D4.2: Inventory of historical and open EO data and techniques to be used

October/2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870378.

Author(s)/Organisation(s)	Nikolaos Tsakiridis, Nikolaos Tziolas (i-BEC)
Contributor(s)	Nikiforos Samarinas (i-BEC)
Work Package	WP4
Delivery Date (DoA)	31/10/2020
Actual Delivery Date	31/10/2020
Abstract:	A report describing (i) the overall DIONE soil methodological framework with more emphasis to a novel methodological approach which combines satellite imagery data, soil spectral libraries and local spectral measurements based on a portable and handled Micro electronical Systems (MEMS) (ii) the current status in soil spectroscopy, (iii) the modelling procedures that can transform spectral signatures into soil properties and (iv) the Soil Organic Carbon (SOC) map resulted form the proposed methodology.

Document Revision History			
Date	**Version**	**Author/Contributor/ Reviewer**	**Summary of main changes**
21/09/2020	V0.1	I-BEC	Initial ToC
12/10/2020	V0.6	I-BEC	Submitted for internal review
29/10/2020	V0.9	I-BEC	Small modifications according to reviewers' feedback
31/10/2020	V1.0	ICCS	Approved, final version submitted

Dissemination Level		
PU	Public	**X**
CO	Confidential, only for members of the consortium (including the EC)	

DIONE Consortium			
Participant Number	**Participant organisation name**	**Short name**	**Country**
1	INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	ICCS	EL
2	DIABALKANIKO KENTRO PERIBALLONTOS	i-BEC	EL
3	SINERGISE LABORATORIJ ZA GEOGRAFSKE INFORMACIJSKE SISTEME DOO	SINERGISE	SI
4	CORE INNOVATION AND TECHNOLOGY OE	CI	EL
5	NATIONAL PAYING AGENCY	NMA	LT
6	INOSENS DOO NOVI SAD	INO	RS
7	GILAB doo Beograd-Palilula	GILAB	RS
8	Cyprus Agricultural Payments Organisation (CAPO)	CAPO	EL

Funding Scheme: Innovation Action (IA) ● Theme: DT-SPACE-01-EO-2018-2020
Start date of project: 01 January, 2020 ● Duration: 30 months

Table of Contents

List of figures

List of Abbreviations and Acronyms	
CAP	Common Agricultural Policy
CEC	Cation exchange capacity
CLMS	Copernicus Land Service
CNN	Convolutional Neural Network
DIAS	Data and Information Access Services
EO	Earth Observation
ESDAC	European Soil Data Centre
EU	European Union
IS	Imaging Spectroscopy
GEOSS	Global Earth Observation System of Systems
GLOSOLAN	Global Soil Laboratory Network
GSSL	GEO-CRADLE Soil Spectra Library
LDCM	Landsat Data Continuity Mission
LUCAS	Land Use and Cover Area frame Survey
MEMS	Microelectromechanical Systems
MIR	Mid-wavelength infrared
MSI	Multispectral Instrument
NASA	National Aeronautics and Space Administration
NDVI	Normalized difference vegetation index
NIR	Near Infrared
OLI	Operational Land Imager
OTA	Over The Air
RGB	Red Green Blue
SAR	Synthetic Aperture Radar
SDG	Sustainable Development Goal
SOC	Soil Organic Carbon
SOP	Standard Operating Procedures
SSL	Soil Spectra Library
SWIR	Short-Wave infrared
TIRS	Thermal Infrared Sensor
TLS	Transport Layer Security
USGS	United States Geological Survey
VSNIR	Visible and Near-Infrared

1 Introduction

1.1 Context and Background

Environmental degradation and specifically land degradation is caused by various factors, including climatic variations and human activities such as agriculture. At the same time, as the world population continues to increase, pressure on soil also escalates and the natural capital of soil faces continuing decline (Koch et al., 2013, Montanarella et al., 2016), for example Soil Organic Carbon (SOC) loss is one of the main causes of soil degradation on the planet (Lal, 2004). International policy makers have recognized this and a range of initiatives to address it have emerged over recent years. Several EU policies and worldwide efforts have underscored the importance of soil monitoring to maintain healthy soils. Land and soil degradation is a global concern and land degradation neutrality is one of the targets of the UN Sustainable Development Goals.

Considering the above, the spatiotemporal monitoring of soil properties and the provision of a cost effective method, for the estimation of the environmental indicators is of vital importance. In this regard, and since conventional methods for soil monitoring are time consuming and expensive, we propose a novel framework aiming to enable the combination and synergistic use of **Earth Observation (EO) systems** with **open-access soil spectral libraries (SSLs)** and **local spectral measurements** to be used within the scope of the DIONE project. Using memory-based learning algorithms the above combination can provide an alternative to contemporary approaches that require laborious soil sampling and use of analytical techniques, that is inexpensive and accurate and can provide targeted large scale mapping of soil properties such as SOC, pH, electrical conductivity, etc.

In the past decade, researchers in the EU and throughout the world have been focusing on finding an easy, time-efficient, and inexpensive way to map, monitor and estimate soil use and soil property changes by developing beyond the state of the art, innovative methodologies, services and tools. EO data arising from satellite sensors and in-situ measurements (archived or updated) are a unique source of knowledge, whereas the combination of them with state-of-the-art (SotA) methods can give information regarding the soil properties, even in remote locations. The role of the Global Earth Observation System of Systems (GEOSS) is instrumental for this: the GEOSS Platform promotes the use of common technical standards so that data from thousands of different instruments can be exploited free of charge and combined into coherent data sets.

Traditionally, geostatistical techniques have been used to map soil properties by spatially interpolating the properties obtained through analytical techniques in discrete points. This task however is not cost-effective, while the quality of the resulted layer depends on either the amount of data and the harmonised geographic distribution. By exploiting EO data, the various satellite imagery band combinations in the visible and near-infrared (VNIR) and short-wave infrared (SWIR) may be used to identify the concentration of various soil properties. The image optical spectral data, and particularly the hyperspectral data, have been proven good indicators for the prediction and mapping of the spatial variability of soil properties such SOC (Ben-Dor et al., 2009; Chabrillat et al., 2019). The variation of the different spectral profiles and the contribution of artificial intelligence models (and in particular through the use of machine and deep learning models), can give an insight on differences occurring in soils due to the various factors, such as the slope inclination, the land use categories and others. Apart from the satellite imagery, also field data is of crucial importance in order to provide field information, mainly for calibration and validation purposes. In this regard, contemporary soil spectral libraries (SSLs) in combination with local in-situ measurements, collected through portable/handled spectrometers built with microelectromechanical systems (MEMS), can be transformed in soil properties and used as reference data, enabling the prediction of the soil property maps, through the exploitation of the satellite data.

1.2 Purpose of the deliverable

The overarching object of the current deliverable is to present the methodological framework of the soil monitoring in which DIONE operates. Additional information is presented regarding the requirements that is mandatory for the integration of existing open SSLs and the spectral profiles from the multispectral satellite data, in order to deliver quantitative and spatially explicit soil indicators, and thus enable the continuous monitoring of the soil ecosystem.

In DIONE deliverable *D4.1. Technical specifications of the in-situ soil scanning system (SSS), data processing system and farmer's geo-tagged photos framework* the in-situ soil scanning system was described, which encompasses: (i) the MEMS spectrometers which capture in-situ spectra, and (ii) the database to which these data are delivered. The present deliverable thus continues with the processing of these data, in conjunction with other historical and open EO data from other sources, to arrive at the end-product (i.e. the soil property maps).

The deliverable aims to cover the synergistic use of Earth Observation techniques (spaceborne and in-situ data) with a clear focus on the estimation of soil properties. These innovative tools and techniques can support the monitoring of the soil ecosystem and thus address goals and targets of major international frameworks, such as the Sustainable Development Goals (SDGs), the Common Agricultural Policy (CAP), and the new European Green Deal.

The rest of the document is organised as follows. Section 2 presents a literature review of the most widely used methods on soil spectroscopy analysis and soil property mapping. Section 3 illustrates the novel approach, in which DIONE soil monitoring framework will operate. Technologies and modelling procedures are explored to achieve a reliable and cost effective scalable approach to soil property map production. Finally, in section 4, we provide a short summary of the current work.

2 Review on past methodologies for soil property mapping

The availability of soil information in the Member States of the European Union varies greatly in many regards; evidently, as the soil ecosystem requires constant monitoring to prevent its degradation and promote its sustainable management, this lack of in-situ data in some regions may hinder the accurate estimation of its status. This section focuses on past techniques that have been developed in the literature that pertain to the mapping of soil properties in order to assess its spatio-temporal variance.

2.1 Factorial models, environmental similarity, and environmental correlation

Traditionally, the process to develop a digital soil map (depicting the spatial distribution of a given physicochemical property or of a class) for a region entails the following steps (Keskin and Grunwald, 2018; Minasny and McBratney, 2016):

1. Sampling locations for the region to be mapped are selected via stratification.
2. A field campaign takes place in order to extract the physical soil samples from the pre-selected locations.
3. The field samples are sent to a chemical laboratory which estimates the properties of the samples via analytical techniques.
4. A geostatistical model is then applied which makes use of ancillary data (e.g. topographic indices, climatic, geomorphological, and other pedological data, or a single band or an index from satellite data) to correlate them with the analytically measured properties, and thus be able to estimate the properties in the whole region.

However, this approach is not cost efficient and cannot provide a scalable and interoperable model, which could be exploited by anyone. In other words, it has low transferability. Moreover, the cost associated with the chemical laboratory is considerable. It should also be noted that in order to generate a new map to examine the temporal differences, a new field campaign needs to be implemented as this approach only presents *"a static snapshot"* of the current time of sampling. Thus, a more systematic approach is necessitated to generate soil maps on larger temporal and spatial scales.

2.2 Techniques based on VNIR–SWIR spectroscopy

2.2.1 Laboratory-based point spectroscopy

As a first step, it should be noted that the cost associated with the chemical laboratory may be substantially decreased through the use of laboratory-based spectroscopy, (Nocita et al., 2015). The traditional chemical analysis of soil samples to estimate their physicochemical properties is laborious and costly in terms of resources. Diffuse reflectance spectroscopy, especially visible and near infrared spectroscopy (VNIR–SWIR, 400–2500nm) is a very popular technique, widely recognized as a simple, accurate, time- and cost- efficient method of soil analysis, (Stenberg et al., 2010). In essence, the soil samples instead of being sent for chemical analyses first undergo a pre-treatment, entailing (i) the drying of the sample so that there is no effect of soil moisture on the spectrum, (ii) the breaking up of large soil particles, and (iii) the sieving so that only the fine earth fraction (< 2 mm) is retained. Then, the diffuse reflectance spectrum of the sample is collected (either in VNIR–SWIR or in the mid-wavelength infrared or MIR), which can be used to estimate the composition of the sample. A single spectrum may contain comprehensive information on various soil components, and it can thus be used to predict these simultaneously. To infer the properties, a suitable chemometric model is developed, (Figure 1).

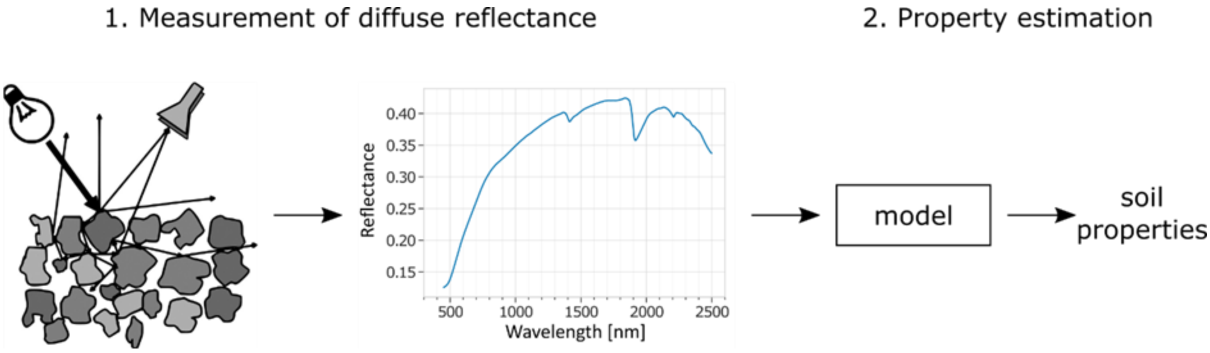

Figure 1: High level overview of diffuse reflectance soil spectroscopy

This research area is very-well studied due to the advantages that the controlled laboratory conditions offer, enabling models to attain high precision, (Tsakiridis et al., 2020). The disadvantage is that physical samples need still to be collected and transferred to a laboratory, and that the samples need to undergo a treatment (drying and sieving) which may take some time spanning from a few days to couple of weeks) depending on the number of samples. An example of a spectrum acquisition using a contact probe is presented in Figure 2.

Figure 2: Measurement of diffuse reflectance spectroscopy in the laboratory using a contact probe bearing an optical fiber

2.2.2 In-situ spectroscopy

Moving from laboratory-based point spectroscopy to in-situ point spectroscopy where soil spectral measurements are acquired in the field, we can reduce the cost associated with the aforementioned processing steps, while on the contrary any offered advantage will be missed. More concretely, the steps that include (i) performing soil sampling, (ii) sending the samples to the laboratory, and (iii) processing the samples (drying and sieving), are longer necessary.

Of course, this means that the advantages offered by the laboratory are now no longer offered by this methodology. Two factors heavily effect the spectral reflectance when it is acquired in-situ: (i) the soil moisture content (Figure 3) (Lesaignoux et al., 2013; Lobell and Asner, 2002) and (ii) the soil roughness. Soil moisture reveal a continuous effect in the spectrum terms of darkening the colour (reducing the albedo) and increasing the spectral absorption mostly on 1400 and 1900 nm, due to the presence of the O-H bond of the water molecule. If models do not account for this effect, then the prediction results may be prone to large errors. Several methodologies have been developed to overcome these effects (Minasny and McBratney, 2016; Nocita et al., 2013; Rodionov et al., 2014). The most popular ones rely on External Parameter Orthogonalization (EPO) and Direct Standardization (DS), and they both have shown to be effective strategies to mitigate the effects of soil water content (Roudier et al., 2017).

Evidently, these approaches do not generate maps; they can only generate point-like data where the soil properties at distinct locations (i.e. those corresponding to the sampling locations or the locations where the in-situ measurements took place) are known. Thus, they replace only the chemical laboratory; in other words, in order to generate maps from these point data another step is still necessary.

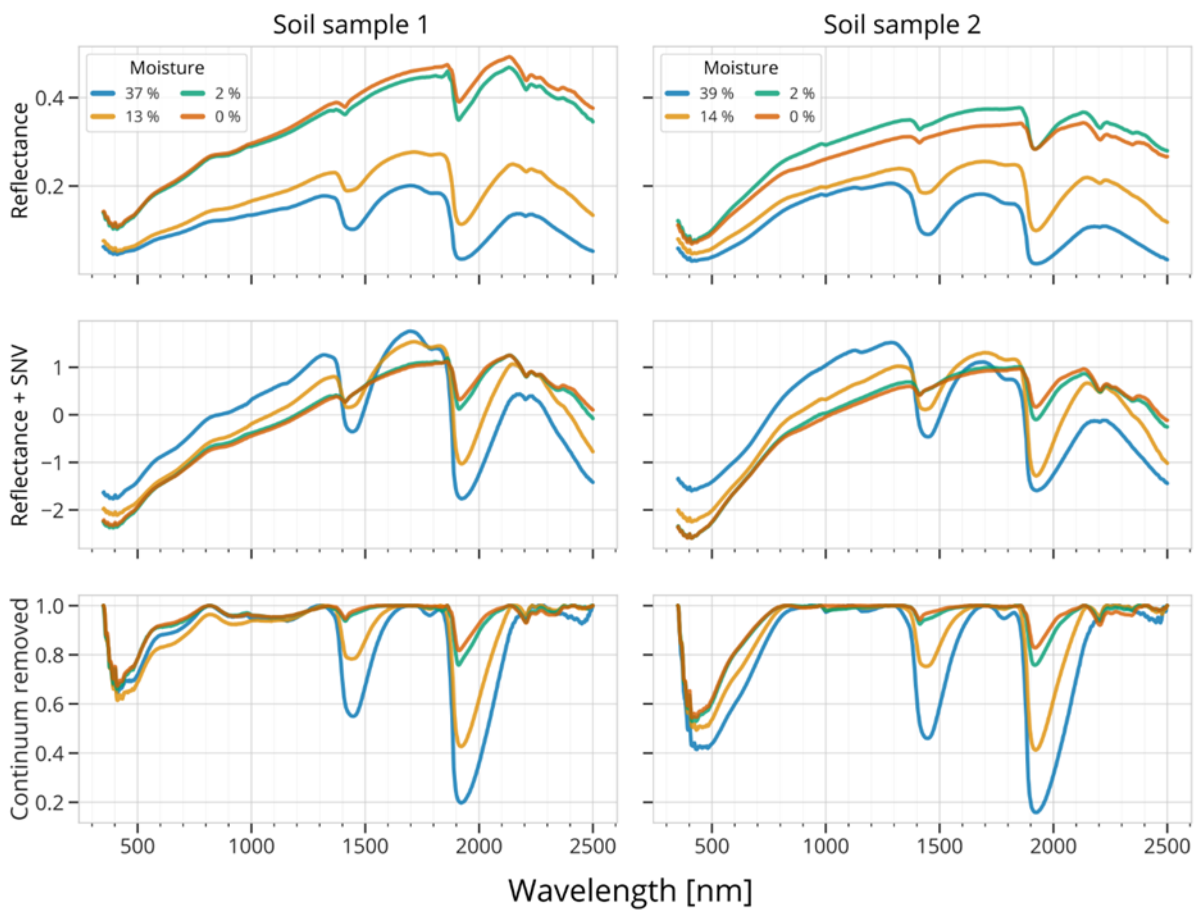

Figure 3: The effect on soil moisture on the reflectance spectrum of two soil samples in VNIR–SWIR (spectral data taken from i-BEC's data repository)

2.2.3 EO-based image spectroscopy

On the past decades, satellite remote sensing has been proven a valuable tool and a cost-effective solution, and thus the continuous provision of such kind of data with high revisit time and with high spatial and spectral acqurary, lead to innovative and scalable sollutions in various scales (local-national scale). In particular, hyperspectral imaging (HSI) has enabled a variety of applications in Earth studies, providing significant improvements on spectral measurement capabilities over conventional systems. It contributes to the identification of different materials, especially retrieving, and quantifying their chemical and structural characteristics as well as their temporal and spatial variations. Numerous of studies (Ben-Dor et al., 2009; Chabrillat et al., 2019) have analysed the critical importance of reflectance spectroscopy, as it gives continuous measurements over the monitored area, and thus reveal information of the soil properties and their temporal variation. It provides important information for integrated approaches combining satellite data with specific tools and modelling techniques. Satellite imagery is a cost-effective evaluation over extensive areas which can provide data with high revisit times, whereas in-situ measurements are more resource demanding but important to provide ground truth and validation data.

The use of optical remote sensing observations and in particular reflectance spectroscopy at the remote sensing scale, referred to as imaging spectroscopy (IS) or hyperspectral imaging, have been shown to be powerful techniques for the quantitative determination and modelling of a range of soil properties. These soil properties include topsoil mineralogical composition such as SOC content, textural composition, iron or carbonate content, and other physicochemical properties.

HSI opposed to the multispectral imaging (MSI), (Figure 4) have the ability to exploit the spectral information received in tens or hundreds narrows of the spectral bands, from the visible to the short-wave infrared (400-2500nm) part of the electromagnetic spectrum (EMS). Therefore, the combination of the HSI datasets with in-situ measurements can estimate the presence of certain soil properties in additional areas. The attractiveness of imaging spectroscopy is that measurements are rapid and estimates of soil properties are inexpensive compared to conventional soil analyses, as it exploits the information carried out by the visible and near-infrared (Vis–NIR or VNIR: 400–1100 nm) and shortwave infrared (SWIR: 1100–2500 nm) part of the electromagnetic spectrum. Especially, in the SWIR several overtones and combinations of absorbance features (whose fundamental absorption bands reside in higher wavelengths) may be found due to the presence of a number of soil chromophores.

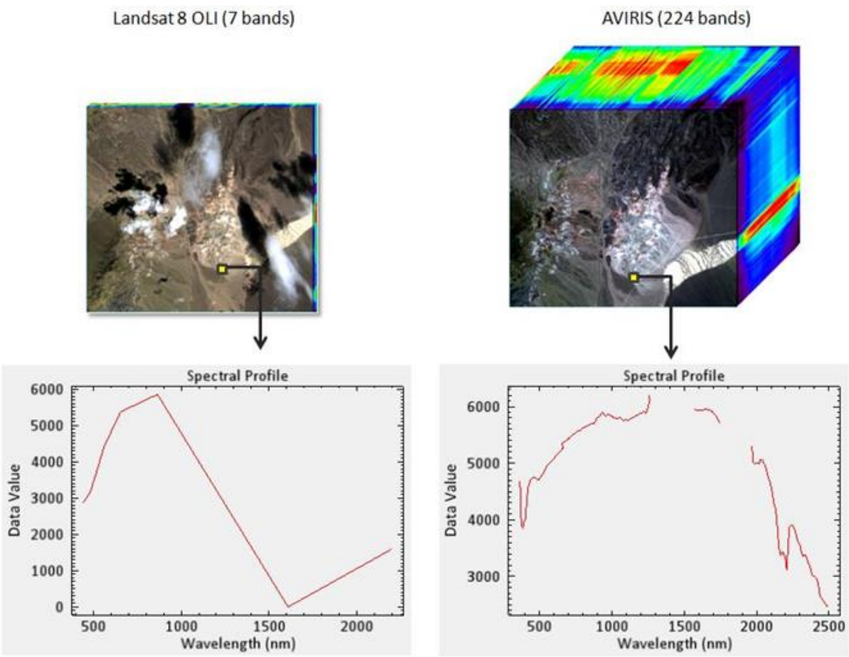

Figure 4 Comparison of MSI (left) and HSI (right) spectral signatures, (source: https://www.harrisgeospatial.com)

In the case of the identification of the bio-physico-chemical properties of soils in croplands, (Figure 5), a cloud-free EO image is used and the following processing steps are implemented. In particular, the workflow processing chain is initiated through the identification of the bare soil areas in a cloud-free EO-based image of an area, which can be achieved through various techniques as shown in section 3.1.1. A recent study has showcased how often this takes place in the European continent for cropland soils throughout the year (Figure 5). After that, the spectral signatures of the bare soil fields may be extracted, where for each different parcel it is possible to have multiple acquisitions within each year. If multiple image acquisitions are examined, then this gives rise to another significant issue vis-à-vis the selection of which spectral signature (or which combination thereof) should be used. In any case, the spectrum recorded is then prone to the same effects as the in-situ proxy spectra acquired during field visits: those of soil moisture and soil roughness, that pose a challenge to the inference process.

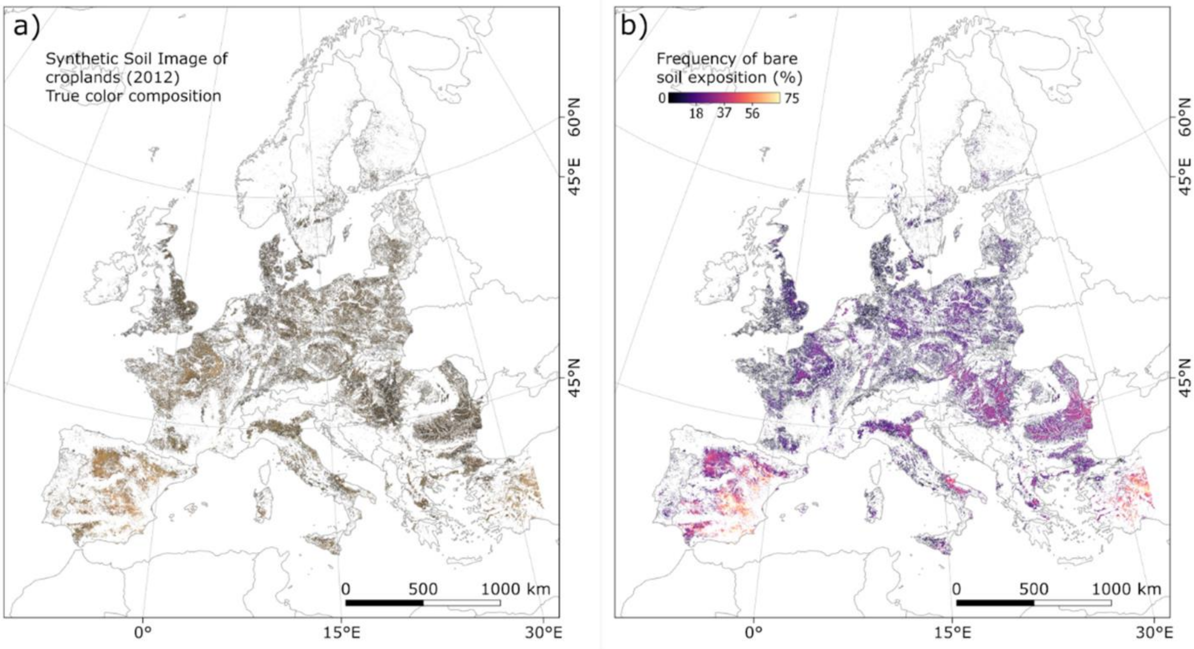

Figure 5: (a) A full synthetic soil image of croplands over the European extent; (b) the equivalent bare soil frequency (Safanelli et al., 2020).

3 DIONE Methodological Framework for Soil Monitoring

The following flowchart (Figure 6) illustrates the DIONE's methodological framework for soil monitoring, in which the approach operates as follows. On the basic layer historical and open EO data will be used (data tier), -interaction between the tools and the data (modelling tier) as well as how everything is brought together in order to determine soil properties (knowledge tier part) in map output format.

The present section outlines the approach in three distinct tiers:

1. **Data tier:** Exploitation of historical and open EO data, detailed in Section 3.1 which describes on the one hand the space-borne data that will act as the basis on which the soil maps will be developed, and on the other hand all relevant data from the in-situ component;
2. **Modelling tier:** Described in Section 3.2 where the processing of the data takes place and;
3. **Knowledge tier:** and Described in Section 3.3, presenting the final products.

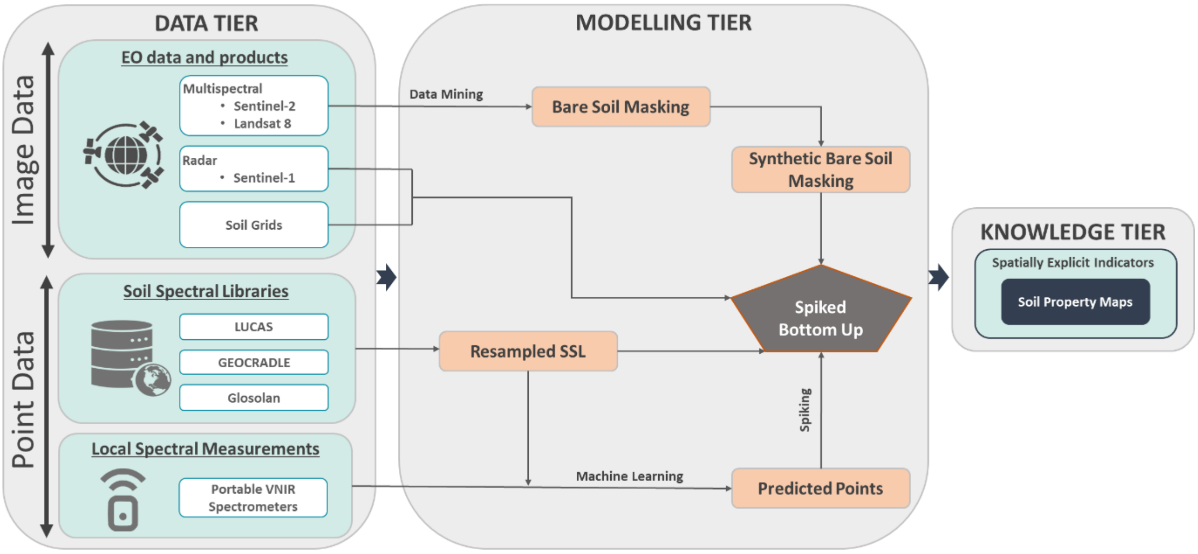

Figure 6: DIONE soil monitoring methodological framework schematic diagram conscerning the derivation of spatially explicit soil carbon indicator based on the spiked bottom-up approach.

3.1 Historical and open EO data tier development and processing

3.1.1 Earth Observation data and products

In the context of DIONE, a fusion of various data is implemented using either archive or open-accessed EO data of both multispectral sensors (Landsat-8 and Sentinel-2) and SAR sensors (Sentinel-1), (Figure 7). Geospatial information from the SoilGrids global soil repository, is additionally used.

In general, Copernicus Sentinel missions which have been launched by the European Commission since 2014, have been proven the cornerstone of the recent scientific community, providing the most accurate and cost-efficient solution in land applications, which enables the development and deployment of various models and services in multiple scientific domains (e.g. risk management, human health, environmental resilience, urban sustainability, etc.). The aforementioned satellite missions were deployed under the same rationale with the open-accessed information of Landsat missions. In particular, Landsat satellite missions (4TM-8OLI) are part of the Landsat Data Continuity

Mission (LDCM) and were developed as a collaboration between the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey (USGS) in order to provide services in numerous domains and scientific fields. Subsequently, SoilGrids is a global digital soil mapping system that makes use of global soil profile information, covariate data and state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe[1].

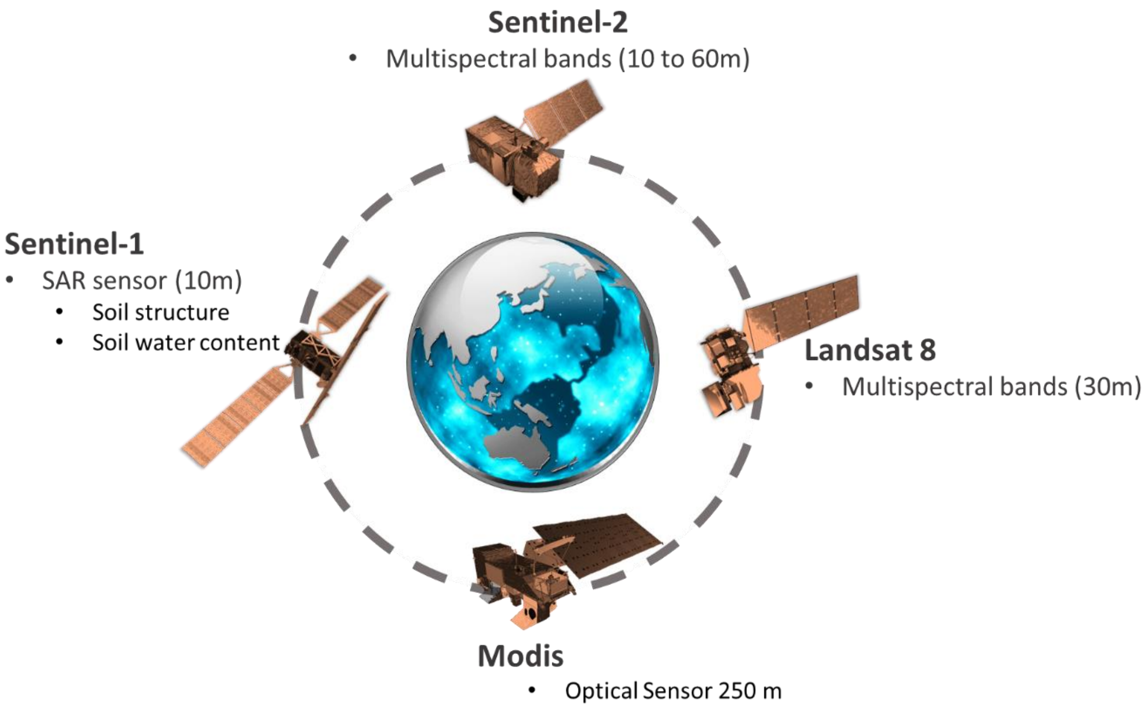

Figure 7: The EO ecosystem in which DIONE methodological framework operates

3.1.1.1 SoilGrids

Recently the international community has paid an increasing attention to improving legacy soil data resources in support of sustainable development (Folberth et al., 2016; Montanarella and Vargas, 2012). In order to contribute to the Global Soil Partnership initiative and to reduce the gap between soil data demand and availability, the International Soil Reference Information Centre (ISRIC) – World Soil Information, released a Global Soil Information system called SoilGrids[2]. It is a collection of global soil property maps at six standard depth intervals **at a spatial resolution of 250 meters**, produced using state-of-the-art machine learning methods. SoilGrids uses global models that are calibrated using all available input observations and globally available environmental covariates. SoilGrids spatial predictions (layers) are produced using a reproducible soil mapping workflow and can therefore be regularly updated as new soil data or covariates become available, after quality control and data standardization/harmonization. Prediction models are fitted using over 230.000 soil profile observations from the WoSIS database and a series of environmental covariates.

SoilGrids provides global predictions for the following standard numeric soil properties: pH, soil organic carbon content, bulk density, coarse fragments content, sand content, silt content, clay content, cation exchange capacity (CEC), total nitrogen as well as SOC density and SOC stock. Data are

[1] https://www.isric.org/explore/soilgrids
[2] https://soilgrids.org/

available and visualizing through a web GIS platform (soilgrids.org) and can be downloaded in raster format via FTP, or through RESTful services of the Soil Info App. Ledo et al., 2019 used the aforementioned datasets in order to take information about the soil texture and the chemical properties. Undoubtfully, SoilGrids dataset is proven to be the most successful approach resulted from the synergistic use of satellite imagery (primarily derived from MODIS land products) and machine learning methods, in order to produce digital soil maps at global-scale. Therefore, until now there isn't any upgrade regarding the spatial resolution of the delivered soil products, as they are available in in MODIS coarse resolution (250m), (Figure 8). Thus, although it is a great attempt as end-product it does not meet the needs presented in *D5.1 Environmental Monitoring Framework* for monitoring soil properties at a parcel level. Considering the above, in DIONE we aim to exploit the open-access satellite data and further leverage on the super spectral imagery at finer spatial resolution in order to estimate the soil properties in the examined areas, at finer spatial resolution.

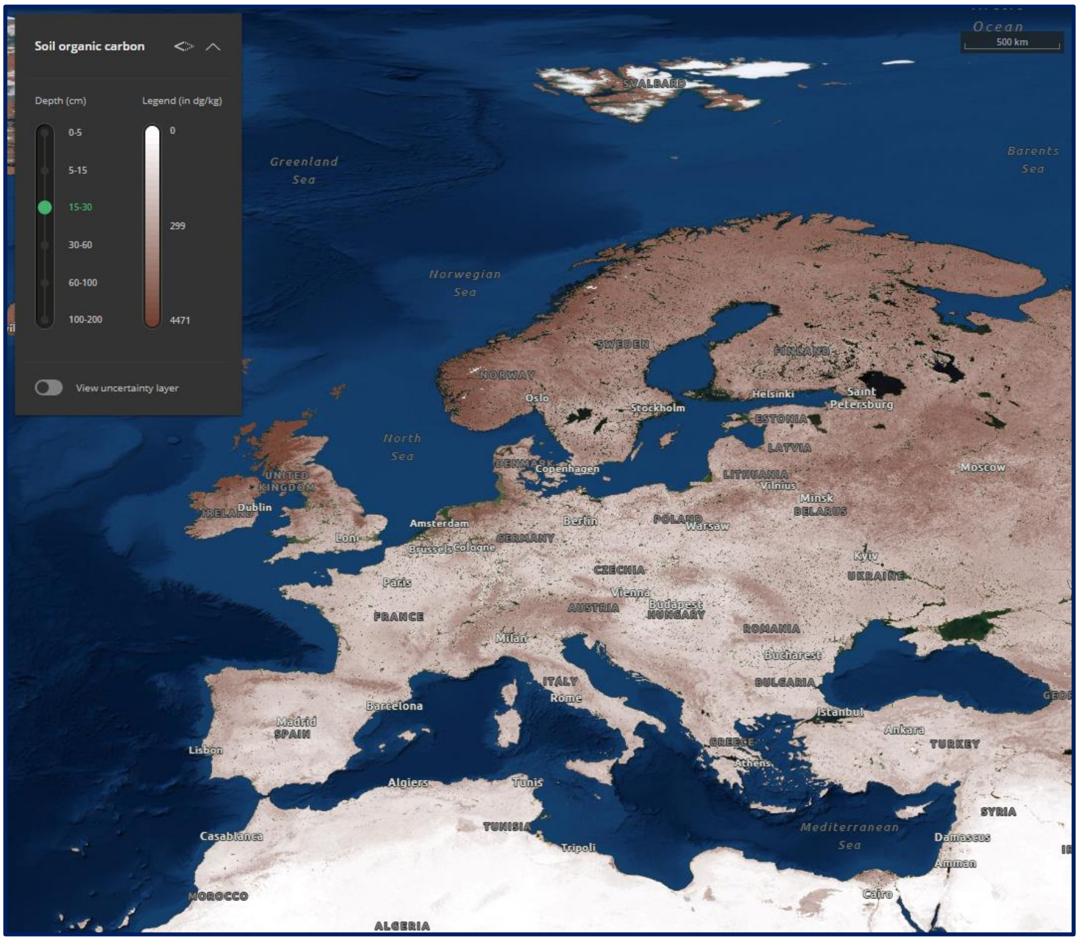

Figure 8: SOC concentration around Europe using the SoilGrids open access platform. https://soilgrids.org/

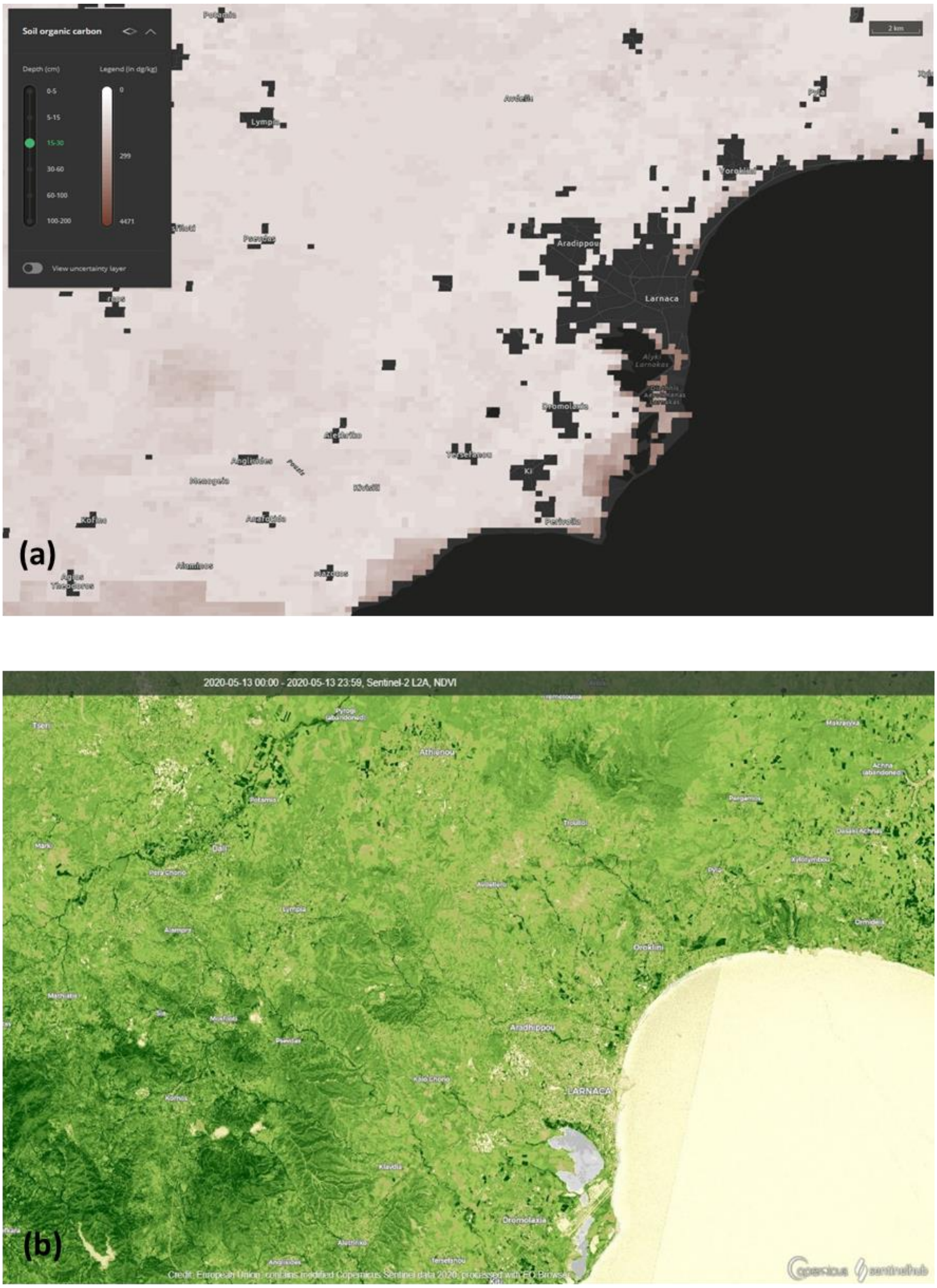

Figure 9: Cyprus area with (a) Soil Grids SOC concentration (b) Sentinel-2 NDVI from EO broswer.

3.1.1.2 Satellite optical Multispectral data

Multispectral satellite sensors capture spectral data in different parts of the EMS, but usually in VNIR-SWIR parts. MSI data received from Landsat-8 and Sentinel-2 allow the implementation of a timeseries analysis, revealing the diachronic evolution of the organic matter, whereas based on this detailed maps of soil properties (e.g. SOC maps) can be retrieved in finer spatial resolution. In this regard, sensors on Sentinel-2 and Landsat 8 satellites have considerable potential for detailed mapping of soil properties like SOC (Žížala et al., 2019). Studies have already confirmed this potential (Ben-Dor et al., 2008; Castaldi et al., 2019b, 2016; Gholizadeh et al., 2018). Considering the above, Castaldi et al., (2019b) evaluate the validity of the spatial resolution of Sentinel-2 for SOC mapping both in local and regional scale. In the context below additional information is given regarding the two satellite missions, their characteristics and capabilities, as well as their synergic use in terms of estimating the soil properties in various examined areas.

Description of Landsat-8 data

Landsat 8 provides datasets over Earth's terrestrial and polar regions in VNIR–SWIR and thermal infrared. Landsat 8 includes a multispectral image scanner-OLI (Operational Land Imager) and a thermal image scanner-TIRS (Thermal InfraRed Sensor), receiving data at the spatial resolution of 30m and 100m respectively. Additionally, the Landsat imagery includes the panchromatic spectral band which covers the visible wavelength at finer spectral resolution (15m)[3]. The OLI provides two additional spectral bands, one tailored especially for detecting cirrus clouds and the other for coastal zone observations. The TIRS collects data for two more narrow spectral bands in thermal region.

Description of Sentinel-2 data

Sentinel-2's Multi-Spectral Instrument (MSI) features 13 spectral bands from the visible and near-infrared (VNIR) to the short-wave infrared (SWIR), featuring four at 10 m, six at 20 m and three at 60 m resolution. The best compromise in terms of user requirements and mission performance, cost and schedule risk, it provides enhanced continuity for Spot and Landsat, with narrower bands for improving identification of features, additional red channels for assessing vegetation, and dedicated bands for improving atmospheric correction and detecting cirrus clouds. This way, recent studies exploits the Sentinel-2 imagery for applications mainly in land monitoring, emergency management and security[4]. Furthermore, the twin Sentinel's (Sentinel-2A and Sentinel-2B) were launched, providing capabilities of continuous monitoring over the globe with high temporal resolution (10 days at the equator with one satellite, and 5 days with 2 satellites under cloud-free conditions which results in 2-3 days at mid-latitudes). In this project, the spectral bands of 1, 9 and 10 will be excluded for the further processing as they mainly contribute to applications related to the atmospheric correction (e.g. precise aerosol and cirrus correction).

Synergies between MSI data and applications The multispectral data from the Landsat-8 has been utilized into a promising framework to map soil texture based on a powerful data mining procedure to retrieve spectral reflectance signatures (Demattê et al., 2018). In addition, recent studies (Castaldi et al., 2019b; Taghadosi et al., 2019; Vaudour et al., 2019) have paved the way to the exploitation of Sentinel-2 data which along with advanced regression analytics reveal similar results as the hyperspectral data from HySpex, on estimating the soil properties (e.g. SOC) of the examined area, (Figure 10).

[3] https://www.mdpi.com/2072-4292/8/3/180/htm
[4] https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/applications

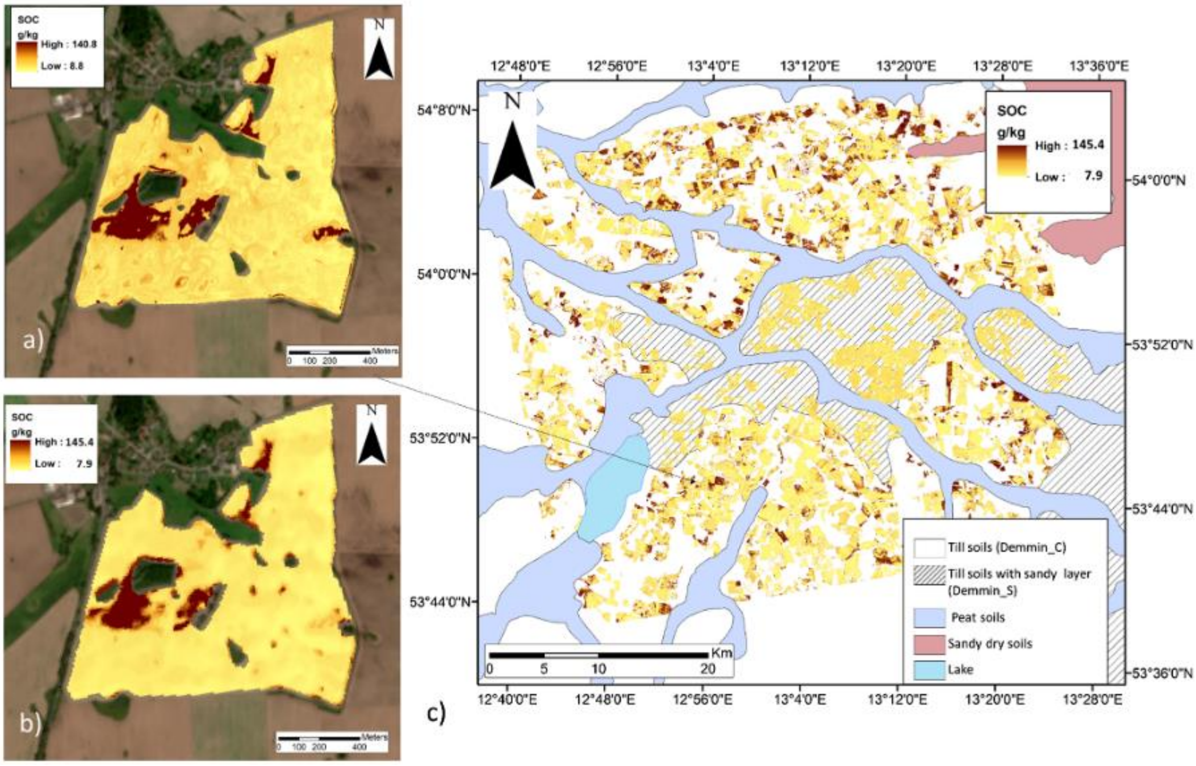

Figure 10: Soil organic carbonm (SOC) maps of a singular field in Demmin-C area obtained by HySpex (a) and real Sentinel-2 (b) data. On the right hide side (c), the SOC map at regional scale retrieved from Sentinel-2 data in the Demmin_C and Demmin_S sub regions (Castaldi et al. 2019).

Additionally, synergies of Sentinel-2 and Landsat 8 offer a great opportunity for accurate estimation of soil properties, (Figure 10) as they cover the VNIR–SWIR region, a spectral region where the soil characteristics (e.g. colour, grain size, etc.) can be identified through the presence of chromophores which have been noted to absorb radiation in various wavelengths[5]. Continuing, Gholizadeh et al., (2018) extracted spectral signatures from Sentinel-2 bands, from four independent study areas involving different types of agricultural areas and used a support vector regression algorithm to produce spatial distribution maps of soil properties. Each pixel was enriched with additional features resulted from various band combinations, while this enhanced dataset provide high-quality information on variations in SOC comparing to airborne sensors.

Sentinel-2 constellation offers great functionalities in data mining techniques, while in collaboration with the low cost, portable spectrometers (MEMS), predictions of soil organic matter can be achieved with great accuracy. In details, the RGB bands can improve of bare soil extraction and give better coverage of the study area (i.e. more accurate parcel discrimination). It should be noted that the Sentinel 2 data can be super-resolved so that the bands of coarser spatial resolution can match the finer ones (see Deliverable *D3.1 Analysis of the software specifications*). Landsat-8 with the additional source of thermal information illustrate a promising capabilities on the estimation of the soil moisture content, which alters the reflectance values (presence of spectral absorptions, or reduce of the reflectance response), and effects on the prediction accuracy, (Karyotis et al., 2020).

[5] https://www.researchgate.net/publication/200458942_Soil_Reflectance

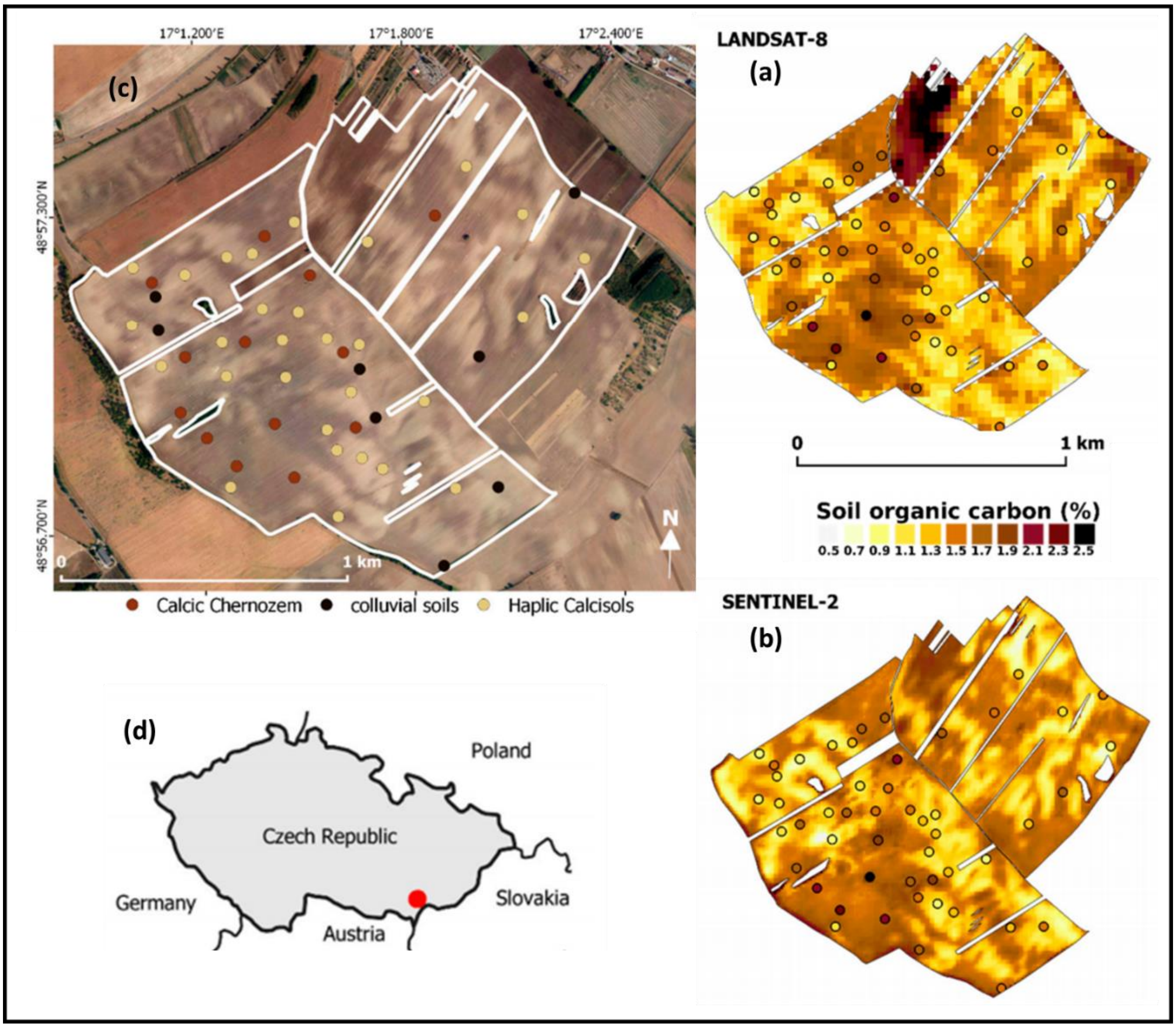

Figure 11: Spatial distribution SOC based on prediction using (a) Landsat 8, (b) Sentinel-2. Where (c) and (d) study site location (Žížala et al., 2019)

3.1.1.3 Satellite Radar Data

The **Sentinel-1** is a Synthetic Aperture Radar (SAR) instrument. It operates C-Band in four exclusive imaging modes with different resolution (down to 5 m) and coverage (up to 400 km). It provides dual polarization capability, very short revisit time and rapid product delivery. The mission is composed of a constellation of two satellites, Sentinel-1A and Sentinel-1B, sharing the same orbital plane[6].

Sentinel-1 have the ability to transmit and receive the backscatter signal either with single polarisation (VV or HH) or with cross-polarisation (VH or HV). Sentinel-1 data are sensitive in three physical factors, the surface roughness and the local topography and the **dielectric constant**, where the last on make them very promising indicators for the estimation of the **soil structure** and the **soil moisture content**.

In details, the estimation of soil moisture over bare soil areas or areas with sparse vegeration is implemented using the SAR signal as an input in physical or statistical models. The procedure of site-specific calibration is not required for physical models making them a method to simulate the backscattering coefficients from radar configuration and soil parameters, (Hajj et al., 2017). Recent

[6] https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-1-sar/sar-instrument

studies have evaluated the capabilities of Sentinel-1 data in the estimation of soil moisture and soil structure. (Ambrosone et al., 2020; Ayehu et al., 2020; Tziolas et al., 2020a). Additionally, their ability of penetrating the clouds and the receiving data both in day and night, prevents significant drawbacks in model's performance due to the cloud presence. Considering the above, Sentinel-1 data will be exploited in DIONE, as they will contribute in reducing of cloud artifacts over the examined areas and in the identification of areas affected by the presence of the soil moisture and soil structure.

3.1.2 Soil Spectral Libraries

Moving from the remote sensing domain to the in-situ component, the soil libraries and soil spectral libraries (SSLs) are further used. Along with the growing availability of EO data, the open-access SSLs have enabled a data-driven approach to effectively describe soil, both qualitatively and quantitatively, finding robust statistical associations between laboratory spectral signatures and soil properties (Ballabio et al., 2016; Nocita et al., 2015). Furthermore, SSLs provide, in addition to spectral data, analytical data on a number of soil variables, allowing the calibration of multivariate models covering larger soil variability than the models calibrated using local libraries. Among the most up-to-date and historical European datasets are the European Soil Data Centre (ESDAC) Land Use and Cover Area frame Survey (LUCAS) and the GEO-CRADLE Soil Spectral Library (GSSL) which in this work will be used for implementation based on the major advantages that can offer in the framework of this project. Moreover, Global Soil Laboratory Network (GLOSOLAN) SSL which is a future globally calibrated soil dataset will be explored for potential use since data are not yet available.

3.1.2.1 GLOSOLAN

The GLOSOLAN was established in 2017 to build and strengthen the capacity of laboratories in soil analysis and provide a modern solution to the need of soil spectral data harmonization. GLOSOLAN is working to improve the proficiency of soil laboratories in both wet and spectroscopy, while the areas of interest are extended over the two following pillars:

- the harmonization of **Standard Operating Procedures** (SOP) for known methods of wet chemistry and
- the development of a global, and representative **Soil Spectral Library** over MIR spectra.

The developed SSL will contain Global MIR Soil Spectral measurements with matched conventional soil property data. Both spectral and wet chemistry data will measured in one gold standard laboratory, and will provide estimates of soil properties of increased accuracy. The library will be freely available to laboratories around the world. This effort will leverage agricultural productivity and mitigate land degradation through providing information regarding soil health.

Although, GLOSOLAN can be characterized as more than worthwhile, unfortunately, the product is still not available, making its use within the DIONE's framework, impossible. Therefore, the use of existing SSLs will be utilized as described in the following sections.

3.1.2.2 LUCAS

The statistical office of the European Union (EUROSTAT) organizes a triennial survey of land use, land cover and changes over time across the member-states of the European Union, known as the Land Use and Coverage Area Frame Survey (LUCAS), with the latest survey conducted in 2018. The LUCAS Programme started in 2001 as an area frame survey organized and managed by Eurostat and is considered the largest harmonized open-access tool of topsoil properties at global scale, with data freely available from the ESDAC. In 2009, the European Commission extended LUCAS to additionally collect topsoil samples and analyse their key topsoil properties in 23 member-states. This survey is based on the visual assessment of parameters that are deemed relevant to agricultural policy. The topsoil sampling locations were selected using a Latin hypercube-base stratified sampling design from the LUCAS master sample grid of 2 km by 2km, (Castaldi et al., 2019a). This topsoil survey was an attempt to build a consistent database using standard sampling and analytical techniques, where the analysis of all soil samples was carried out by a single chemical laboratory.

The LUCAS 2009 topsoil dataset is a large European soil spectral library (about 20.000 samples from 25 EU countries), which was collected in the framework of the European land use/cover area frame statistical survey in 2009 (Tóth et al., 2013). The latest LUCAS surveys were undertaken in 2015 and 2018 and covered the 28 EU Member States with sampling expanded to cover locations at altitudes above 1000m and with 21.859 total number of data points, (Figure 12). The 2015 LUCAS data became publicly available during the final stages of the preparation of this document (early October 2020). It should be pointed out that this delay between the actual data sampling and the time the data are available for research use is another obstacle inserted by the traditional approach; a modern approach should provide results in a more timely fashion, which can enable end-users to take the appropriate action swiftly. The output format of LUCAS SSL is in CSV files to facilitate use of the data and XLS and ESRI point shapefile.

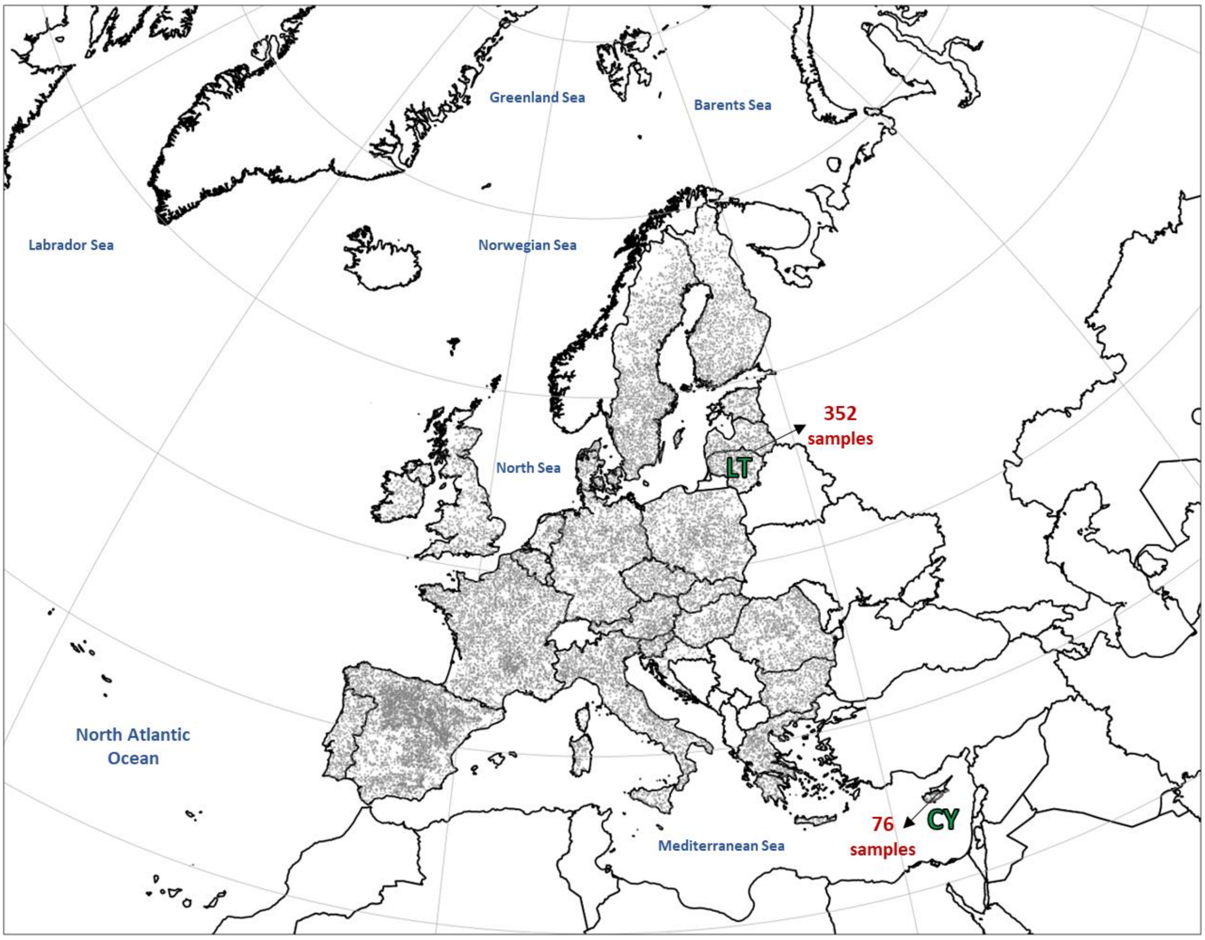

Figure 12: Overlap of LUCAS 2015 points around Europe region. DIONE pilot areas (Lithuania and Cyprus) are also appear in the map with the total number of samples.

Given the sufficient spatial distribution and representatives of LUCAS data archive across the European continent, we leverage on the reference soil data as the source for the soil texture information. It is an available library under the Open Data Base License with a big number of samples and in more specific context will be used for large scale mapping needs. In addition, the dataset includes laboratory soil spectral samples (VNIR–SWIR) that covers the same spectral range of the MEMS, and thus enables the model calibration of soil properties estimation. In the same context, the modelling procedure is largely supported, by allowing the resampling process to be performed in order to achieve the spectral configuration of the MEMS in accurately and reliably way.

3.1.2.3 GEO-CRADLE

The GEO-CRADLE Soil Spectral Library (GSSL) was developed in the context of the GEO-CRADLE EU funded project[7]. For the first time an open and standardized regional SSL was developed, as a complement to the EU Soil Sample Data Base. The GEO-CRADLE SSL further provided detailed thematic soil maps by analysing soil spectra with Sentinel-2 data.

The GSSL was created quite recently and contains 1754 soil samples and their corresponding soil properties (SOC, Texture, $CaCo_3$, CEC, NO_3, pH) in a very well spatial distributed area across nine countries in the North Africa, Middle East and Balkan region (Figure 13). The samples were selected from national soil data archives or collected through field surveyors during the project’s lifetime, following general guidelines, standards, and protocols to ensure consistent data collection and analysis. The GSSL is in compliance with GEOSS data principles and Open Database License standards and is publicly available through the GEO-CRADLE’s project regional datahub[8]. The GSSL follows a standardization process, which allows spectral measurements from different spectrometers to be comparable, thus it is future-proof and expandable. The need for the application of standard measurement protocols and internal soil standard methods amongst the collaborating actors, to optimally allow the processing of the collected set of heterogeneous data due to diverse sources and spectrometers, towards the enhancement of predictive performance has been highlighted in previous studies (Kopačková and Ben-Dor, 2016; Romero et al., 2018). Contrarily with the GSSL, LUCAS SSL follows no such inter-calibration procedure.

The development of the GEO-CRADLE SSL facilitates further development of a global soil property and spectral database and supports the regional contribution to global soil mapping activities and other regional GEOSS hubs. In this regard, GSSL shows the way to take advantage of other efforts and actions that are in process such as the H2020 Soils for Africa program that is going to develop a soil spectral database for the African continent but only if there is the same structure for data development.

The GSSL datasets are provided in csv formats per country SSLs or as a complete standalone file. The data in the library are diverse, comprising soils of 18 soil classes of the world. Approximately 58% of the soil samples belong to the top soil layer (0-30cm), 20% at depths within 30-60cm, and 20% from 60 to 100cm, while the rest 2% originate from samples collected at depths >100cm. The soil samples were collected with geographical coordinates in the World Geodetic System (WGS84).

Taking the above advantages that the GSSL can provide into consideration, and with combination of I-BEC’s team expertise on know-how of how to best apply the SSL and derive products, GSSL’s use is considered more than determinant for the implementation of this work as it will take an active part in the modelling processes by feeding the MEMS with the same measurement standardization protocol.

[7] http://geocradle.eu/en/
[8] http://datahub.geocradle.eu/dataset/regional-soil-spectral-library

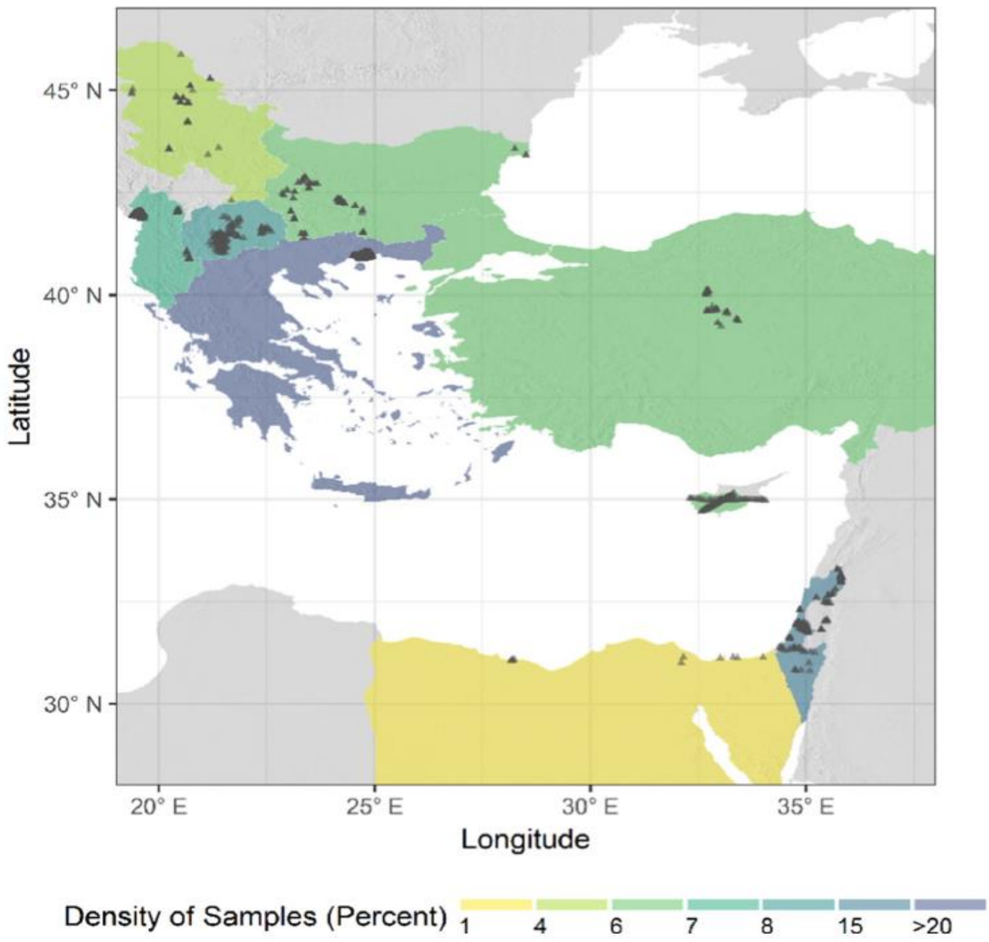

Figure 13: Location of the 1754 sample sites with reflectance spectra in the GEO-CRADLE SSL (Tziolas et al., 2019).

3.1.3 Micro Electro Mechanical Systems

Moving from the laboratory to in-situ measurements, i.e. captured directly at the field during the visits, the next component is the spectral measurement using the low-cost MEMS sensors. **Micro Electro Mechanical Systems** (MEMS), which are low-cost portable and handled spectrometers in the visible and near-infrared (VNIR) range, can contribute with an effective way in soil monitoring. They offer some advantages over the contemporary approaches, and thus they support decision makers in agricultural systems by providing spatial explicit information, which can assist in the protection of the soil ecosystem, help farmers by maximizing yields, and also promote the sustainable production. Because of their portable use, MEMS are able to improve the farming productivity by bringing real-time information of soil status, through a wireless connection with the sensors, while this will dramatically reduce the cost associated with the laboratory soil analysis. The aforementioned soil data in collaboration with novel machine learning techniques can estimate with efficient accuracy levels the physical and chemical status of soils, including: SOC, particle size distribution (sand, silt, clay), electrical conductivity, pH, total nitrogen, etc.

A big advantage of this tool is that VNIR spectra and the associated metadata can be transmitted over the air (OTA) using TLS technologies to a central database, which can confirm their integrity and evaluate their quality (novelty detection). For more information the reader is referred to *D4.1.*

Technical specifications of the in-situ soil scanning system (SSS), data processing system and farmer's geo-tagged photos framework. Moreover, their integration with satellite and other EO data via transfer learning techniques can then enable the large-scale mapping of soil properties. Finally, the continuous data flow can further enable the analysis on the temporal domain, which can e.g. identify normal/abnormal situations, perform novelty detection, and map and monitor land degradation intensity. A very interesting notice, regarding the MEMS cost, is that it is still unclear if spectrometers at lower costs can provide sufficient prediction accuracy in certain soil analysis applications, as no comprehensive analysis has been performed regarding the usefulness of these instruments, (Tang et al., 2020). However, based on our preliminary analysis, we ensure that the soil properties will be predicted at great levels of accuracy.

Figure 14 presents a typical representation of a MEMS point field measurement with wireless connection to a mobile.

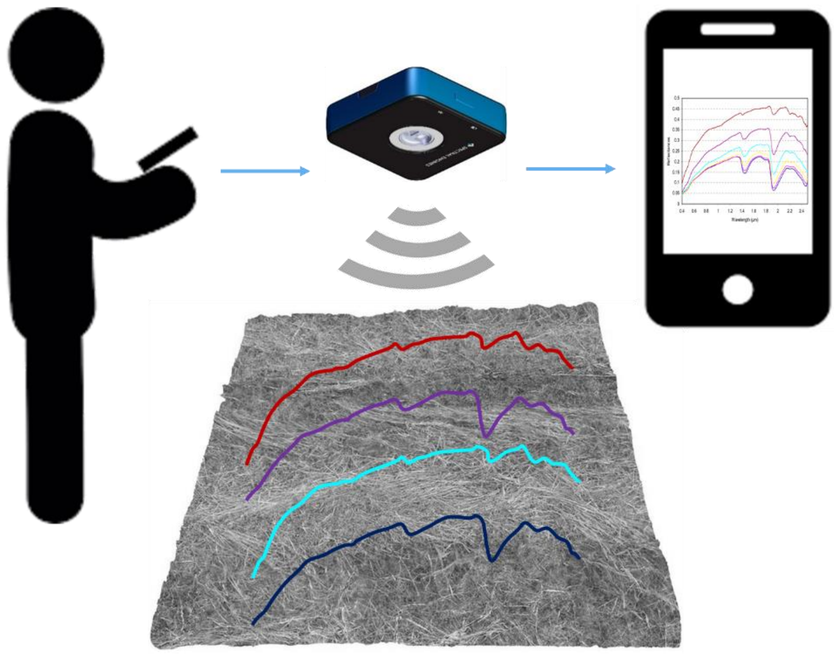

Figure 14: A typical represention of a MEM point field measurement with wireless connection to a mobile device.

One of the most important thing that must be taken into account is the distributive nature of the significantly lower cost and portable MEMS VNIR spectrometers and their in-situ usage will diminish the large cost associated from expensive VNIR spectrometers and expensive soil sampling campaigns. In addition, the need to calibrate, verify, and supplement the information provided from satellite data in a cost effective way makes the use of the MEMSs crucial in the whole DIONE's soil monitoring methodological framework, particularly pertaining to the spiked bottom up approach (see section 3.2.2) in order for the satellite imagery to provide accurate and reliable data in near real-time.

As already mentioned in the DIONE's *D4.1. Technical specifications of the in-situ soil scanning system (SSS), data processing system and farmer's geo-tagged photos framework*, the important wavelength regarding the soil properties to be predicted, are from 1750 to 2150nm giving in that way the selection of the suitable sensor as well as the simple and quick protocol for the acquisition of the spectra.. Moreover, in the same Deliverable technical specifications of the data pre-processing system, including how the server ensures data integration and validity, are provided. In short, the suitable sensor is Spectral Engines NIRONE Sensor S (D2.2) works at 1750 to 2150nm with a quick measurement protocol that increases the default measurement height to 2.4cm and ensures inter- and intra-calibration between measurements and instruments.

3.2 Modelling tier development and processing

3.2.1 Processing EO data

With the rapid development of EO technology and continuous launch of new satellites, the amount of the EO data is getting higher and thus the need of new processing techniques seems mandatory, (Xia et al., 2018). From the technological perspective, focus has been given on the development of service-oriented architectures, facilitating the linkage between data resources and processing. The following sections (3.2.1.1 and 3.2.1.2) present the proposed algorithms that are proposed to implement regarding the bare soil masking as well as methods to handle the big EO data.

3.2.1.1 Bare soil masking

Considering the extraction of the soil properties maps, first we need to identify and process areas described with vegetation absence. However, in terms of cropland monitoring, the bare soil areas are usual occurred at least one per annum, (Figure 15).

Figure 15: Variation of semi-covered parcel as captured by Sentinel-2 images (source: Sinergise Ltd)

In order to control this process, a **masking processing step** will be performed to differentiate the vegetated areas from the exposed soil. A wide range of activities will be performed to mask out bare soil pixels. These range from research works (Demattê et al., 2018). An example of this process is the Demattê et al, (2018) analysis which used spectral indices derived from Landsat datasets (NDVI & NBR2) and a thresholding process to construct a bare soil map at global scale, as a proxy for soil monitoring. Similarly, SCMaP processor (Rogge et al., 2018) delivers a geospatial layer, where the permanently vegetated areas are identified, and compared with the non-vegetated areas. This product is assigned to areas with temporary exposed soils, such as croplands. SCMaP used a modified NDVI that also integrates the visible blue for the reduction of cloud presence.

Furthermore, for the spectral modelling approach over semi-bare soil areas, additional indices can be evaluated for best results depending on chlorophyll content for the correction of saturated conventional vegetation indices such as NDVI. In that context, Sinergise LTD developed the bare soil marker to identify ploughing events by detecting exposed bare soil. A decision tree classifier is trained to identify the bare-soil areas, using a set of EO derived features[9] (NDVI, Bare soil index-BSI,

[9] medium.com/sentinel-hub/create-useful-and-beautiful-satellite-images-with-custom-scripts-8ef0e6a474c6

Normalized Difference Vegetation Index with Red Edge 3 – $NDVI_RE_3$, and chlorophyll index with Red Edge - CL_RE).

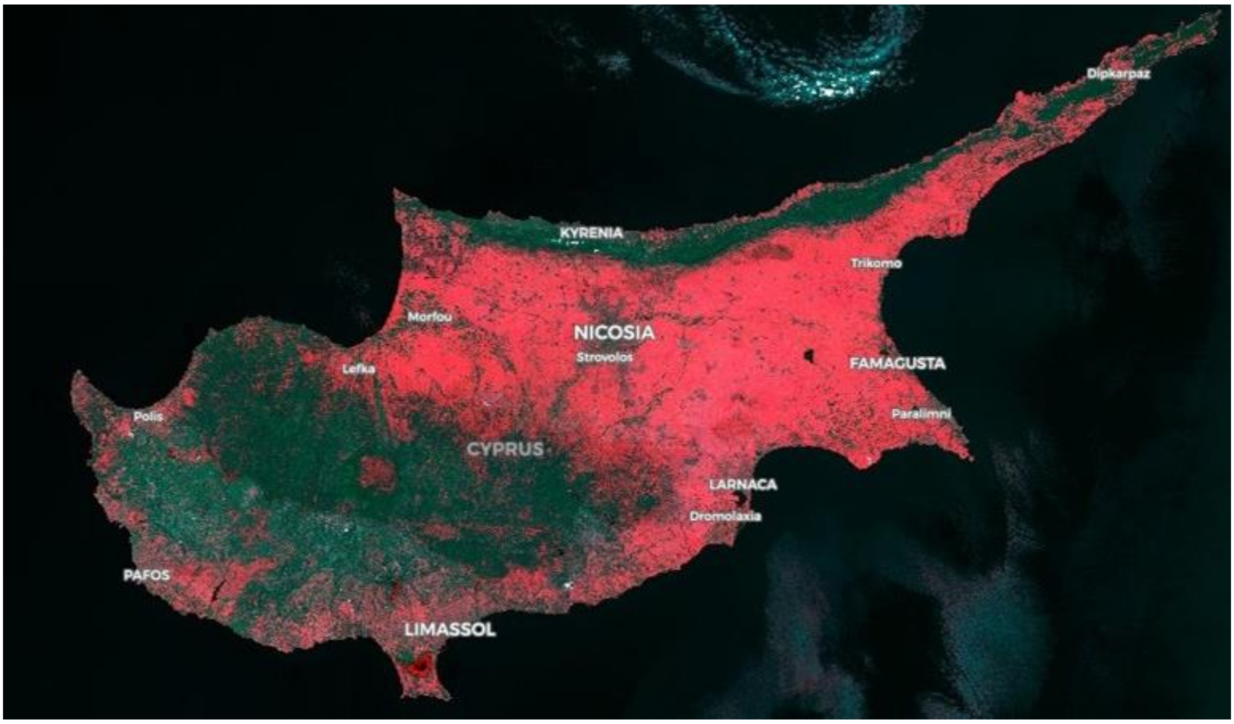

Figure 16: An overview of the bare-soil classifier overCyperus, taken on 2020–05–13. The red color marks the detected bare soi (credit: European Union, contains modified Copernicus Senitnle data 2020, processed with EO browser)l.

Within the framework of this project we are going to perform a bare soil filtering based on NDVI and NBR2 thresholds to acquire the raw soil surface spectrum. In this context, to exclude vegetated and mixed pixels, NDVI values will be derived from the Sentinel-2 imagery data, using the B4 and B8 bands. Additionally, the differences between B3 and B2 as well as B4 and B3 will be derived to remove some of the erroneous data by keeping only the ones with positive differences in these bands. Lastly, the NBR2 related to the dry vegetation presence in the pixels has been used in support of the selection of representative bare soil areas (Demattê et al., 2018). The optimum values of NDVI and NBR2 indices were selected by examining the correlation statistics between the corresponding resampled LUCAS soil spectral signature and the median surface reflectance of Sentinel-2 values from the selected bare soil pixels, in a previous study (Tziolas et al., 2020a). In that regard, we defined the NDVI and NBR2 thresholds eqal to 0.25 and 0.075, respectively.Further, for the spectral modelling approach over semi-bare soil areas, additional indices will be evaluated for best results depending on environmental conditions for the correction of disturbing factors such as soil moisture detection with the Non-Dimensional Water Index (NDWI), possible soil roughness albedo correction measures to minimize introduced uncertainties.

It should be noted that the synergetic use of index based approaches, such as SCMaP, and Sinergise's classification product is expected to be more advantageous than the single ones, hence, difference between both approaches and their synergy should be tested during the implementation of deliverable "D4.4 Implementation and development of systems; SSS, data processing and geo-tagged photos framework alpha versions".

3.2.1.2 Big data management

In the era of Big Data, where a tremendous wealth of information is captured daily, the combined impact of new computing resources and techniques with an ever-increasing avalanche of large datasets, is transforming many research areas. The Earth Observation has a pivotal role in the Big data era, and is a prime example of the recent Big Data technologies. The services of Copernicus and Landsat produce over 15 TB of open data daily that need processing and storage. Pursuing exploitation of these data as part of the EO downstream sector requires innovative application of mature ICT solutions; they are essential to address the related issues pertaining to both EO data processing, also characterized by the Big Data four Vs: volume, velocity, variety, veracity.

The Copernicus **Data and Information Access Services (DIAS)** are cloud-based platforms providing centralized access to Copernicus data and information, as well as to processing tools, thus facilitating and standardizing the access to the open Copernicus EO data. The DIAS[10] online platforms allow users to discover, manipulate, process and download Copernicus data and information. All DIAS platforms provide access to Copernicus Sentinel data, as well as to the information products from Copernicus' six operational services, together with cloud-based tools (open source and/or on a pay-per-use basis). In essence, the DIAS platforms are solve the accessibility problems to these large data repositories by providing easy and fast access as well as a variety of sophisticated processing tools and resources for users, without the need to download vast amounts of satellite data to a local computer (Figure 17).

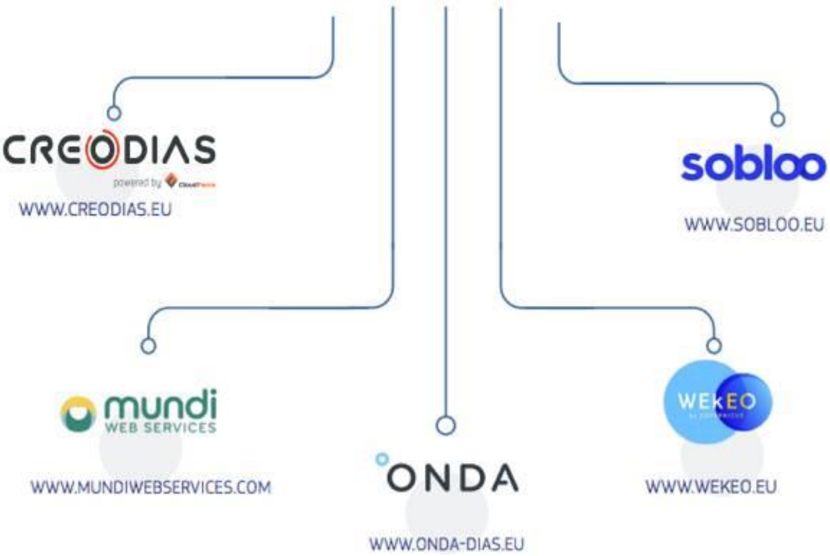

Figure 17: The DIAS platforms and where to reach them

Google Earth Engine is a geospatial processing service deployed on the Google Cloud Platform, enabling users to perform geospatial processing at scale. Its purpose is to (i) provide an interactive platform for geospatial algorithm development at scale, (ii) enable high-impact, data-driven science, and (iii) make substantive progress on global challenges that involve large geospatial datasets. The Earth Engine's public data archive includes more than forty years of historical imagery and scientific datasets, updated and expanded daily. Some of its datasets include the Copernicus data catalogue (Sentinel data, CORINE land cover, etc.), data from Landsat and MODIS, climate and weather data, and other geophysical data.

[10] https://www.copernicus.eu/sites/default/files/Copernicus_DIAS_Factsheet_June2018.pdf

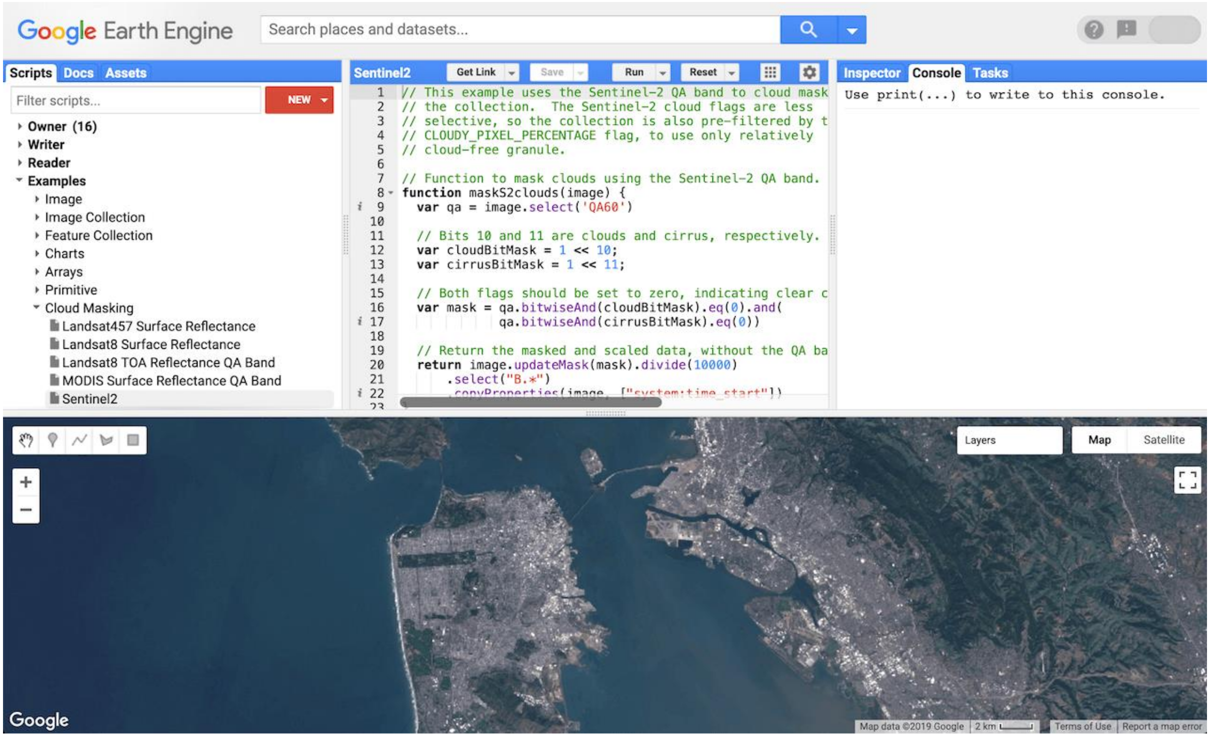

Figure 18: The Google Earth Engine platform

The **Open Data Cube (ODC)** is an Open Source[11] Geospatial Data Management and Analysis Software project that enables user to harness the power of satellite data. At its core, the ODC is a set of Python libraries and PostgreSQL database that helps user work with geospatial raster data. It provides the foundation of several international, regional to national scale data architecture solutions, such as Digital Earth Australia, Africa Regional Data Cube, and others. The Data Cube works well with Analysis Ready Data (ARD), pre-processed, as well as ready to use data made available by data providers. Any data available to the users can be installed to their own cube instance, including commercial, in situ, or otherwise derived products, without the need to share sensitive information and deploy them elsewhere. The Open Data Cube system is designed to: (i) catalogue large amounts of Earth Observation data, (ii) provide a Python based API for high performance querying and data access, (iii) give scientists and other users easy ability to perform Exploratory Data Analysis, (iv) allow scalable continent scale processing of the stored data, and (v) track the provenance of all the contained data to allow for quality control and updates.

[11] https://www.opendatacube.org/

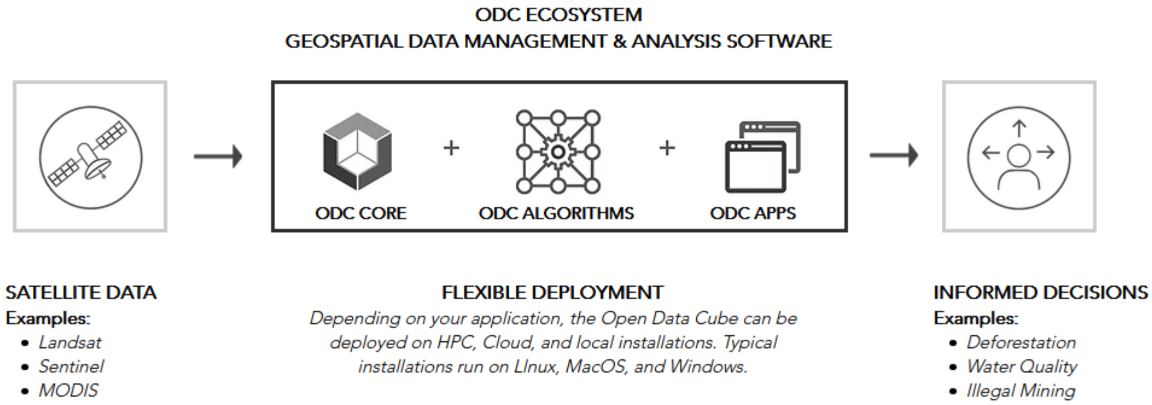

Figure 19: The ODC ecosystem

3.2.2 Spiked bottom-up approach

The European's Union Copernicus programme already highlighted an important problem which refers to the necessity of design effective integration strategies of data from satellite and airborne with ground-based measurements. In this context, a methodological framework is proposed leveraging on the data mining techniques with soil datasets will be estimated with the integration of in-situ and remotely sensed imagery data.

A **spiked bottom-up approach** is proposed to be implemented in the context of DIONE, integrating information from existing open SSLs and spectral imagery data, towards the delivery of quantitative and spatially explicit soil indicators (Tziolas et al., 2020b). This scheme is built upon the works of Castaldi et al., (2018) and data spiking, Brown (2007), in which local spectroscopic calibrations of soil properties are implemented based on novel machine learning NN and the combination of archive and updated local spectral information. In addition, a comparison between the spiked bottom-up approach methodology and other approaches such as *traditional approach* and a *bottom up approach* without spiking of site-specific samples, has been performed and present in Figure 20 (Tziolas et al., 2020b). Moreover, a study from Ward et al., (2019) built models based on the LUCAS soil spectral library and on few spectral measurement of local samples in order to develop an approach which is applicable to remote sensing imagery. Their approach benefit was the replacement of expensive wet chemistry analyses of soil samples from local study site, needed to calibrate remote sensing models, by soil spectroscopy.

The important point that we must underscore is that the spiked bottom up approach in its original format uses for the part of local spectral measurements the procedure of a laboratory soil analysis.

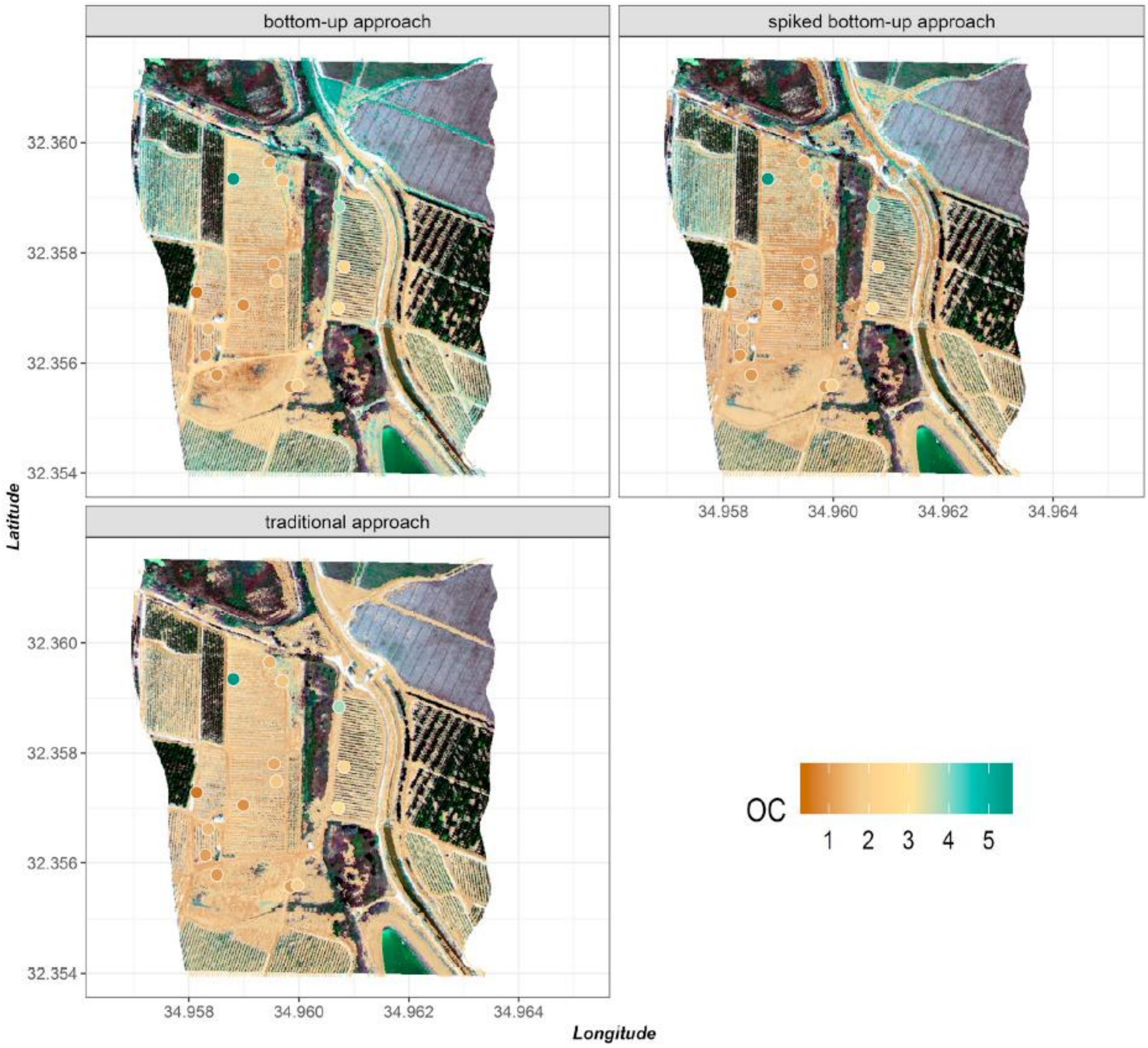

Figure 20: Comparison maps of SOC as generated by the (i)bottom-up, (ii)spiked bottom-up and (iii) traditional aprroaches from the AisaFenix image, with the points indicating ground soil samples used for validation purposes (Tziolas et al., 2020b).

The extra advantage in this approach is the novel use of the MEMS in the whole methodology instead of the spectroscopy/chemical laboratory soil analysis. This contribution is beyond the state of the art, making in that way the final approach a methodology without any laboratory soil analysis, spectra or chemical, and giving further speed to the observation but also even lower cost of use.

The spiked bottom-up approach scheme that will be utilized in DIONE, (Figure 21) involves two main parts:

1. Prediction of soil properties using MEMS point spectral signatures at the specific soil areas.
2. Prediction of soil properties using imaging spectroscopy across the whole site, using the predicted soil properties resulted from the part 1.

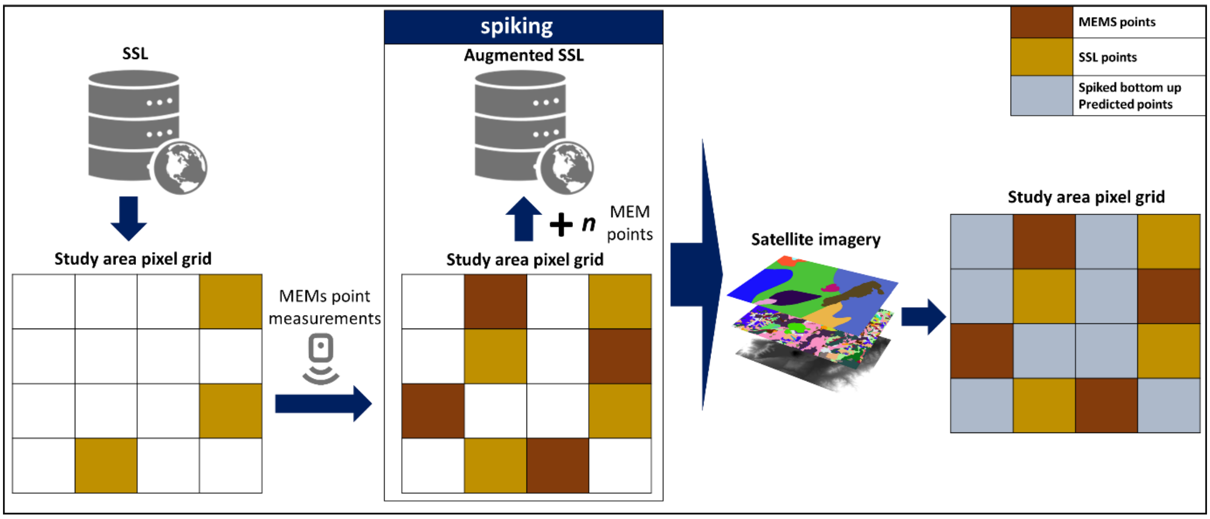

Figure 21: Overall DIONE spiked bottom-up approach. The number of pixels in figure's study area pixel grid is indicative.

<u>First step: point data for spiking</u>

The first step in DIONE's spiked bottom up approach is the MEMS spectral signatures in the field under certain protocols (see DIONE Deliverable D4.1). After this procedure and by using modelling techniques (machine/deep learning), spectral signatures will be transformed in soil properties.

One of the most promising families of machine learning algorithms are the deep learning approaches, with the convolutional neural network (CNN) being one of its most prominent representatives. Tsakiridis et al., (2020) proposed a localized multi-channel 1-D CNN which scored significantly better than four other machine learning methodologies commonly employed in soil spectroscopy, namely the PLS, Cubist, SVR and SBL algorithms (Figure 22).

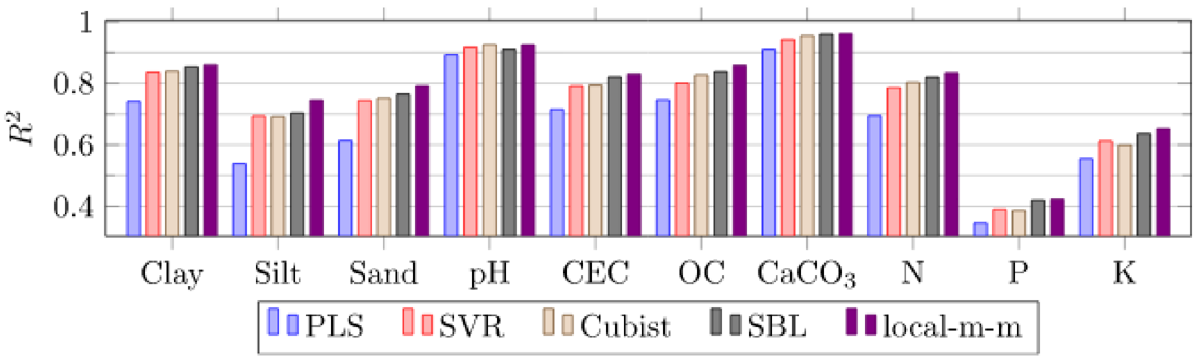

Figure 22: Visualization of the precision results (R^2, higher is better) for the 1-D CNN and the competing approaches (Tsakiridis et al., 2020).

A strong advantage of this CNN is that has some degree of interpretability, and allows the identification of key spectral signatures as well as interesting feature regions. Considering the above, we propose to use this approach in order to provide soil properties from MEMS spectral signatures.

In general, the main goal of the proposed algorithm is to incorporate the extra MEMS point soil data to the existing SSL (spiking approach) in order to augment it and thereby assist the machine learning techniques to provide more accurate predictions.

Second step: development of maps from spiked datasets

In the second step of spiked bottom-up approach, the estimated values from the MEMS will be used as an input to the calibration a model for the estimation of soil properties, such as SOC concentration. In details, a data-driven technique will be applied in order to extract the bare soil areas, based on approaches presented in Karyotis et al., (2020) and Tziolas et al., (2020a). To begin with, the model consists of a neural network architecture which learns higher level features from the raw bands of the satellite data, which enable to better identify the relation between the EO data and the examined soil properties. Following the aforementioned procedure, the additional processing of both Sentinel 1 and Sentinel 2 data will be examined to overcome the effects of soil moisture and roughness on VNIR–SWIR (Figure 23).

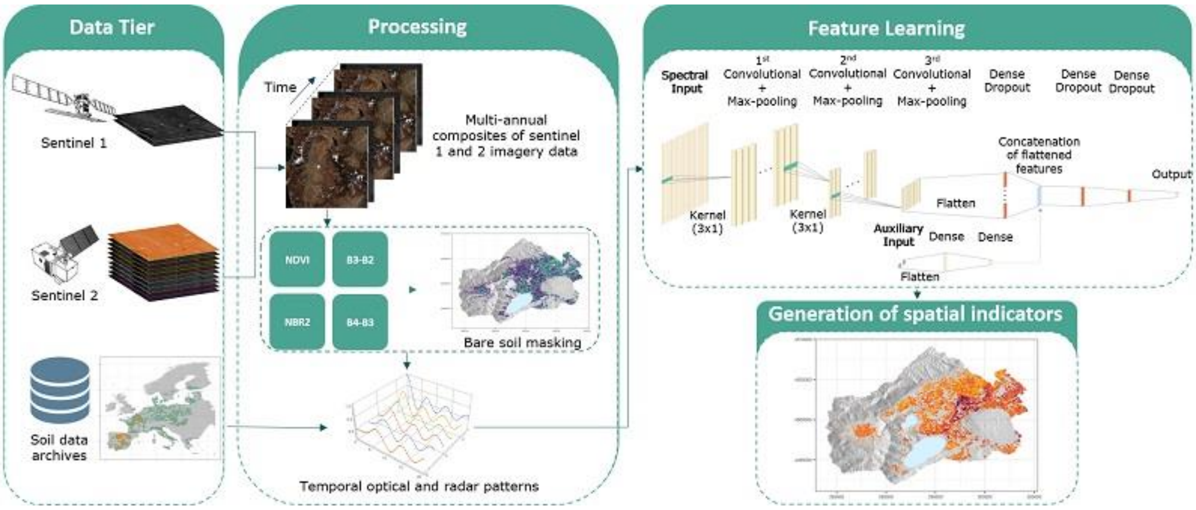

Figure 23: Development of soil maps from a combined use of Sentinel 1, Sentinel 2, and existing SSLs Tziolas et al., 2020a..

3.3 Knowledge tier development and processing

Understanding, the need for an accurate prediction of soil properties, we propose a methodological framework of cost-effective soil monitoring using various data from multiple data sources with the ultimate goal of producing **soil property maps**.

DIONE will finally generate spatially explicit indicators of the soil properties as **raster data (e.g GeoTIFF)**, that will be ingested into the DBMS described in *D4.1. Technical specifications of the in-situ soil scanning system (SSS), data processing system and farmer's geo-tagged photos framework* alongside with its relevant metadata, enabling external actors to access it and integrate it into their own applications. Using the interface of the DBMS for example, these soil maps will be integrated into the environmental performance tool of DIONE (Task 5.3 Development of the DIONE Environmental Performance Tool).

Among the above indicators, we must highlight the importance of SOC. SOC is recognized by the European Union, who considered the decline of SOC in European soils, as one of the main threats for soil degradation (CEC, 2006). In addition, understanding its spatial distribution is necessary to maintain food security and improve environmental quality in the context of global environmental change (Gholizadeh et al., 2018; Zeraatpisheh et al., 2019).

Moreover, we should also mention that DIONE deliverable *D.5.1 – Environmental metrics methodology for ML-system*, focused on addressing environment and climate priorities within the framework of CAP implementation and recognized that the **soil organic matter** is the primary and fundamental indicator of soil's quality in regard to land productivity and climate change. It is also an indicator that gets affected by mismanagement, thus reflecting the management practices in arable land.

4 Conclusions

The present deliverable provides a methodological framework describing which historical and open EO data will be exploited in DIONE, with what means and to what end, for the generation of EO-driven soil maps. This work provides a concrete insight of the different EO techniques that can be integrate, in order to help the EU's directives to meet their objectives. More specifically, a synergistic use of open access satellite data, SSLs and local spectral measurements is presented as an environmental performance tool in order to provide spatially explicit soil indicators.

The current deliverable presents a novel scheme via the exploitation of the spiked bottom-up approach, offering advantages in near real-time soil monitoring. The proposed methodology is achieved through the use of the portable and low-cost MEMS sensors, which provide in-situ measurements, the augment of the existing SSL, and the integration of the aforementioned data with the open-accessed satellite data. Last but not least, the big EO data management and storage is of great importance for the successful implementation of the DIONE's soil monitoring methodology and here we present also methods to achieve accurate and fast management of the Big satellite data.

References

Ambrosone, M., Matese, A., Di Gennaro, S.F., Gioli, B., Tudoroiu, M., Genesio, L., Miglietta, F., Baronti, S., Maienza, A., Ungaro, F., Toscano, P., 2020. Retrieving soil moisture in rainfed and irrigated fields using Sentinel-2 observations and a modified OPTRAM approach. Int. J. Appl. Earth Obs. Geoinf. https://doi.org/10.1016/j.jag.2020.102113

Ayehu, G., Tadesse, T., Gessesse, B., Yigrem, Y., M.melesse, A., 2020. Combined use of sentinel-1 sar and landsat sensors products for residual soil moisture retrieval over agricultural fields in the upper blue nile basin, ethiopia. Sensors (Switzerland). https://doi.org/10.3390/s20113282

Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma. https://doi.org/10.1016/j.geoderma.2015.07.006

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using Imaging Spectroscopy to study soil properties. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2008.09.019

Ben-Dor, E., Taylor, R.G., Hill, J., Demattê, J.A.M., Whiting, M.L., Chabrillat, S., Sommer, S., 2008. Imaging Spectrometry for Soil Applications. Adv. Agron. https://doi.org/10.1016/S0065-2113(07)00008-9

Brown, D.J., 2007. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. Geoderma. https://doi.org/10.1016/j.geoderma.2007.04.021

Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B., 2019a. Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects. Remote Sens. https://doi.org/10.3390/rs11182121

Castaldi, F., Chabrillat, S., Jones, A., Vreys, K., Bomans, B., van Wesemael, B., 2018. Soil organic carbon estimation in croplands by hyperspectral remote APEX data using the LUCAS topsoil database. Remote Sens. https://doi.org/10.3390/rs10020153

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019b. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS J. Photogramm. Remote Sens. https://doi.org/10.1016/j.isprsjprs.2018.11.026

Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2016.03.025

CEC, 2006. Thematic strategy for soil protection. Com.

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B., 2019. Imaging Spectroscopy for Soil Mapping and Monitoring. Surv. Geophys. https://doi.org/10.1007/s10712-019-09524-0

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 Sens. Environ. https://doi.org/10.1016/j.rse.2018.04.047

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L.B., Obersteiner, M., Van Der Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. Nat. Commun. https://doi.org/10.1038/ncomms11872

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.09.015

Hajj, M. El, Baghdadi, N., Zribi, M., Bazzi, H., 2017. Synergic use of Sentinel-1 and Sentinel-2 images for operational soil moisture mapping at high spatial resolution over agricultural areas. Remote Sens. https://doi.org/10.3390/rs9121292

Karyotis, K., Tziolas, N., Tsakiridis, N., Samarinas, N., Chatzimisios, P., Demattê, J.A.M., Zalidis, G., 2020. Digital soil mapping using Sentinel-2 imagery supported by ASTER thermal infrared bands. https://doi.org/10.1117/12.2570821

Keskin, H., Grunwald, S., 2018. Regression kriging as a workhorse in the digital soil mapper's toolbox. Geoderma. https://doi.org/10.1016/j.geoderma.2018.04.004

Kopačková, V., Ben-Dor, E., 2016. Normalizing reflectance from different spectrometers and protocols with an internal soil standard. Int. J. Remote Sens. https://doi.org/10.1080/01431161.2016.1148291

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science (80-.). https://doi.org/10.1126/science.1097396

Ledo, A., Hillier, J., Smith, P., Aguilera, E., Blagodatskiy, S., Brearley, F.Q., Datta, A., Diaz-Pines, E., Don, A., Dondini, M., Dunn, J., Feliciano, D.M., Liebig, M.A., Lang, R., Llorente, M., Zinn, Y.L., McNamara, N., Ogle, S., Qin, Z., Rovira, P., Rowe, R., Vicente-Vicente, J.L., Whitaker, J., Yue, Q., Zerihun, A., 2019. A global, empirical, harmonised dataset of soil organic carbon changes under perennial crops. Sci. Data. https://doi.org/10.1038/s41597-019-0062-1

Lesaignoux, A., Fabre, S., Briottet, X., 2013. Influence of soil moisture content on spectral reflectance of bare soils in the 0.4-14 µm domain. Int. J. Remote Sens. https://doi.org/10.1080/01431161.2012.743693

Lobell, D.B., Asner, G.P., 2002. Moisture Effects on Soil Reflectance. Soil Sci. Soc. Am. J. https://doi.org/10.2136/sssaj2002.0722

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma. https://doi.org/10.1016/j.geoderma.2015.07.017

Montanarella, L., Pennock, D.J., McKenzie, N., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Aulakh, M.S., Yagi, K., Hong, S.Y., Vijarnsorn, P., Zhang, G.L., Arrouays, D., Black, H., Krasilnikov, P., Sobocká, J., Alegre, J., Henriquez, C.R., Mendonça-Santos, M. de L., Taboada, M., Espinosa-Victoria, D., AlShankiti, A., AlaviPanah, S.K., Mustafa Elsheikh, E.A. El, Hempel, J., Arbestain, M.C., Nachtergaele, F., Vargas, R., 2016. World's soils are under threat. SOIL. https://doi.org/10.5194/soil-2-79-2016

Montanarella, L., Vargas, R., 2012. Global governance of soil resources as a necessary condition for sustainable development. Curr. Opin. Environ. Sustain. https://doi.org/10.1016/j.cosust.2012.06.007

Nocita, M., Stevens, A., Noon, C., Van Wesemael, B., 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. Geoderma. https://doi.org/10.1016/j.geoderma.2012.07.020

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthes, B., Dor, E.B., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., DemattÃa, J.A., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Wetterlind, J., 2015. Chapter Four - Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Adv. Agron.

Rodionov, A., Pätzold, S., Welp, G., Pallares, R.C., Damerow, L., Amelung, W., 2014. Sensing of Soil Organic Carbon Using Visible and Near-Infrared Spectroscopy at Variable Moisture and Surface Roughness. Soil Sci. Soc. Am. J. https://doi.org/10.2136/sssaj2013.07.0264

Romero, D.J., Ben-Dor, E., Demattê, J.A.M., Souza, A.B. e., Vicente, L.E., Tavares, T.R., Martello, M., Strabeli, T.F., da Silva Barros, P.P., Fiorio, P.R., Gallo, B.C., Sato, M.V., Eitelwein, M.T., 2018. Internal soil standard method for the Brazilian soil spectral library: Performance and proximate analysis. Geoderma. https://doi.org/10.1016/j.geoderma.2017.09.014

Roudier, P., Hedley, C.B., Lobsey, C.R., Viscarra Rossel, R.A., Leroux, C., 2017. Evaluation of two methods to eliminate the effect of water from soil vis–NIR spectra for predictions of organic carbon. Geoderma. https://doi.org/10.1016/j.geoderma.2017.02.014

Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Visible and Near Infrared Spectroscopy in Soil Science, Advances in Agronomy. https://doi.org/10.1016/S0065-2113(10)07005-7

Taghadosi, M.M., Hasanlou, M., Eftekhari, K., 2019. Retrieval of soil salinity from Sentinel-2 multispectral imagery. Eur. J. Remote Sens. https://doi.org/10.1080/22797254.2019.1571870

Tang, Y., Jones, E., Minasny, B., 2020. Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia. Geoderma Reg. https://doi.org/10.1016/j.geodrs.2019.e00240

Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environ. Monit. Assess. https://doi.org/10.1007/s10661-013-3109-3

Tsakiridis, N.L., Keramaris, K.D., Theocharis, J.B., Zalidis, G.C., 2020. Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. Geoderma. https://doi.org/10.1016/j.geoderma.2020.114208

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020a. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sens. https://doi.org/10.3390/RS12091389

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2019. A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved VIS-NIR-SWIR soil properties estimation. Geoderma. https://doi.org/10.1016/j.geoderma.2018.12.044

Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020b. An integrated methodology using open soil spectral libraries and Earth Observation data for soil organic carbon estimations in support of soil-related SDGs. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2020.111793

Vaudour, E., Gomez, C., Fouad, Y., Lagacherie, P., 2019. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2019.01.006

Ward, K.J., Chabrillat, S., Neumann, C., Foerster, S., 2019. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered LUCAS soil database. Geoderma. https://doi.org/10.1016/j.geoderma.2019.07.010

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma. https://doi.org/10.1016/j.geoderma.2018.09.006

Žížala, D., Minarík, R., Zádorová, T., 2019. Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions. Remote Sens. https://doi.org/10.3390/rs11242947