigns that took place at early September of 2018 and 2019, from a group of field experts. At this period, the fields of the study area were mostly cultivated while high levels of vegetation were observed. The data collection points were chosen after thorough study of recent (i.e. immediately prior to the field visit) spaceborne imagery of the area, and identification of bare soil fields or fields with low vegetation. A local Soil Spectral Library (SSL) consisting of 100 samples was assembled by pooling together data from topsoil samplings (0–30 cm) collected from different fields, under a thorough soil sampling procedure.

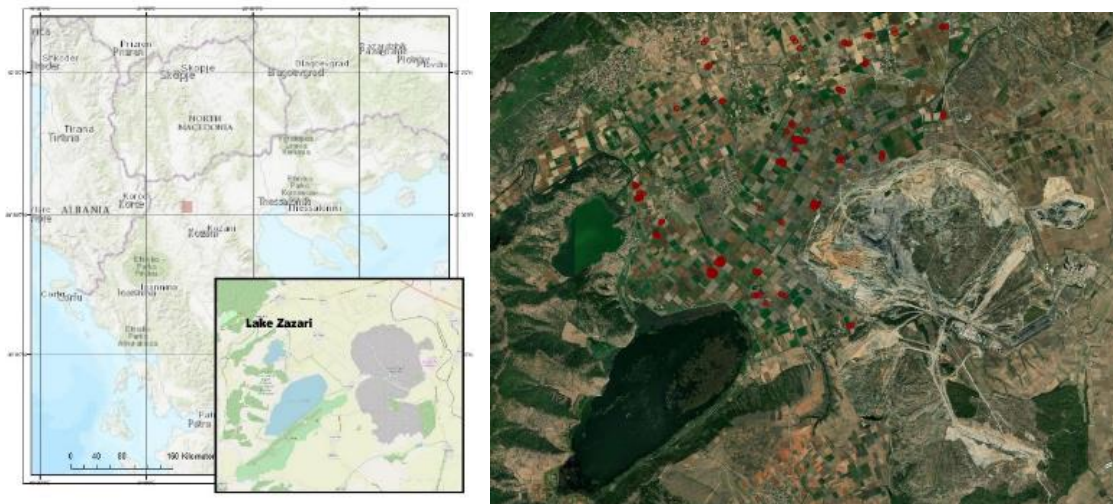


Figure 1. Study area and sampling points' spatial distribution

The samples, after their collection, were air-dried, crushed and sieved resulting to fine earth samples. Stones and non-soil material were removed. The powdered samples were sifted, keeping particles with caliber less than 2 millimeters. These samples were split into two equal parts. The first part was used to assess the SOM concentration, while the second one was used for soil spectra analysis.

The spectroscopic analysis was carried out according to the protocol proposed in (Ben Dor, et al., 2015). The spectral measurement range of the equipment used for spectral signature extraction (PSR+3500 spectrometer from Spectral Evolution) is 350 to 2500nm. For the internal calibration, two different sandy soil samples (Willy Bay and Lucky Bay) with known characteristics and spectral signatures were used. The pulped samples were deposited inside a totally transparent container and placed inside a dark box. The reflectance for each sample was measured three times and the average value of each spectral band was calculated and assigned as the measured value. The calibration procedure was performed after the completion of every five soil samples measurements (Kopačková & Ben-Dor, 2016).

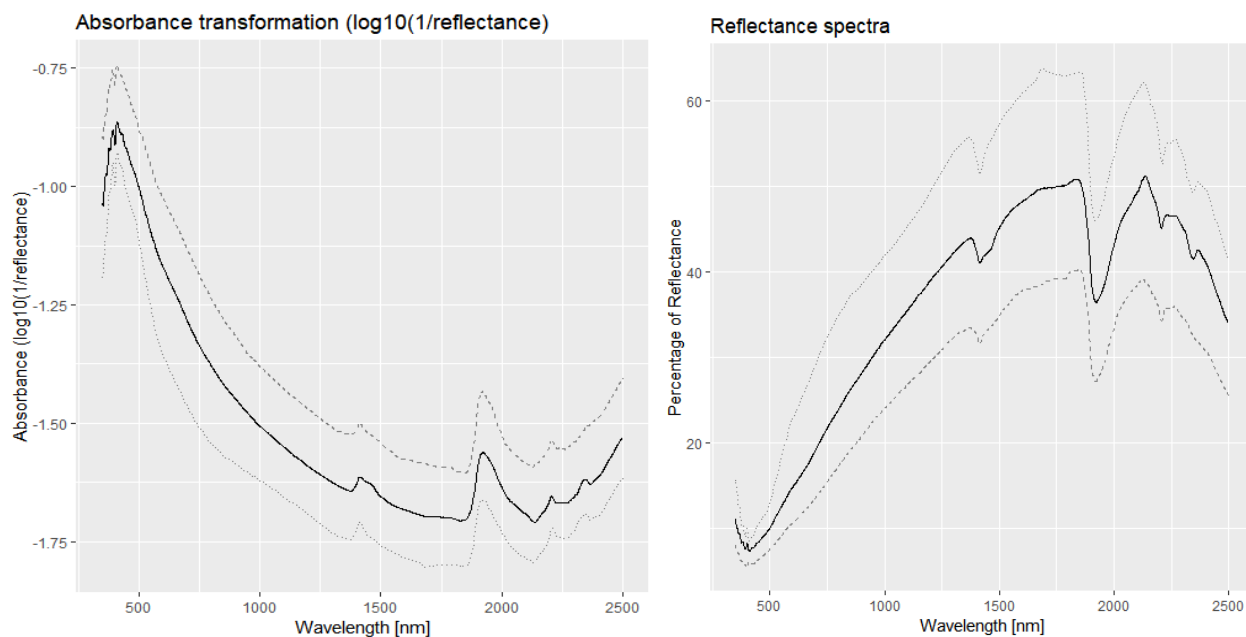


Figure 2 - Laboratorial spectral signatures and spectral reflectance –Median (solid line), first quartile (dotted line) and third quartile (dashed line) of absorbance and percentage of reflectance of sampling set.

### 1.2 2.2 Laboratorial analysis

The soil property analysis was performed by the certified chemical laboratory of the interBalkan Environment Center (I-BEC). The SOM was measured with the application of Walkley-Black method (Heanes, 1984). The key statistical measures are presented in Table 1 - SOM laboratory analyses key statistics.

Table 1 - SOM laboratory analyses key statistics

Property	Min	Q25	Median	Mean	Q75	Max	SD	Skewness	Kurtosis
SOM [%]	0.9	1.73	5.55	4.69	6.98	8.66	2.57	-0.11	-1.63

Most soils in the studied area are characterized by moderate or higher SOM content for croplands. The results show low skewness indicating the absence of outliers, thus there do not exist instances with relatively high SOM values. Since the absolute difference between mean value and median is relatively small, the distribution curve is rather symmetrical, with high variability, suitable for statistical modeling of hyperspectral products.

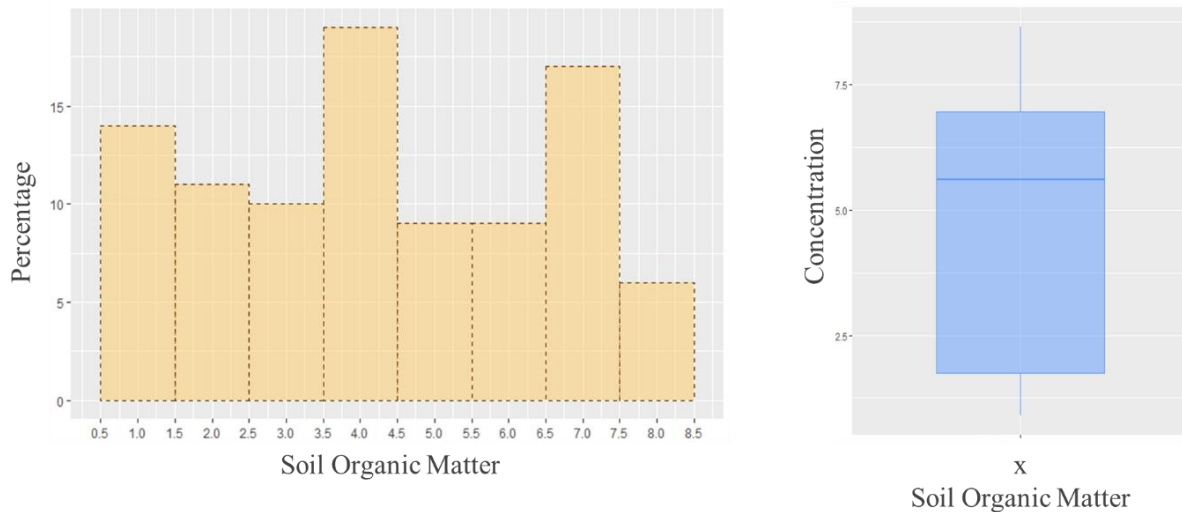


Figure 3 - Histogram and boxplot of SOM concentration

### 1.3 2.3 Spaceborne imagery acquisition and preprocessing

A three year time series (starting from January 2017, until December of 2019) of Sentinel-2 L2 corrected images covering the surrounding area of Lake Zazari were obtained from the Copernicus Sci-Hub platform. In order to determine the agricultural sub-areas of the study area, the normalized difference vegetation index (NDVI) and the normalized burn ratio 2 (NBR2) vegetation indices were calculated where

$$NDVI = \frac{(NIR - Red)}{(NIR + Red)}, NBR2 = \frac{(SWIR1 - SWIR2)}{(SWIR1 + SWIR2)}$$

The areas that their datapoints had  $NDVI \geq 0.25$  and  $NBR2 \geq 0.075$  (Demattê, et al., 2018) were classified as bare agricultural soil, extending to more than 400 distinct agricultural fields. Correlation analysis was conducted between spectral bands of Sentinel-2 imagery and resampled laboratory signatures. As shown on main diagonal of the upper right quarter-block of Figure 4, there is strong correlation between laboratory spectra and Sentinel-2 imagery, thus the development of robust statistical modeling can be conducted.

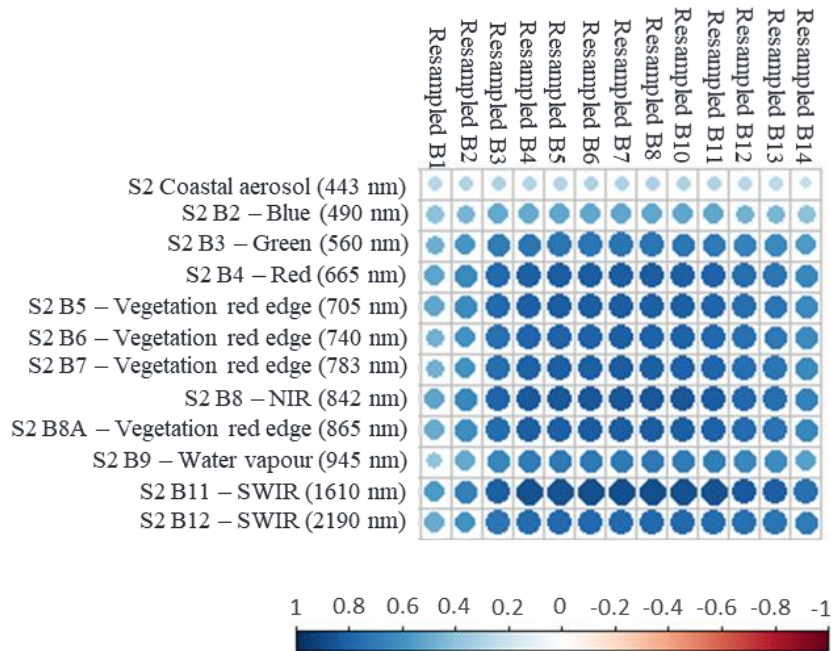


Figure 4 – Correlation between S2 (horizontal) spectral bands and resampled laboratorial spectral bands (vertical)

ASTER Level-2 corrected images from the nearest date to the field campaign were also collected from NASA Land Processes Distributed Active Archive Center (LP DAAC). The mask filtering procedure was conducted through the collation of Sentinel-2 products and ASTER of the same date resulting to the extraction of bare soil of the studied area. The conclusive part of data acquisition contains the collection of Sentinel-2 and ASTER spectral signatures of the sampling dataset for a time frame of 3-years depth. NDVI and NBR-2 filtering was applied, through a local version of the Committee on Earth Observation Satellites Open Data Cube (Killough, 2018), on top of cloud filtering with the same thresholds as previous, while for cloud coverage filter, the value of 10% was selected as maximum tolerated.

#### 1.4 2.4 Model fitting

For the estimation of the depended variable, namely the chemical attribute SOM, a modified version of Partial Least Square Algorithm was applied, through pls package of the Caret package wrapper of R. This Local PLS version follows the memory-based learning approach and develops a single model for each unknown pattern individually using only its neighboring (in terms of spatial and spectral distance) patterns. The experimentation set-up involved a 5-fold cross-validation procedure, where each fold was successively used as test set and the rest as calibration. The hyperparameters of the models where determined through an internal 5-fold cross-validation procedure within each calibration set (Tsakiridis, et al., 2017). The accuracy metrics were calculated as the average results in the independent set across all folds. The fitted model evaluation was assessed through the calculation of the values of the following measures, between the estimated values  $\hat{y}$  and the independent test set:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, RMSE = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{N}}$$

$$RPIQ = \frac{\text{Interquartile distance}}{RMSE}$$

Where,  $y$  is the laboratory measured values,  $N$  the total number of samples consisting the independent set,  $\bar{y}$  is the average value of the independent set, and  $\hat{y}_i$  is the predicted property through the fitted model.



### 3. RESULTS

Local PLS modeling was fitted first to the dataset containing only Sentinel-2 bands while the second part of the modeling was to conjugate the ASTER Thermal InfraRed bands by matching the exact data acquisition dates of the two platforms. The performance of the two models is assessed through the calculation of the average of the above metrics, and is presented in Table 2

Table 2. Local PLS results for the independent test regression modeling

Dataset	R <sup>2</sup>	RMSE	RPIQ
Sentinel-2 bands	0.5274	1.8545	1.2987
Sentinel-2 bands and ASTER TIR Bands	0.6478	1.6492	1.8243

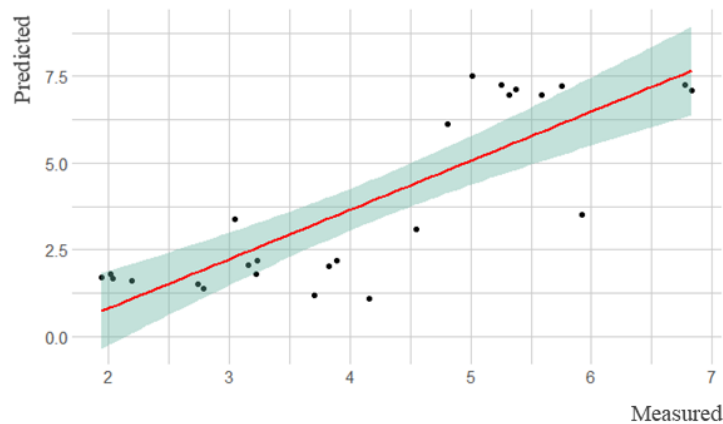


Figure 5 – Linear fitting of validation set of Local PLS model for the conjugated dataset including Sentinel-2 and ASTER bands.

The above values signify that the model fitted to the extended dataset leads to increased accuracy, compared to the Local PLS model that was fitted only to Sentinel-2 bands. There is indication that ASTER products can enhance the use of Sentinel-2 imagery for SOM assessment through the conjugation of TIR bands. The TIR region responded in an encouraging way suggesting further investigation of its potential of assessing a wider variety of chemical soil properties.

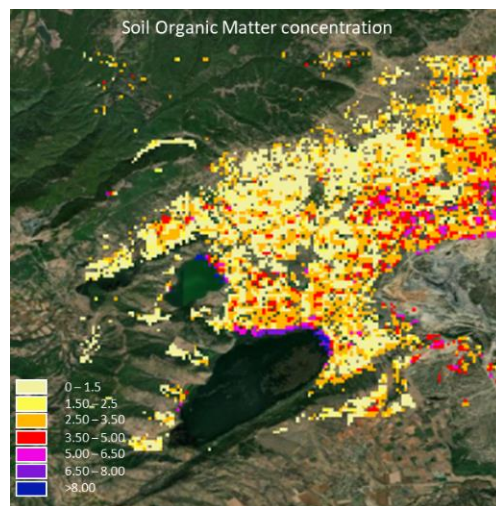


Figure 6. Soil Organic Matter concentration, as predicted from Local PLS modeling. Spectral signatures' source: Sentinel-2 L2A product with sensing date 2018-09-04, and ASTER atmospherically corrected with sensing date 2018-09-09

## 4. CONCLUSION

The proposed approach which is based on the composite image sourced from heterogeneous satellite sources, overcomes the hindrances of utilizing spectroscopy to evaluate soil variables from a single satellite image due to fractional vegetation cover by exploiting the seasonal variation. Therefore, it could greatly improve soil mapping and pave the way for the establishment of an innovative and efficient monitoring system of soil variables. According to (Nanni & Demattê, 2016), it is possible to estimate clay, Fe<sub>2</sub>O<sub>3</sub>, and TiO<sub>2</sub> contents with stepwise models developed from satellite spectral data derived from Landsat-5 Thematic Mapper, since the performed regression modeling returned R<sup>2</sup> value as high as 0.72. The above finding emboldens the application of the proposed methodology on an extended soil attributes set.

For instance, the continuous spatial and temporal monitoring of the SOM can facilitate the reporting of organic carbon content in agricultural soils that is crucial from the environmental perspective, as well as for economic terms to ensure that the beneficiaries of the Common Agricultural Policy respect their cross-compliance obligations for the period post-2020. An integrated in-situ and spaceborne framework as proposed in the DIONE project would be of great benefit for progressing an earth observation based soil monitoring system.

Overall, this study is an interdisciplinary work combining Big Data, machine learning and the Greek DC and aims to support SDGs, and can be realized to a cluster of other areas of comparable terrestrial ecosystems to solve critical problems such as the assessment of topsoil SOM in Greece and even in wider European region. Many adaptations can be explored for further development. Different distance metrics can be assessed for the developed models. Furthermore, with the significant expansion of the training set, Convolutional Neural Networks could be employed (Tsakiridis, et al., 2020) for more precise estimations.

## ACKNOWLEDGMENTS

This research work was funded by the project “DIONE – An integrated EO-based toolbox for modernizing CAP area-based compliance checks and assessing respective environmental impact”. DIONE has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No.870378.

## REFERENCES

- Angelopoulou, T. et al., 2019. Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing*, 11(6), p. 676.
- Ben Dor, E., Ong, C. & Lau, I. C., 2015. Reflectance measurements of soils in the laboratory: standards and protocol. *Geoderma*, 245-246, pp. 112-124.
- Ben-Dor, E. et al., 2009. Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment*, pp. 38-55.
- Centre, J. R., 2012. *JRC Annual Report 2012*, s.l.: Joint Research Centre.
- Commission, E., 2019. *The European Green Deal*, Brussels: European Commission.
- Demattê, J. A. M., Fongaro, C. T., Rizzo, R. & Safanelli, J. L., 2018. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sensing of Environment*, Volume 212, pp. 161-175.
- F.A.O, F. a. A. O. o. t. U. N., 2015. *Status of the World's Soil Resources*, Rome: Food and Agriculture Organization of the United Nations.
- Hartemink, A. E. & McBratney, A., 2008. A soil science renaissance. *Geoderma*, pp. 123-129.
- Heanes, D., 1984. Determination of total organic-C in soils by an improved chromic acid digestion and spectrophotometric procedure. *Commun. Soil Sci. Plant Anal.*, pp. 1191-1213.
- Killough, B., 2018. *Overview of the open data cube initiative*. s.l.
- Kopačková, V. & Ben-Dor, E., 2016. Normalizing reflectance from different spectrometers and protocols with an internal soil standard. *International Journal of Remote Sensing*, 37(6), pp. 1276-1290.



- Paustian, K. et al., 2016. Climate-smart soils. *Nature*, Volume 532, pp. 49-57.
- Ramesh, T. et al., 2019. Chapter One - Soil organic carbon dynamics: Impact of land use changes and management practices: A review. *Advances in agronomy*, Volume 156, pp. 1-107.
- Rubio, E., Caselles, V. & Badenas, C., 1997. Emissivity measurements of several soils and vegetation types in the 8–14  $\mu\text{m}$  wave band: analysis of two field methods. *Remote Sens. Environ.*, Volume 59, pp. 490-521.
- Ruddiman, W. F., 2003. The anthropogenic greenhouse era began thousands of years ago. *Clim. Change*, pp. 261-293.
- Sawut, M. et al., 2014. Estimating soil sand content using thermal infrared spectra in arid lands. *International Journal of Applied Earth Observation and Geoinformation*, Volume 33, pp. 203-210.
- Tsakiridis, N., Keramaris, K., Theocharis, J. & Zalidis, G., 2020. Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma*, Volume 367.
- Tsakiridis, N. et al., 2017. *Predicting soil properties for sustainable agriculture using vis-NIR spectroscopy: a case study in northern Greece*. Paphos
- Tziolas, N. et al., 2019. A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved VIS-NIR-SWIR soil properties estimation., *Geoderma*, Volume 340, pp. 11-24.
- UNEP, 2013. *UNEP 2012 annual report*, s.l.: United Nations Environment Programme.