Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East, and Balkans and Developing Links with GEO related initiatives towards GEOSS

D4.6 – Pilot Activity Report Improved Food Security – Water Extremes Management

Contract Number	H2020 SC5-18b-2015, Project GA number: 690133	**Acronym**	GEO-CRADLE
Full title	Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East, and Balkans and Developing Links with GEO related initiatives towards GEOSS.		
Project URL	http://geocradle.eu		
EC Project Officers	Ms Gaëlle LE BOULER & Ms Malgorzata ROGIVAL		

Deliverable/Document	**Number**	D4.6	**Name**	Pilot Activity Report IFS-WEM		
Work package	**Number**	WP4	**Name**	Pilot towards regional challenges		
Date of delivery	**Contractual**		M26	**Actual**	24.07.2018	
Status	Final					
Nature	Report					
Distribution Type	Public					
Authoring Partner	i-BEC (inter-Balkan Environment Center)					
Prepared by	i-BEC, TAU, and CIMA					
Quality Assurance	Dr. Charalampos Kontoes, Lefteris Mamais, Alexia Tsouni					
Contact Person	e.g. Dr Charalampos (Haris) Kontoes	Project Coordinator				
	Metaxa & Vas. Pavlou Str. • 152 36 Penteli, Greece					
	Email	kontoes@noa.gr	**Phone**	+30-2108109113	**Fax**	+30-2106138343

Executive Summary

The objective of this deliverable is to present the pilot activities of Task 4.2, namely Improved Food Security and Water Extremes Management, and report on its outcomes. The GEO-CRADLE network encompasses the Balkans, the Middle East, and North Africa (including Cyprus and Turkey). This task involved the knowledge transfer actions of Earth Observation tools and practices from countries in the region that are more advanced and mature in the utilization of Earth Observation (EO) data towards countries that are lacking behind. This document entails a description of tools and practices which were used and have been distributed to all partners, having as an end goal their uptake and future utilization.

Having reflected upon the outputs of Work Packages (WP) 2 and 3, namely the existing gaps, needs, relevant EO capacities and existing maturity in the region of interest (RoI), a refined scope and objective was agreed upon and was presented in D4.2. This scope encompassed the steps which ought to have been taken in order to ensure a concrete and established uptake of EO services by the partners, with respect to the improvement of food security and water extremes management. The state-of-the-art technologies disseminated entail: a) on the hand the development of a regional Soil Spectral Library (SSL), granted as noted in WP 2 and 3 that the region was severely under-represented in other contemporary SSLs, which is necessary for the transformation of EO data to end-user data; and b) on the other hand the combined use of EO data and EO-derived maps with other sources (such as e.g. meteorological and topographical data) in an integrated web-GIS platform for monitoring of natural risks, projecting potential of risk in the future (forecasting), and evaluating the effect of prevention measures.

As far as the establishment of the regional SSL is concerned, a concentrated effort was realized to collect soil samples from the whole region. The effort was met largely with a success, and a total of 1754 distinct soil samples (records) comprise the regional SSL, which is made available through the regional datahub. It must be underscored that particular care was taken to ensure that the library conforms and adheres to other spectral libraries, assuring its future extension and compliance with other contemporary libraries. In other words, the library is expandable and can be used in conjunction with other similar libraries conforming to the same specifications. With regards to the spatial coverage, 1327 soil samples are from the Balkans, while 427 are from the Middle East and North Africa (including Cyprus and Turkey), with North Africa being the least represented region. To ultimately derive to this SSL, the methodology and techniques used to a) perform soil sampling (i.e. collect the soil samples), b) prepare the samples and chemically analyze them, c) perform the spectral measurements in a standardized, uniform, and error-free way, d) develop spectral models using machine learning algorithms, and finally e) applying the models to Copernicus EO data, were all disseminated to all project partners. This was achieved through a series of lectures, meetings, webinars, hands-on demonstrations, training days, and dissemination of software. These were all put to the test and culminated through their real-world application, i.e. the establishment of the regional SSL and production of EO-related products utilizing the SSL.

With respect to the management of water extremes, a web-GIS (Geo-Information System) platform was developed which enables the integration of static and dynamic geo-spatial data (including data derived from Earth Observation sources) as well as models for forecasting. This platform (named myDEWETRA) was populated with significant data for the RoI including past weather and climatic data, and forecast results of models assessing the floods and droughts risk. Moreover, products derived from Earth Observation techniques were integrated. Besides the general data of the region, a more concrete example was considered with more thorough data and results. A case study was considered, encompassing the Drin river basin in Northern Albania. A hydrological model was subsequently implemented in the region, whereby a comparison was made between modelled soil moisture and satellite soil moisture. The use of clay content maps from satellite data was investigated for the evaluation of the hydrogical model’s parameters. In addition, two soil maps generated from satellites of the Copernicus mission with the help of the regional SSL were furthermore integrated into the myDEWETRA platform and into the hydrological model of the region, showcasing how the outcomes of this pilot activity may be exploited in the future by other researchers.

The current pilot aims to improve the capacity building through a detailed program to strengthen EO project partners (and beyond) knowledge and capabilities. Moreover, the pilot outcomes are relevant to improve decisions of policy makers, concerning agricultural and natural processes in general. It has a strong potential to help EU agricultural policy meeting their objectives, concerning enhanced competitiveness of the agricultural sector and improved sustainability and effectiveness through reduced environmental impacts and utilization of natural resources in more sustainable and high efficient manners.

Project Information

This document is part of a research project funded under the **European Union Horizon 2020 Programme - Coordination and Support Action.**

Call Identifier: SC5-18b-2015, Integrating North African, Middle East and Balkan Earth Observation capacities in GEOSS.

Project GA number: 690133

Project Title: GEO-CRADLE - Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East, and Balkans and Developing Links with GEO related initiatives towards GEOSS.

Project Beneficiaries:

ID	Participant Organisation Name	Country	Logo
1	National Observatory of Athens (NOA) - **Coordinator**	Greece	
2	Interbalkan Environment Center (IBEC)	Greece	
3	Center for Environment and development for the Arab Region and Europe (CEDARE)	Egypt	
4	Research and Studies Telecommunications Centre (CERT)	Tunisia	
5	Tel Aviv University (TAU)	Israel	
6	Cyprus University of Technology (CUT)	Cyprus	
7	TUBITAK UZAY Space Technologies Research Institute (UZAY)	Turkey	
8	Space research and technology institute (SRTI)	Bulgaria	
9	National Institute of R&D for Optoelectronics (INOE)	Romania	
10	University of Ss Cyril and Methodius (USCM)	FYROM	
11	Institute for Nature Conservation in Albania (INCA)	Albania	
12	Institute of Physics Belgrade (IPB)	Serbia	
13	CIMA Research Foundation (CIMA)	Italy	
14	Academy of Athens (AOA)	Greece	
15	INOSENS (INS)	Serbia	
16	European Association of Remote Sensing Companies (EARSC)	EU	
17	EURISY	EU	
18	EuroGeoSurveys (EGS)	EU	
19	World Radiation Center (PMOD/WRC)*	Switzerland	

*Note: Switzerland is not requesting financial contribution from the EC

Table of Contents

Acronyms and Abbreviations

Acronym	Description
CIMA	CIMA research foundation (International Center for Environmental Monitoring)
DEM	Digital Elevation Map
EC	European Commission
EO	Earth Observation
ENET	Elastic Net
EU	European Union
FAO	Food and Agriculture Organization of the United Nations
GA	Grant Agreement
GA	Genetic Algorithm
GEOGLAM	Group on Earth Observations Global Agricultural Monitoring
GPR	Gaussian Process Regression
i-BEC	Inter-Balkan Environment Center
IP	Intellectual Property
kNN	k-Nearest Neighbors
LUCAS	Land Use / Land Cover Area Frame Survey
MENA	Middle-East and North Africa
MS	Milestones
NOA	National Observatory of Athens
PC	Project Coordinator
PLS	Partial Least Squares
QC	Quality Control
RD	Research Domain
RSG	Reference Soil Group
SSL	Soil Spectral Library

SVM	Support Vector Machine
TAU	Tel-Aviv University
TWI	Topographic Wetness Index
Vis-NIR	Visible – Near Infra-red
WP	Work Package
WPL	Work Package Leader
WRB	World Reference Base

List of Tables

List of Figures

1. Introduction

1.1. Regional importance

During the first months of GEO-CRADLE's implementation, an inventory of regional capacities and an end-user need analysis was conducted (WP2). This was achieved through a series of interviews carried out with end-user organizations, as well as by utilizing the existing knowledge base of the partners themselves. These two invaluable sources of information were subsequently analyzed and a number of **priorities in relation to the regional challenges** were identified (WP3, D3.3). This is complemented by a common support actions entailing a) the awareness raising of the value of the EO operational services, in order to promote the use of EO data by more actors, b) the improvement of human capacity to develop value-added services and support decision-driven management systems, by disseminating knowledge of how EO data can be exploited, c) the development of a regional data hub operating on the principles of free and open data, allowing unhindered and unconstrained access to EO data for everyone, d) and the establishment of a regional coordination mechanism.

With respect to the food security and water extremes management, the user-need and gap analysis highlighted that for countries with initial, intermediate, and advanced maturity degrees alike, the application of EO for the monitoring of agriculture and for managing water extremes were both **identified as key applications** by the end-users. Additionally, food insecurity is highly related to risks and uncertainty that might be linked, at several timescales, to several factors, from a natural disaster or a humanitarian problem, to ongoing climate change. In that context, considering that agriculture is very prominent in the RoI, and that many countries within it (such as Greece, Albania, Serbia, Tunisia, and Turkey) were subjected in the recent past to water extremes (i.e. floods) with the subsequent adverse impacts and damages to agriculture and crop production, it is easy to infer that open EO data and EO applications are of high importance for the end-users.

As far as the Balkans are concerned, agriculture is a very prominent theme across all the countries, being an important source of economic revenue. This underscores the importance of open EO data with respect to agriculture, which assist in a number of ways such as soil nutrient mapping, crop growth assessment, mapping of vegetation stress, and yield prediction. Likewise, the importance of EO systems, as accurate source of data, cannot be over-stated, in order to improve soil inventories, reporting and accounting activities within the region. Moreover, it becomes apparent why monitoring of water extremes is of pivotal importance; the adverse effects of e.g. a flood event are significant for the regional economies. In addition, the gap analysis showcased that both for the western Balkans (such as FYROM and Albania) the space capacities are limited, and the EO sector is mostly dominated by the public sector. Even for more intermediate and advanced countries (such as Greece and Romania) that exhibit developed EO capacities, structural gaps and lack of funding prohibit the private sector from acquiring high quality commercial EO data, and open Copernicus data are not sufficiently exploited. The availability of open EO data and release of tools that help utilize them, may pave the way to future investments in this region, leading to sustainable development of resources and activities, in addition to strengthened competitiveness and performance in the key sector of agriculture.

As far as the Middle East and North Africa regions are concerned, there are a number of setbacks in meeting food security. They are mostly due to the adverse effects of climate change (such as soil desertification), land and water scarcity, and discrepancy between high rates of population growth and low rates of food production growth. Despite the fact that some countries have attained an advanced maturity degree (such as Israel, and Turkey), having a large and mature public and private EO sector, other countries are lagging behind. This is attributed mostly to structural gaps (e.g. lack of data sharing, and lack of coordination) and capacity gaps (i.e. human resource limitations).

Taking into consideration the above, the IFS-WEM pilot comes to meet these regional needs. Its aims are to improve agricultural management and assist the decision process regarding agricultural practices by providing enhanced insight of the agricultural lands, while also implement open and efficient tools for water extremes management.

1.2. Overview of the pilot activities

Healthy soil is the foundation of the food system. The organisms living in the soil perform many vital functions including converting dead and decaying matter as well as minerals to plant nutrients, which are vital for the growth of plants and ergo food. It is the processes that occur within soil, most of which are driven by the life that is found within, which drive ecosystem and global functions and thus help maintain life above ground. It is a fundamental and non-renewable natural resource which people rely on for the production of food, fibre and energy and forms as a basis for all life on Earth. Soils provide the basis for agriculture, being the medium on which plants grow, in addition to providing the nutrients to the plants. It is estimated that over 95% of food is produced from the soil ecosystem. They are thus instrumental for securing the food supply, as without healthy soils it is impossible to sustain the increased demand for food. Hence, without robust soil ecosystems, the world's food web and in extension the humans would be in trouble, since soil produces healthy crops that in turn nourish people.

However, in the recent years, soils have become one of the world's most vulnerable resources in the face of climate change, land degradation, biodiversity loss and increased demand for food production [1]. Maintaining a healthy soil demands care and effort from farmers because farming is not benign. By definition, farming disturbs the natural soil processes including that of nutrient cycling - the release and uptake of nutrients. Farming systems mine the soil for nutrients and reduce the soil organic matter levels through repetitive harvesting of crops and inadequate efforts to replenish nutrients and restore soil quality. In addition, there exist a number of unprecedented pressures on soil from degradation and urbanisation, which are threatening the above-mentioned functions, and impend agro-ecological balances and food security. Thus, a profound change is necessary in order to move towards more sustainable ways of production so that they can provide in a sustainable way food for the additional 2 billion people expected by 2050.

In that context and since the launch of Global Soil partnership (Food and Agriculture Organization, 2012), a wide range of activities with the main aim to strengthen and support soil ecosystem and its functions by improving the use of EO tools for site specific projections or forecasting has been implemented. These activities have made great strides towards developing methodologies and management practices, fostering ecosystem creation and raising awareness amongst the main stakeholders (e.g. the 5 pillars of action – GSP). Along these lines, the G20 Ministerial Declaration launched the Group on Earth Observations Global Agricultural Monitoring (GEOGLAM) initiative in June 2011, with the aim of strengthening GLobal Agricultural Monitoring by improving the use of remote sensing tools for crop production projections. More recently, the role of soil ecosystem has been recognized as central to sustainable development since it contributes to a wide range of critical functions. In this line, the importance of monitoring land use and land use change for monitoring GHG emissions from agricultural activities has been recognized and prioritized in the context of the 2030 EU climate and energy framework (Decision No 529/2013/EU). Finally, the importance of monitoring a range of essential variables for the implementation of Sustainable Development Goals (SDGs) has been recognised and prioritized.

It becomes apparent that it is vital to develop a methodology for the consistent monitoring of soil health at continental scale. The traditional approach of soil sampling is inefficient to cover large areas timely and regularly. Moreover, it requires complex analytical methods which are not consistent. A more efficient solution that provides a fast and low-cost method with sufficient accuracy is the application of soil spectroscopy.

Visible and near-infrared (Vis-NIR) diffuse reflectance spectroscopy has proved to be a fast, cost- effective, environmental-friendly, non-destructive, reproducible, and repeatable analytical technique [2]. This technique, although used more often in the laboratory, can be applied both in-situ and using air- and space-borne sensors [3]. The driving idea of this approach, is that the soil physical and chemical properties can be inferred by statistical analysis of the reflectance spectrum of the soil. Considering that soil is a vastly complex material consisting of organic and inorganic mineral matter, and is highly spatially variable, this relationship becomes more difficult to model.

Soil spectral libraries 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. Recently, large soil spectral libraries (SSLs) have been developed at very large scales (e.g. the Australian SSL [4] and the Chinese SSL [5]). In Europe, the LUCAS soil spectral library [6] is comprised of 19,036 soil samples across 23 EU countries, a task which was undertaken within the frame of the Land Use / Land Cover Area Frame Survey (LUCAS). Recently, a global SSL was developed, which gathered all existing and compatible SSLs [7]. These large SSLs can be used not only to infer the soil properties on the continental scale, but can assist in the development of more local (i.e. regional or country-level) models; new data collected (provided that they adhere to the same standards) can be used in conjunction with existing SSLs to build models that are more robust and accurate (e.g. by using the so called spiking strategy [8]).However, in all of these SSLs, the region of interest (encompassing the Balkans, North Africa, Middle East, Cyprus, and Turkey) is severely underrepresented and inhomogeneous; spanning from the absolute absence of samples, to countries having only a handful of samples.

In these recent efforts for the development of SSLs, i-BEC and TAU have in the past years established local SSLs at their country level, adhering to the global standards. TAU has a number of peer-reviewed publications on the development of SSLs and their application using EO data (e.g. [9]–[13]) and has an advanced maturity level in the usage of SSL, as well as in the dissemination of the know-how needed to establish one. On the other hand, i-BEC has recently established a large local SSL in Greece [14], has significant expertise in applied machine learning [15] and has also applied soil spectroscopy to measure soil contamination [16]. Furthermore, i-BEC, being the regional coordinator in the Balkans has significant expertise on know-how of how to best apply the developed SSL and derive products which may be of significant value for end-users and stakeholders alike.

The overarching objective of the Improved Food Security – Water Extremes Management pilot is to develop a regional SSL in the RoI, as well as disseminate all the knowledge and tools necessary to the partners so that they may use it effectively to derive soil thematic maps using laboratory, in situ, and satellite data, thus developing their capacities. To this end, partners were taught how to perform effective soil sampling in the region, and where trained on how to perform spectral measurements following a robust standardization and measurement protocol. In addition, the process of building robust and accurate machine learning models was shared with the partners. Finally, a methodology on how to apply this SSL on a local basis was showcased in a feasibility study in the Drin River basin, located in Northern Albania.

The rest of the deliverable is organized as follows: Section 2 describes the implementation of the pilot activity, covering all technical aspects. In particular, Section 2.1 describes the steps taken to establish the regional SSL and also contains an in-depth description of its components both on a per-country level as well as a whole. In Section 2.2 the myDEWETRA platform is presented, showcasing its ability to depict all information stored in a user-friendly way. The machine learning algorithms and techniques applied to develop soil spectroscopy models may be found in Section 2.3. The feasibility study's methodology and results are described in Section 2.5. Additionally, in Section 2.6 the training sessions and webinars conducted during the course of this activity are detailed. A brief overview of the outcomes may be found in Section 2.7. The pilot implementation is analyzed in Section 3, with its positive outcomes as well as potential shortcomings outlined in Sections 3.1 and 3.2 respectively. The conclusions of this work and future considerations are described in Section 3.3. In Appendix A, the protocol followed during soil sampling from the partners may be found, which can assist future researchers to sample their region and contribute to the SSL.

2. Project implementation report

2.1. Development of regional soil spectral library

This section describes the process of building the regional soil spectral library of the region. Despite the fact that there have been quite a few efforts recently to create large soil spectral libraries (e.g. the LUCAS SSL in the European Union [6] and a global SSL [7]) the Balkans and the MENA region are currently under-represented. During the implementation of this task, a consolidated effort was realized to create a regional SSL with high standards, compatible with other efforts worldwide, and extensible, focusing on this underrepresented region. In the subsection that follows the methodology used to populate this SSL, and the samples comprising it are described.

In the second project meeting, conducted in Limassol, Cyprus in November of 2016, the partners participating in this task were handed a protocol to follow in order to collect and measure the soil samples. This protocol is presented in Appendix A. The instructions presented the partners with two options: a) either find samples already collected and measured by e.g. their local university or any other laboratory who would agree to hand them out to them, b) or collect the soil samples themselves following a detailed protocol. The first option is useful because sampling campaigns are costly, and soil archives stored by universities, research centres, agriculture associations, and government agencies could have provided an economical opportunity to enlarge the regional spectral library; even if the samples had been acquired decades ago, they still contain spectral information which could be utilized to improve the representability of the spectroscopic calibration models.

As far as new samples are concerned, it was decided to follow stratified sampling. In stratified sampling, the population is partitioned into non-overlapping groups, called strata and a sample is selected by some design within each stratum. The principal reasons for using stratified random sampling rather than simple random sampling include:

- Stratification may produce a smaller error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are very homogeneous.
- The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.
- Estimates of population parameters may be desired for subgroups of the population. These subgroups should then be identified.
- If measurements within strata have lower standard deviation, stratification gives smaller error in estimation.
- For many applications, measurements become more manageable and/or cheaper when the population is grouped into strata.
- It is often desirable to have estimates of population parameters for groups within the population.

2.1.1. **Physical and Chemical analytical measurements**

2.1.1.1. *Soil texture*

Soil texture is an important characteristic of mineral soils. It affects water holding capacity, drainage properties, root development and more. Because texture has a great effect on water movement through a soil, it also directly affects the rate at which pesticides and nutrients move through the soil. Soil particles, such as clay, can also bind up certain elements and nutrients, directly affecting the soil's ability to retain the nutrients. Fine textured soils generally have a higher capacity for water retention, whereas sandy soils contain large pore spaces that allow leaching. Soil textures are classified by the fractions of each soil separate (sand, silt, and clay) present in a soil. Classifications are typically named for the primary constituent particle size or a combination of the most abundant particles sizes, e.g. "sandy clay" or "silty clay". A fourth term, loam, is used to describe equal properties of sand, silt, and clay in a soil sample, and lends to the naming of even more classifications, e.g. "clay loam" or "silt loam"; determining the soil texture is often aided with the use of a soil texture triangle (Figure 2-1).

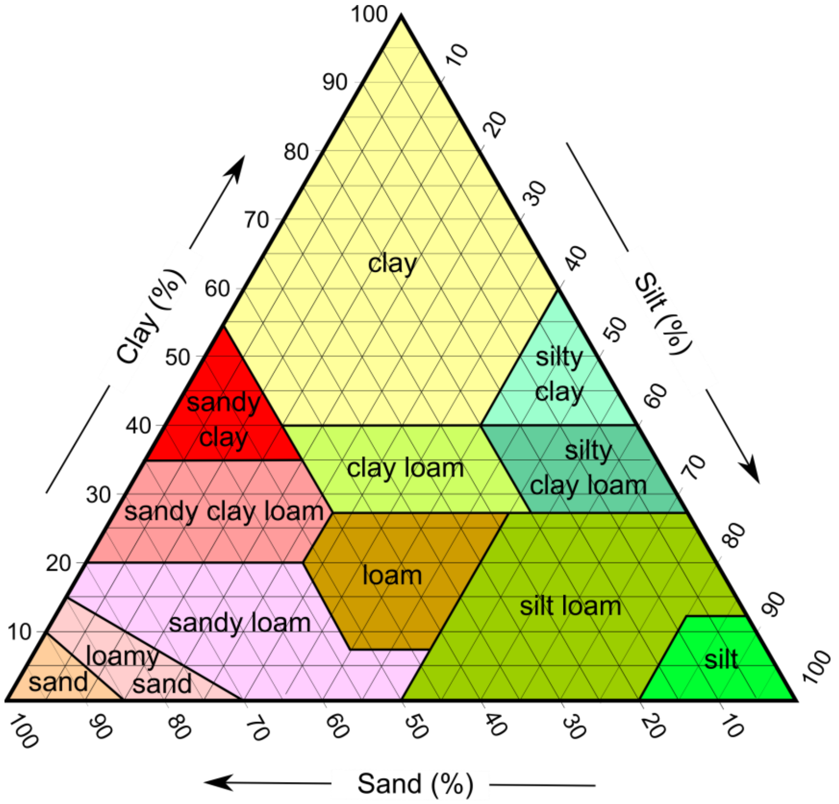

Figure 2-1. Soil characterization according to its texture – soil texture classes (WRB and USDA)

The mineral components of the soil are:

- **Coarse fragments:** Greater than 2 mm and include coarse quartz, rock fragments and cemented material. This is commonly called the 'gravel fraction'.
- **Sand:** Comprise quartz and resistant primary minerals such as mica. Sand particles are between 2 mm and 20 microns in size (Note: there are 1000 microns in 1 mm).
- **Silt:** Silts are typically composed of quartz and small mineral particles such as feldspars and mica and are between 2 and 20 microns in diameter.
- **Clay:** Clays are made up of secondary clay minerals and oxides/oxyhydroxides of iron and aluminium and are less than 2 microns in diameter.

2.1.1.2. *Soil Organic matter*

Soil organic matter (SOM) is the organic matter component of soil. It is the product of on-site biological decomposition and affects the chemical and physical properties of the soil and its overall health. It consists of plant and animal residues at various stages of decomposition, cells and tissues of soil organisms, and substances synthesized by soil organisms [17]. The composition and breakdown rate of SOM affect: the soil structure and porosity; the water infiltration rate and moisture holding capacity of soils; the diversity and biological activity of soil organisms; and plant nutrient availability. As such, SOM exerts numerous positive effects on soil physical and chemical properties, as well as the soil's capacity to provide regulatory ecosystem services. Particularly, the presence of SOM is regarded as being critical for soil function and soil quality. SOM is composed of roughly 58% carbon [18] which corresponds to SOC and is influenced by microbial activity, accessibility of organic residues to microbes, various site conditions and management practices. Managing SOC through sustainable agricultural and land use practices has become a widely acknowledged strategy to restore healthy soil properties to combat land degradation and desertification, and enhance the resilience of agro-ecosystems to environmental shocks. Thus, the importance of SOM in agriculture is paramount.

Two common methods for analysis of soil organic matter are the Walkley-Black acid digestion method and the weight loss on ignition method. The Walkley-Black method, used since the 1930's, uses chromic acid to measure the oxidizable organic carbon in a soil and is more accurate and more precise on soils with less than 2.0% organic matter. On soils very high in organic matter, the Walkley-Black method may result in low test results, due to the incomplete oxidation of the organic carbon in the sample. The Loss on Ignition method is better suited to soils with greater than 6.0% organic matter, however it should be used carefully due to changes in weight measurement caused by absorption of water molecules and changes in mineralogy occurring during the combustion process.

2.1.1.3. *Calcium Carbonate ($CaCO_3$)*

Calcium carbonate is a common substance found in rocks in all parts of the world, and is the main component of shells of marine organisms, snails, coal balls, pearls, and eggshells. Calcium carbonate is the active ingredient in agricultural lime, and is created when calcium ions in hard water react with carbonate ions creating limescale. Agricultural lime, powdered chalk or limestone, is used as a cheap method for neutralizing acidic soil (see Subsection 2.1.1.4), making it suitable for planting.

$CaCO_3$ is measured in soils usually following one of the two below listed methodologies: 1) It is either calculated from the weight of CO_2 lost after treating a sample with excess hydrochloric acid, or 2) from the pH of a suspension of the soil in dilute acetic acid. If a laboratory measurement protocol is closely followed, these methods can be quite accurate and produce reliable results [19].

2.1.1.4. *pH*

Soil pH is a measure of the acidity or basicity of a soil, and measures the concentration of hydrogen ions in the soil solution. The lower the pH of the soil, the greater its acidity. Soil pH is considered a master variable in soils as it affects many chemical processes. Although not a nutrient itself, it relates to plant nutrition. It specifically affects plant nutrient availability by controlling the chemical forms of the different nutrients and influencing the chemical reactions they undergo. Plant growth and most soil processes, including nutrient availability and microbial activity, are favoured by a soil pH range of 5.5–8. Acid soil, particularly in the subsurface, will also restrict root access to water and nutrients.

Soil acidification is a natural process accelerated by agriculture. Soil acidifies because the concentration of hydrogen ions in the soil increases. The main cause of soil acidification is inefficient use of nitrogen, followed by the export of alkalinity in produce. Acid rain and excessive plant growth are also significant sources of acidity. In contrast, the accumulation of alkalinity in a soil (as Na, K, Ca and Mg bicarbonates and carbonates) occurs when there is insufficient water flowing through the soils to leach soluble salts. This may be due to arid conditions, or poor internal soil drainage.

Thus, considering that soil pH effects both nutrient availability and microbial activity, the monitoring of it is significant to ensure sustainable agriculture and enhance the food security of the soils.

2.1.1.5. *Nitrates*

Soil nitrate (NO_3^-) is a form of inorganic nitrogen (N) naturally occurring in soils. Nitrates are used as fertilizers in agriculture, because of their high solubility and biodegradability. Nitrates are a highly mobile nutrient in the soil. The polyatomic ion of nitrate is negatively charged and hence cannot be held on to by negatively charged soil (clay and silt) particles. As a consequence, it is vulnerable to being leached down the soil profile. Nitrogen is a necessity in crop growth as shortage of this nutrient lead to poor crop growth, reduced yield and economic loss to the farmer. By knowing the amount of soil nitrogen in the soil, the farmers can be more accurate in their application of additional nitrogen. Knowledge of soil nitrogen level also helps reduce the amount of environmental pollution that occurs from runoff of excess nitrogen in the soil.

2.1.1.6. *Electrical Conductivity*

Electrical conductivity (EC) is a measurement of the dissolved material in an aqueous solution, which relates to the ability of the material to conduct electrical current through it. EC is measured in units called Siemens per unit area (mS/cm), and the higher the dissolved material in a water or soil sample, the higher the EC will be in that material.

The electrical conductivity of soils varies depending on the amount of moisture held by soil particles. As a general rule, sandy soils have a low conductivity, silts have a medium conductivity, and clays have a high conductivity. Consequently, EC correlates strongly to soil particle size and texture. In addition to EC values separating variations in soil texture, EC has been shown to relate closely to other soil properties used to determine a field's productivity. These are the water-holding capacity / drainage, the cation exchange capacity (CEC) which is related to SOM, soil porosity, and soil salinity.

2.1.1.7. *Cation Exchange Capacity (CEC)*

The cation-exchange capacity (CEC) is a measure of how many cations (positively charged ions) can be retained on soil particle surfaces. The clay mineral and organic matter components of soil have negatively charged sites on their surfaces which adsorb and hold positively charged ions (cations) by electrostatic force. This electrical charge is critical to the supply of nutrients to plants because many nutrients exist as cations (e.g. magnesium, potassium and calcium). In general terms, soils with large quantities of negative charge are more fertile because they retain more cations. It is a significant soil property which influences the stability of the soil structure, nutrient availability, soil pH, and the soil's reaction to fertilisers. This is an inherent soil characteristic and is difficult to alter significantly.

2.1.2. Soil class according to WRB

The World reference base (WRB) for soil resources is an international soil classification system for naming soils [20]. The classification of soils is based on soil properties defined in terms of diagnostic horizons, diagnostic properties and diagnostic materials, which to the greatest extent possible should be measurable and observable in the field. Climate parameters are not taken into account for the classification of a soil sample; it is therefore not subordinated to the availability of climate data, and thus the class attributed to a sample will not become obsolete due to global or local climate change.

The WRB classification system comprises two levels of categorical detail:

- the First Level consisting of 32 Reference Soil Groups (RSGs);
- the Second Level, which is the name of the RSG combined with a set of principal and supplementary qualifiers.

Table 2-1. First Level of the 32 reference soil groups of WRB [20]

Description	Class	Abbreviation
1. Soils with thick organic layers:	Histosols	HS
2. Soils with strong human influence –		
With long and intensive agricultural use:	Anthrosols	AT
Containing significant amounts of artefacts:	Technosols	TC
3. Soils with limitations to root growth –		
Permafrost-affected:	Cryosols	CR
Thin or with many coarse fragments:	Leptosols	LP
With a high content of exchangeable Na:	Solonetz	SN
Alternating wet-dry conditions, shrink-swell clays:	Vertisols	VR
High concentration of soluble salts:	Solonchaks	SC
4. Soils distinguished by Fe/Al chemistry –		
Groundwater-affected, underwater and in tidal areas:	Gleysols	GL
Allophanes or Al-humus complexes:	Andosols	AN
Subsoil accumulation of humus and/or oxides:	Podzols	PZ
Accumulation and redistribution of Fe:	Plinthosols	PT
Low-activity clay, P fixation, many Fe oxides, strongly structured:	Nitisols	NT
Dominance of kaolinite and oxides:	Ferralsols	FR
Stagnating water, abrupt textural difference:	Planosols	PL
Stagnating water, structural difference and/or moderate textural difference:	Stagnosols	ST
5. Pronounced accumulation of organic matter in the mineral topsoil –		
Very dark topsoil, secondary carbonates:	Chernozems	CH
Dark topsoil, secondary carbonates:	Kastanozems	KS
Dark topsoil, no secondary carbonates (unless very deep), high base		

Description	Class	Abbreviation
status:	Phaeozems	PH
Dark topsoil, low base status:	Umbrisols	UM
6. Accumulation of moderately soluble salts or non-saline substances –		
Accumulation of, and cementation by, secondary silica:	Durisols	DU
Accumulation of secondary gypsum:	Gypsisols	GY
Accumulation of secondary carbonates:	Calcisols	CL
7. Soils with clay-enriched subsoil –		
Interfingering of coarser-textured, lighter coloured material into a finer-textured, stronger coloured layer:	Retisols	RT
Low-activity clays, low base status:	Acrisols	AC
Low-activity clays, high base status:	Lixisols	LX
High-activity clays, low base status:	Alisols	AL
High-activity clays, high base status:	Luvisols	LV
8. Soils with little or no profile differentiation –		
Moderately developed:	Cambisols	CM
Sandy:	Arenosols AR	AR
Stratified fluviatile, marine and lacustrine sediments:	Fluvisols	FL
No significant profile development:	Regosols	RG

2.1.3. Spectral measurements

The recording of the laboratory spectral signatures in the vis-NIR (350-2500nm) from the soil samples comprising the GEO-CRADLE SSL was made at two different locations using two different instruments, one in Thessaloniki, Greece by i-BEC and one in Tel-Aviv, Israel by TAU. TAU operates an ASD Fieldspec Pro spectrometer while i-BEC operates a PSR+3500 spectrometer from Spectral Evolution (Figure 2-2). While they both cover the same range with approximately the same spectral resolution, they are different instruments and use different sensors. As such, and to ensure the SSL's ability to be further extendable and compatible with other regional SSL, it was necessary to follow a precise measurement and standardization protocol. The full details of the protocol may be found in [9], [10]. The protocol is time and wavelength independent to account for the ambient conditions and can correct for systematic and non-systematic errors.

Figure 2-2. The PSR+ spectrometer by Spectral Evolution used by i-BEC

To begin with, the protocol requires some initial processing of the soil samples. Considering that the particle size of the soil sample, as well as its water content significantly impacts the recorded spectrum. To eliminate these unwanted artefacts and effects on the recorded spectra, each physical soil sample was ground to approximately 2mm. The sample was subsequently passed through a sieve (<2mm) to ensure that no large particles or plant residues were left in the sample. This process was made carefully, and it was ensured that each sample after this operation was at least 100g. The sample was then air dried for at least one week prior to any spectral measurements, and the room temperature and relative humidity of the room were recorded.

i-BEC used a dark box environment for the measurement of soil reflectance with a bare fibre (and no contact with the soil) using a specific set viewing and illumination geometry (Figure 2-3). It consists of two 12 V and 35 W tungsten halogen lamps that illuminate the sample at 45°. The bare fibre is set up to measure the soil sample from the nadir position, covering a field of view (FOV) of 100 mm diameter of the sample's surface. The measurements are done under a constant distance of the sample to the fore optic (10 cm). The instrument is operated using a PDA. TAU performed the spectral measurements using a contact probe. Each soil sample was measured 3 times and the average spectrum was calculated for further analysis.

Figure 2-3. Inside view of the dark box

Figure 2-4. The PDA used to connect to the spectrometer

Figure 2-5. The setup used by TAU

To standardize the soil spectra, two internal soil standard (ISS) was used. The soil samples used as ISS were acquired from the Lucky and Wylie Bay in Western Australia (Figure 2-6). These samples were found to be stable in space and time and to hold a stable soil structure and spectral response common to soils. They are quite homogeneous and almost monomineralic (quartz). Both i-BEC and TAU are in possession of these samples, and it is the measurements of these ISS that can standardize the soil spectra.

Figure 2-6. Location of Lucky and Wylie Bay from where the ISS were collected

Figure 2-7. The Lucky Bay and Wylie Bay samples placed inside a petri dish

The steps taken to perform a spectral recording now follow.

1. First, the spectrometer and the lamps are warmed up for a minimum of 60 min to ensure that any transient effects are gone, and the instrument and illumination is stable
2. The instrument is set up to perform 20 internal averages for each recorded spectrum, and the integration time is set to auto
3. The reflectance of a white reference (WR) is acquired (Figure 2-8)- this is repeated until the WR measurement is stable
4. A petri dish is used to place the first ISS and its surface is gently smoothed against a dry flat disc
5. The ISS is placed into the dark box, directly beneath the fibre optic
6. The dark box is closed, ensuring no ambient light may enter the chamber
7. The reflectance of the ISS is acquired 5 times – if the coefficient of variation exceeds 5% in any of the wavelengths, the measurements are repeated, otherwise the mean spectrum is retained
8. Steps 4-7 are repeated for the other ISS
9. Steps 4-7 are repeated for the soil sample

Figure 2-8. The white reference plate used closed (a) and open (b)

Having acquired the spectra for each soil sample, the spectra are standardized using the ISS. First a correction factor is calculated for every single wavelength. Since two different ISS are employed, the final correction factor is the average of the two corrections factors (one for each reference soil). The correction factor is calculated as:

$$CF(\lambda) = 1 - \frac{S\rho(\lambda) - SBM\rho(\lambda)}{S\rho(\lambda)}$$

Where:

$CF(\lambda)$	The correction factor per wavelength
$S\rho(\lambda)$	The reflectance of the reference
$SBM\rho(\lambda)$	The reflectance of the soil benchmark

Then, after the two correction factors are calculated a

nd averaged, the standardized reflectance spectrum is calculated as:

$$Rc(\lambda) = Ro(\lambda) \times CF_{avg}(\lambda)$$

Where:

$Rc(\lambda)$	The corrected reflectance
$Ro(\lambda)$	The initial reflectance
$CF_{avg}(\lambda)$	The average correction factor

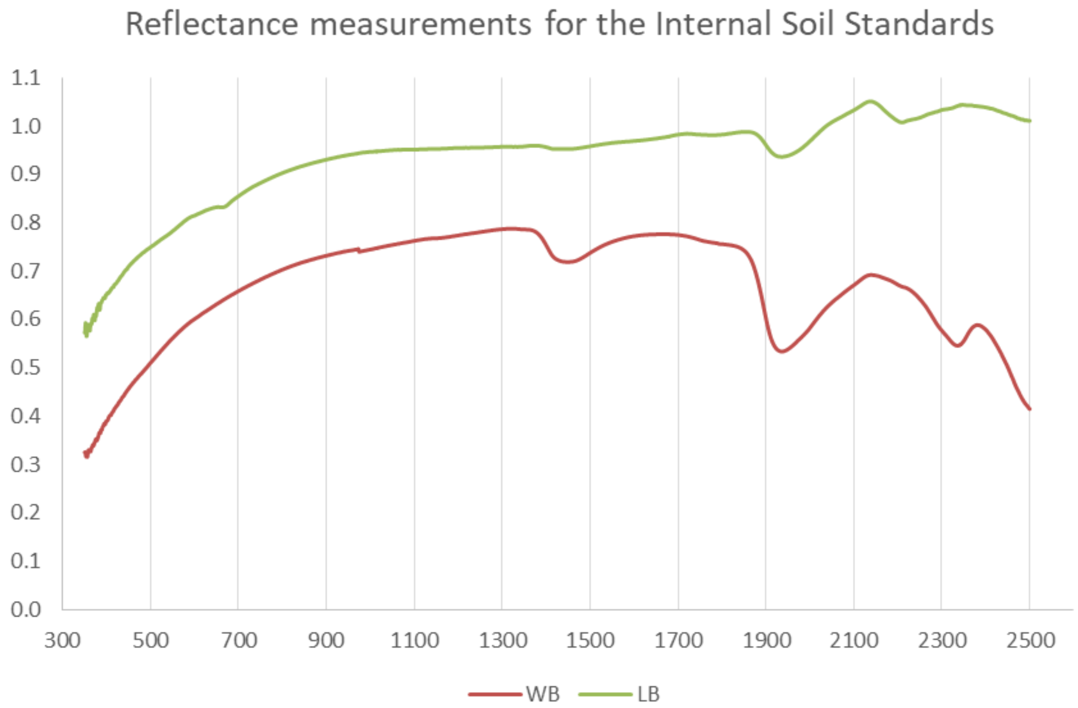

Figure 2-9. Soil benchmark values for each internal soil standard used

2.1.4. **Samples origin**

2.1.4.1. *Albania*

In Albania, INCA was unable to acquire soil samples from other resources (such as other institutions). Furthermore, Albania was chosen to host one of the feasibility studies undertaken in this pilot activity, namely the case of the Drin River basin (see Section 2.5). As such, it was important to obtain new samples from this region, in order to be able to showcase their use in conjunction with the Copernicus EO data. To this end, INCA followed the protocol given to all members by i-BEC and TAU with respect to the process that should be followed to collect the soil samples.

More concretely, first a soil pedological map of Albania was studied. As this provided a static snapshot of the region, a further recently acquired satellite image (18/05/2017) was examined to identify the current soil status of the area. The land cover type of the river basin was examined, in order to find out the exact sites with bare soil land. Moreover, the spatial variability of some remote sensing soil indices was taken into account. Finally, the accessibility of the sampling sites was assessed. All the above criteria dictated, through optimization, the preferred soil sampling locations. A number of 150 potential soil sampling locations were identified, 119 in the Shkodra region, 16 in Kukes, and 15 in Pogradec. This was estimated to take about 7 days of field work to conclude the collection of the samples. The location of the final sampling points is depicted in Figure 2-10.

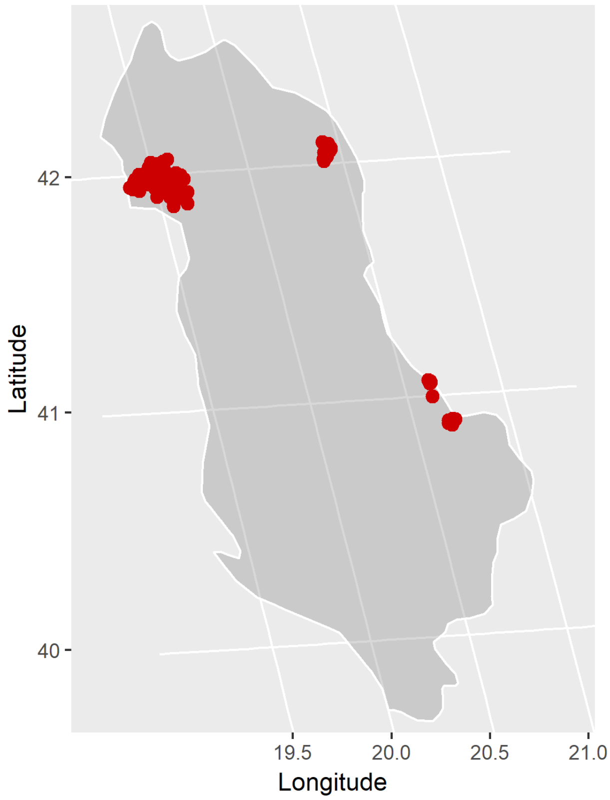

Figure 2-10. Location of the soil samples of Albania

The results of the physical and chemical analysis (including the in-situ measurements of soil moisture) are depicted in Table 2-2 and Figure 2-11 in the form of boxplots. It is noteworthy to highlight the fact that there is a highly outlier soil sample in terms of organic matter. In fact, the value is around 36, which following the FAO definition of organic soil materials [20] classifies the sample as belonging to an organic soil. Organic soils (i.e. soils with more than 20% of organic carbon content) are formed mainly in waterlogged conditions, where the anaerobic soil conditions support the preservation of vegetation residues and litter and their transformation to peat. As such, it is expected that their respective spectral signatures are different than the ones belonging to mineral samples (i.e. soils with less than 20% of organic carbon content).

The corresponding distribution of the patterns to the soil classes according to the WRB soil classification may be found in Figure 2-13. The vast majority of the samples are Fluvisols and Luvisols. The former class is characterized by little profile differentiation and contain fluviatile, marine, and lacustrine sediments. This is more consistent with the Shkodra region. The former class is characterized by high activity clays and soils with clay enriched subsoil, more consistent with the other regions.

The spectral signatures are given in Figure 2-14. Due to the fact that most samples may be classified as silty (Figure 2-12), the spectral signatures are somewhat flat, particularly in the vis-NIR region. The SWIR region, dominated by the clay minerals, is more prominent.

Table 2-2. Major statistical moments of the measured soil properties of Albania

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.71	2.88	2.29	35.78	3.49	7.93	70.86	107
Sand (%)	0.40	22.83	22.10	80.50	13.92	0.83	1.68	107
Silt (%)	6.90	43.08	43.70	67.40	11.92	-0.32	-0.34	107
Clay (%)	12.60	34.08	31.80	72.60	15.18	0.66	-0.61	107
Moisture (%)	1.34	41.78	6.53	12.28	11.55	0.96	3.22	107

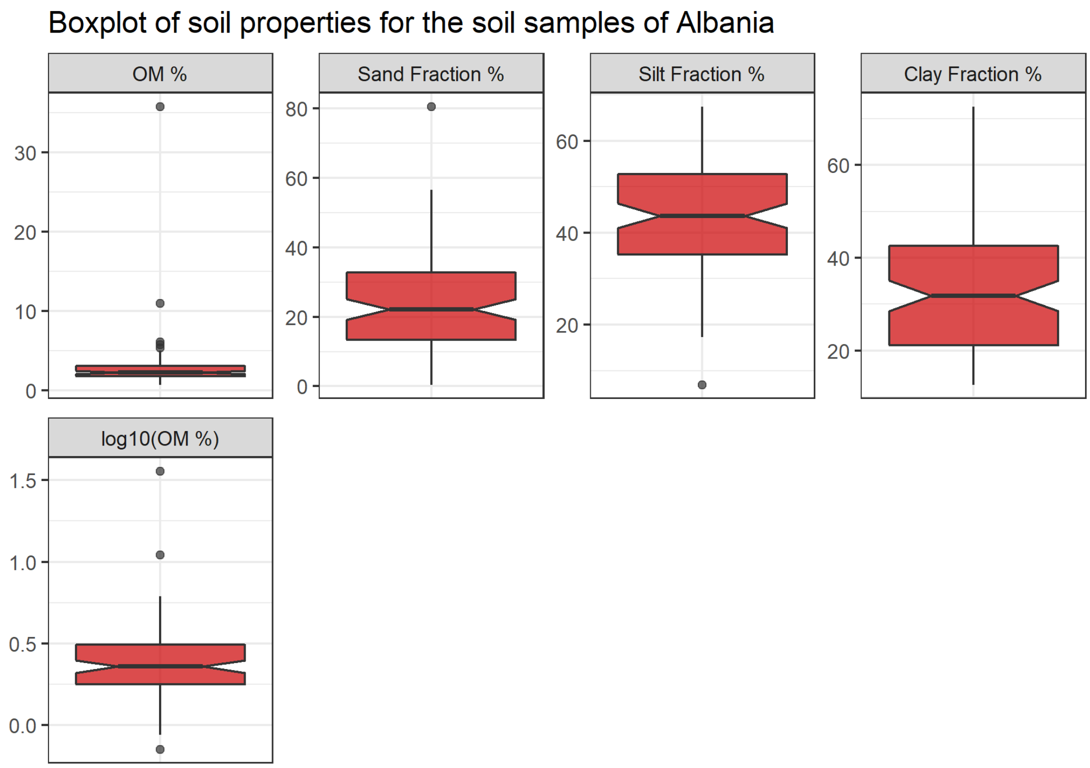

Figure 2-11. Boxplots of the measured soil properties for Albania

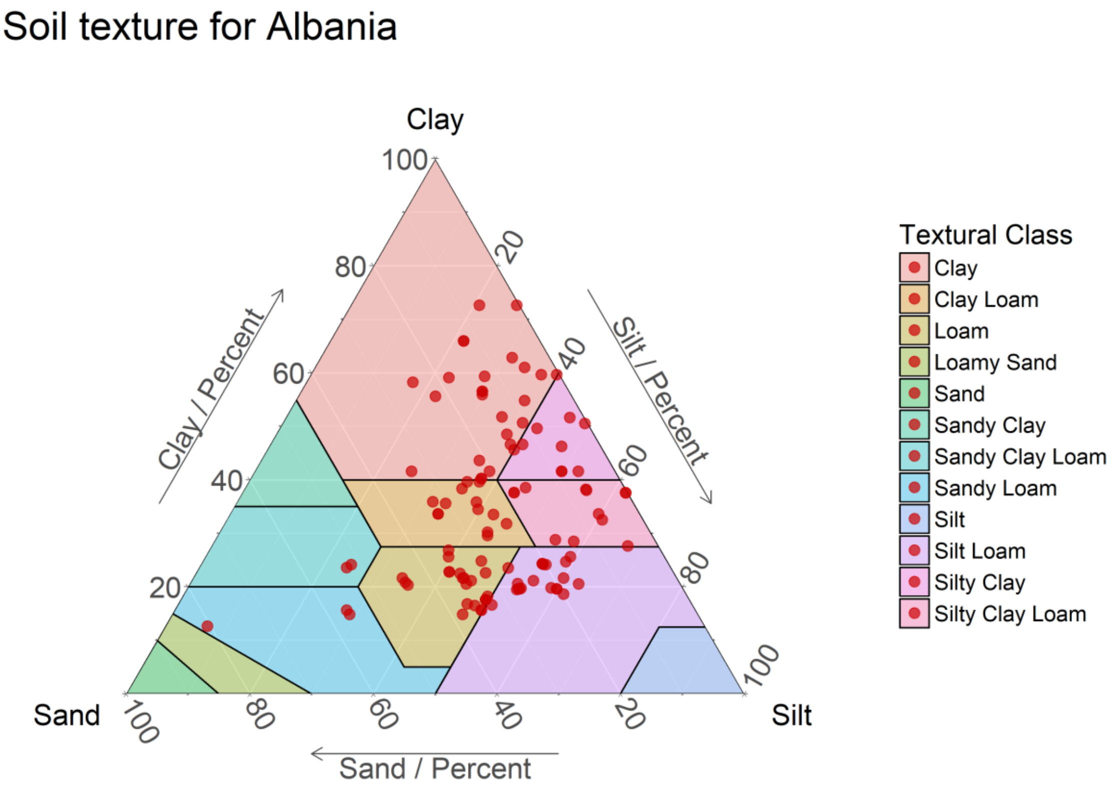

Figure 2-12. Distribution of soil texture class across all soil samples for Albania

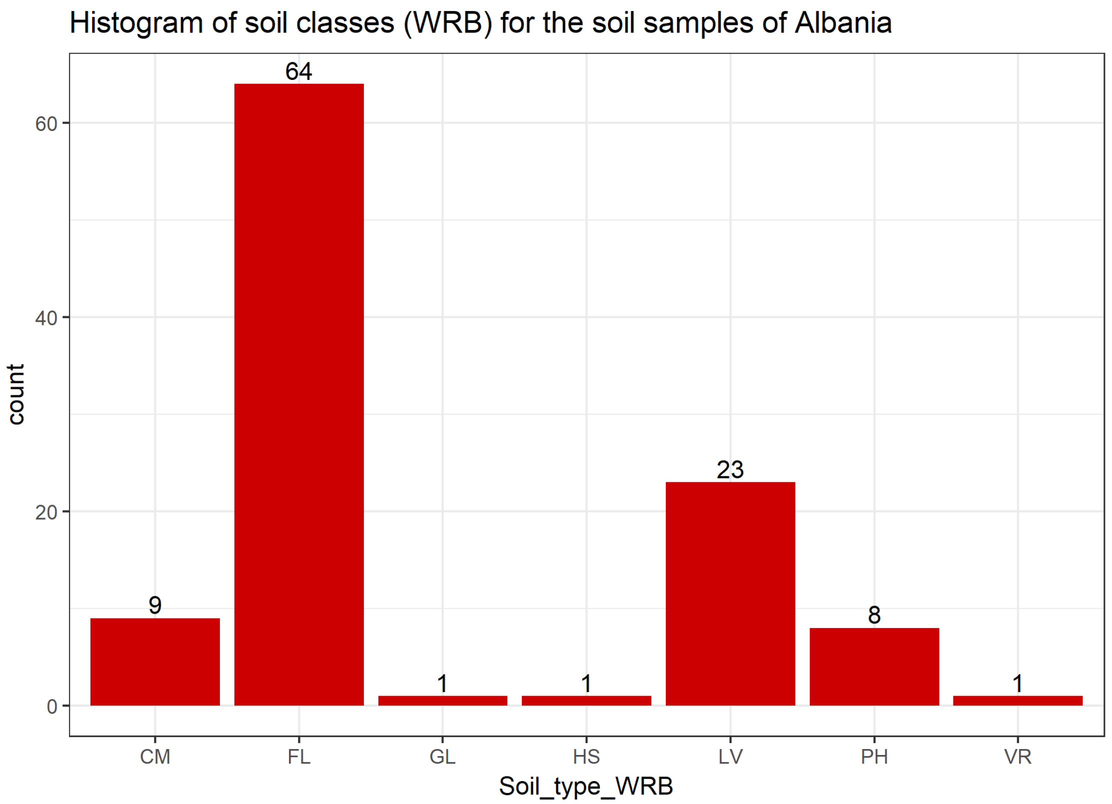

Figure 2-13. Histogram of soil classes of the soil samples of Albania

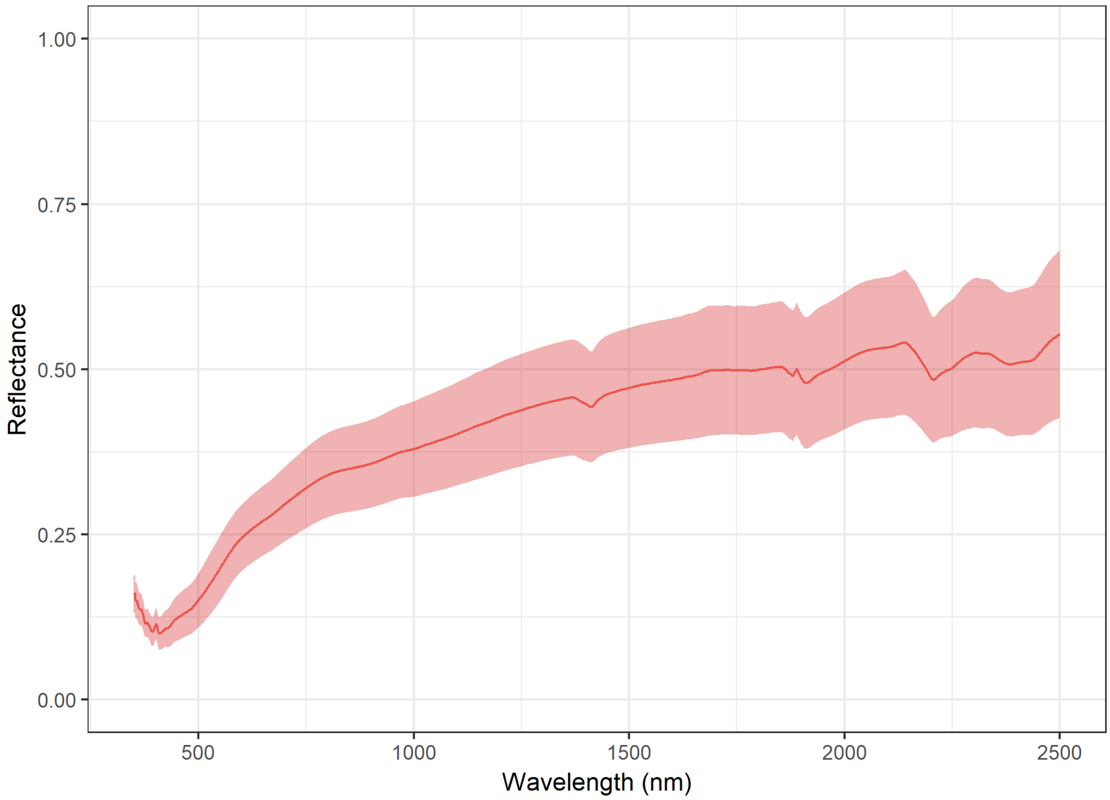

Figure 2-14. The standardized reflectance spectra of Albania (mean ± standard deviation)

2.1.4.2. *Bulgaria*

In Bulgaria, SRTI could not obtain soil samples from previous soil sampling campaigns. Therefore, new soil samples had to be acquired. Due to the severe winter conditions in January-March 2017, the soil sampling campaign was postponed until the weather conditions became more favourable. It was conducted during May-September of 2017. The location of the soil samples is presented in Figure 2-15, whereas Table 2-3 and Figure 2-16 depict the results of the chemical analyses. The textural class of the soil samples is illustrated in Figure 2-17, while Figure 2-18 presents the soil class according to WRB. The samples from Bulgaria are mostly Luvisols, Cambisols, and Fluvisols. The soil classes of CM and FL present little profile differentiation, while the Luvisols contain high activity clays. Indeed, the results indicate that some samples have high clay content. Nevertheless, the samples are well distributed across the dominant soil classes of the region. This is also portrayed through the spectral signatures (Figure 2-19), characterized by well-defined visible regions, and good activity in the SWIR region (dominated mostly by clay minerals).

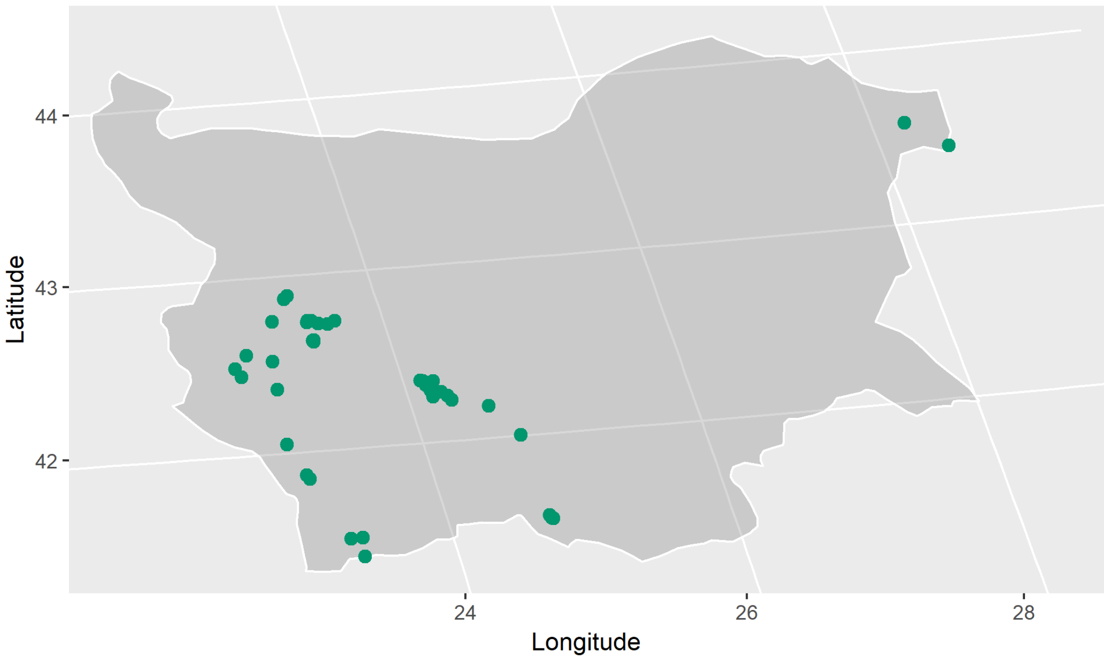

Figure 2-15. Location of the soil samples of Bulgaria

Table 2-3. Major statistical moments of the measured soil properties of Bulgaria

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.00	1.61	1.43	7.86	1.17	1.82	6.97	91
Sand (%)	4.10	41.61	44.00	84.10	21.98	0.03	-1.15	94
Silt (%)	11.30	27.77	24.50	49.30	9.80	0.51	-0.98	94
Clay (%)	4.60	30.67	28.10	77.90	18.11	0.41	-0.76	94
$CaCO_3$ (mg/L)	1.22	6.90	7.51	12.90	4.60	-0.08	-1.86	6
pH (H20)	3.50	5.82	6.00	7.50	1.04	-0.43	-0.73	94
CEC	68.00	87.10	88.00	100.00	8.55	-0.33	-0.77	93

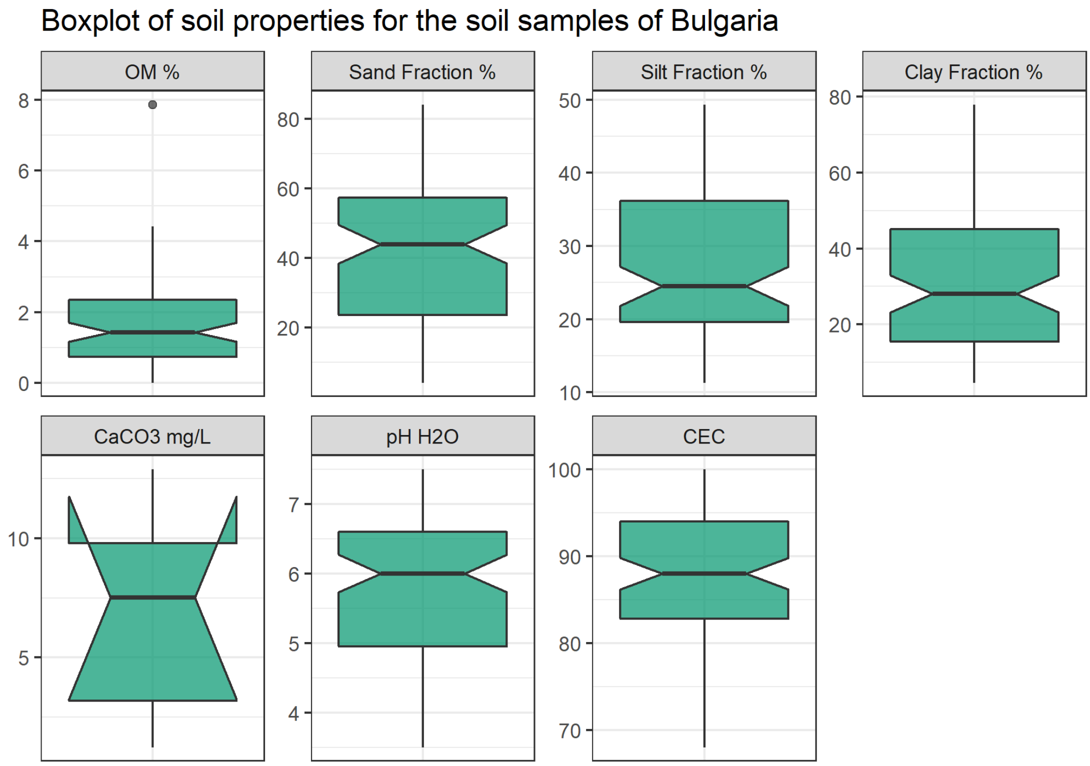

Figure 2-16. Boxplots of the measured soil properties for Bulgaria

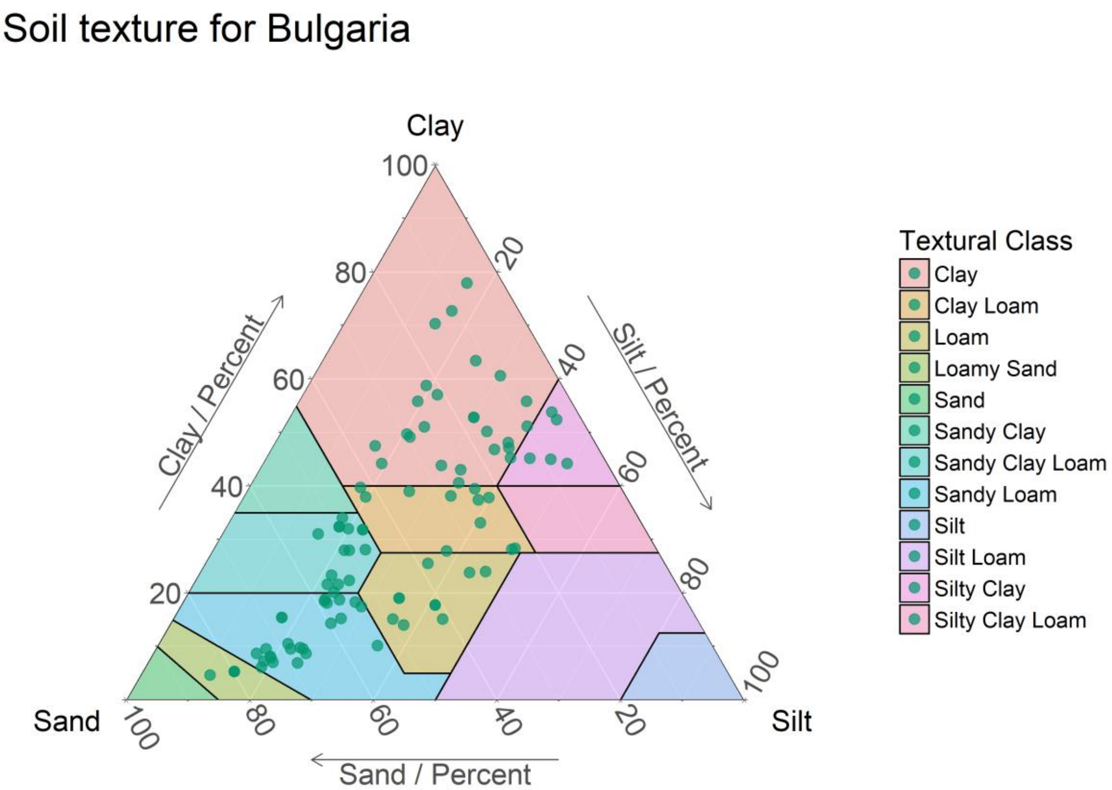

Figure 2-17. Distribution of soil texture class across all soil samples for Bulgaria

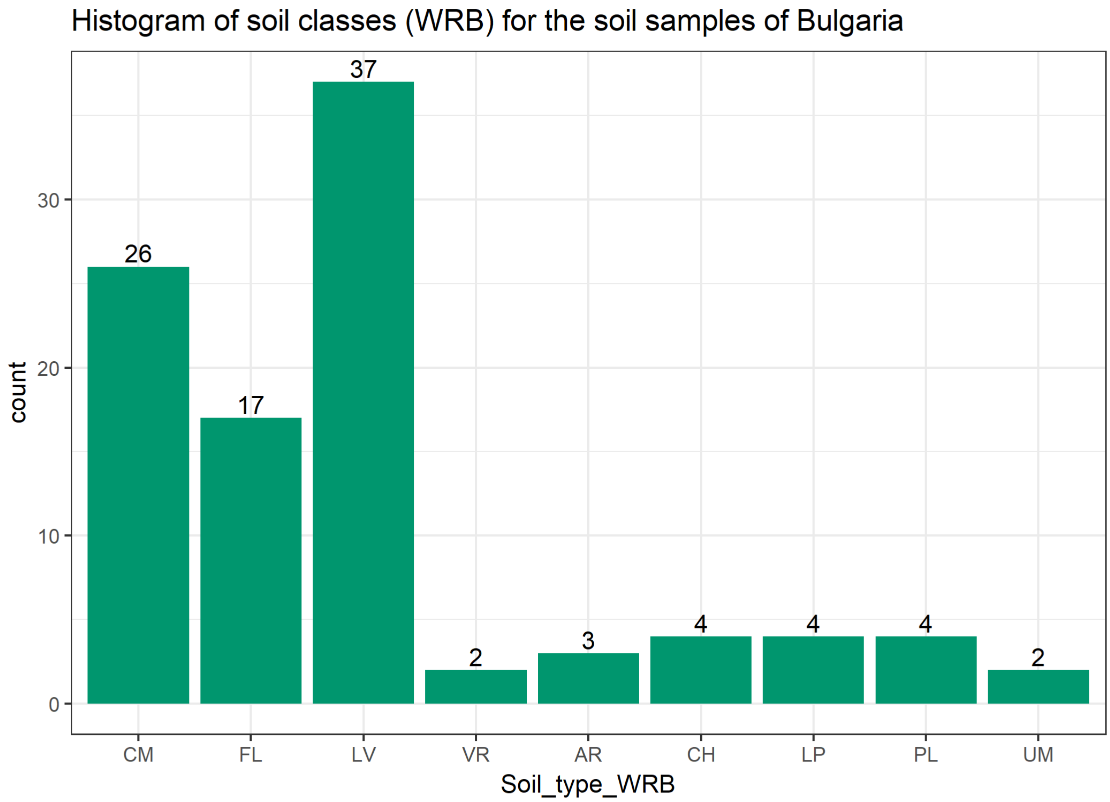

Figure 2-18. Histogram of soil classes of the soil samples of Bulgaria

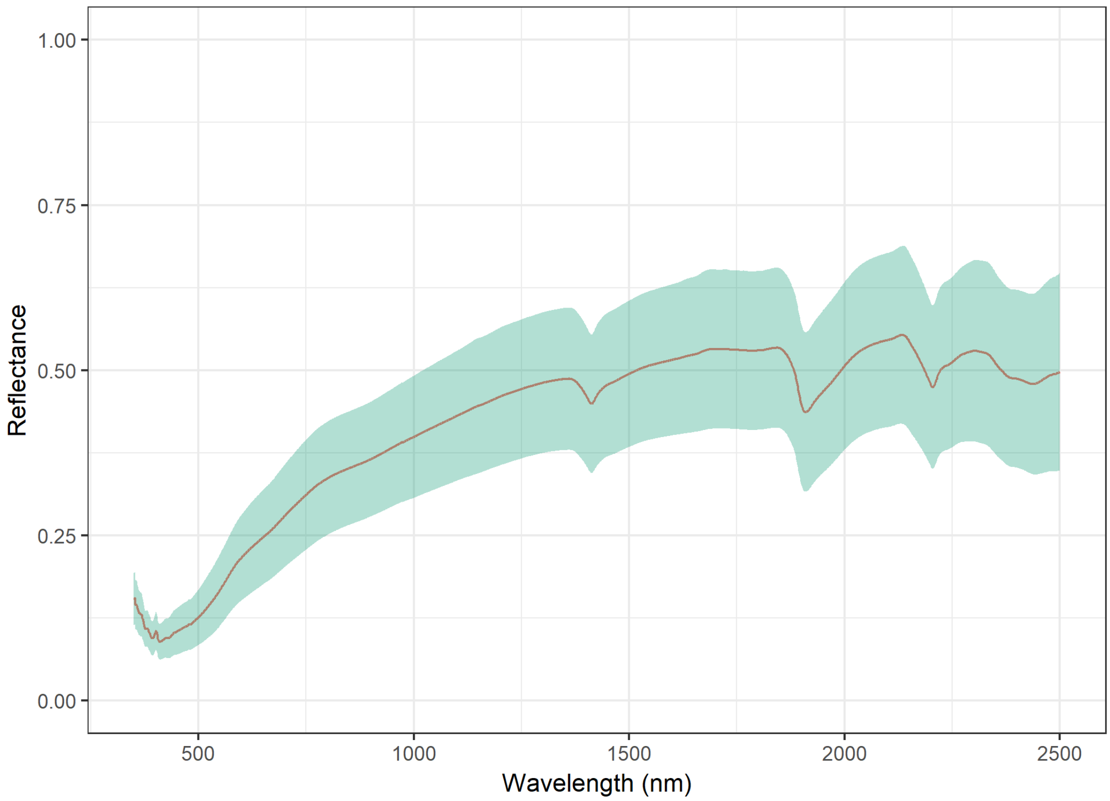

Figure 2-19. The standardized reflectance spectra of Bulgaria (mean ± standard deviation)

2.1.4.3. *Cyprus*

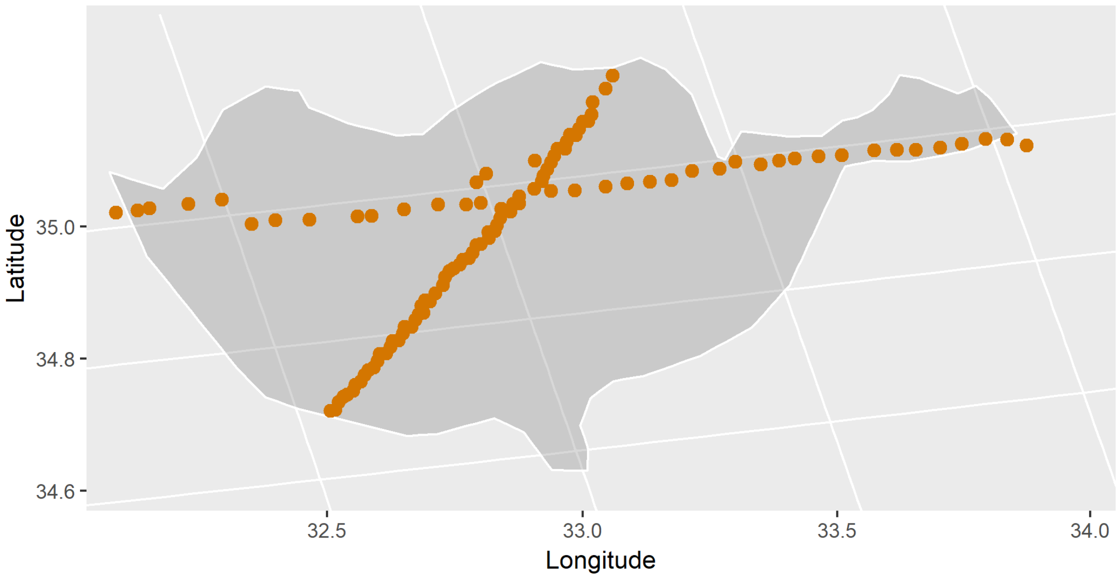

Figure 2-20. Location of the soil samples of Cyprus

Table 2-4. Major statistical moments of the measured soil properties of Cyprus

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.00	0.66	0.08	6.30	1.41	2.51	5.14	96
Sand (%)	25.80	64.14	63.75	88.10	14.95	-0.35	-0.81	94
Silt (%)	10.00	26.36	26.60	46.50	9.22	0.12	-0.98	94
Clay (%)	1.50	9.12	7.10	37.20	7.15	1.51	2.57	94
$CaCO_3$ (%)	1.25	22.47	7.30	81.50	24.96	0.84	-0.93	96
pH (H20)	5.95	7.91	7.97	10.07	0.72	0.08	0.61	96
EC (mS)	0.5	1.52	1.35	6.60	0.95	2.34	11.40	96

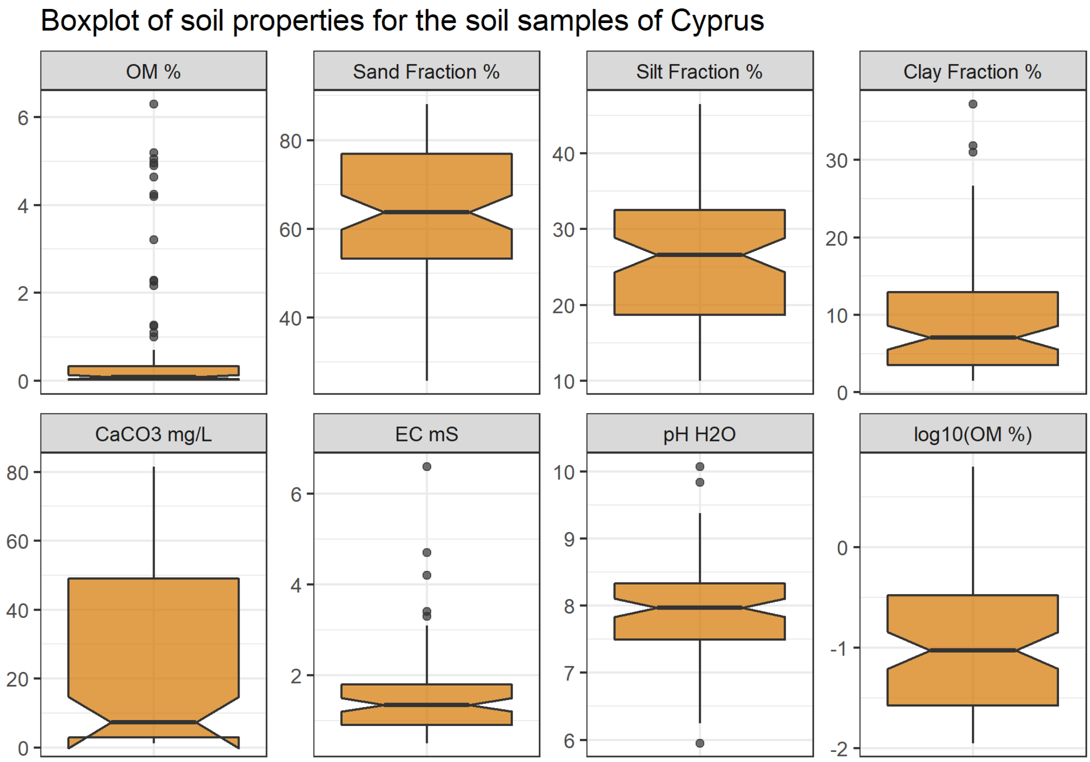

Figure 2-21. Boxplots of the measured soil properties for Cyprus

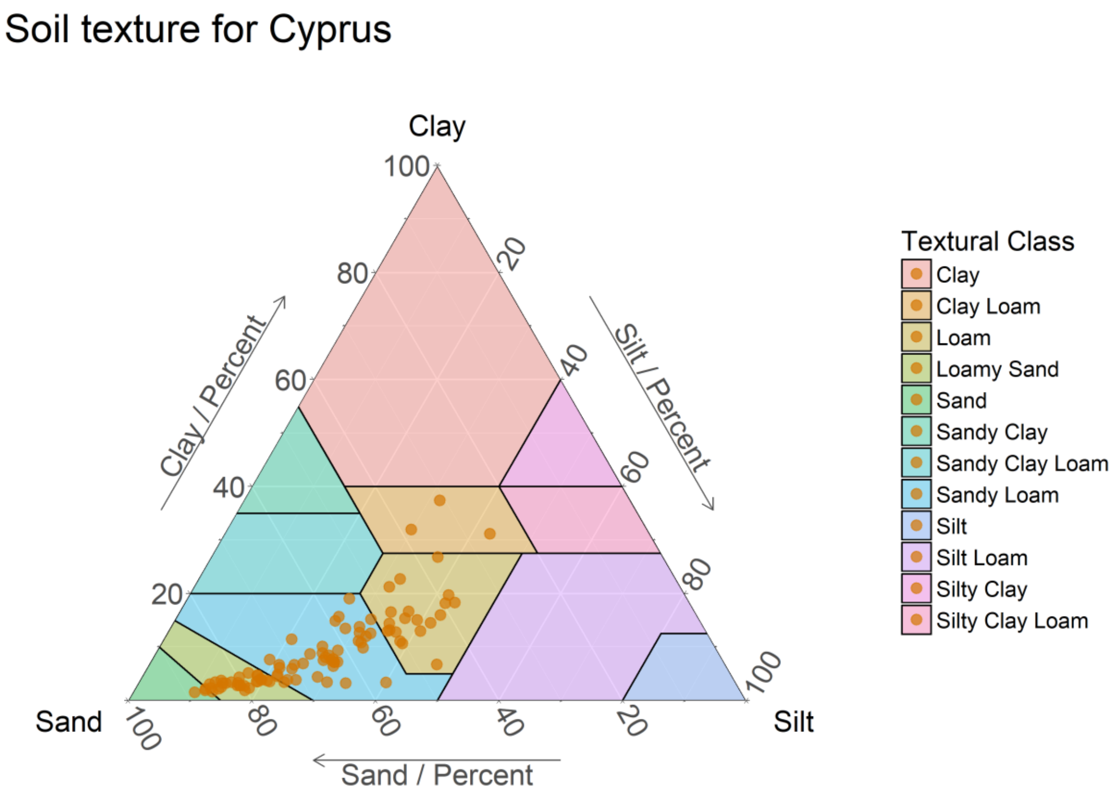

Figure 2-22. Distribution of soil texture class across all soil samples for Cyprus

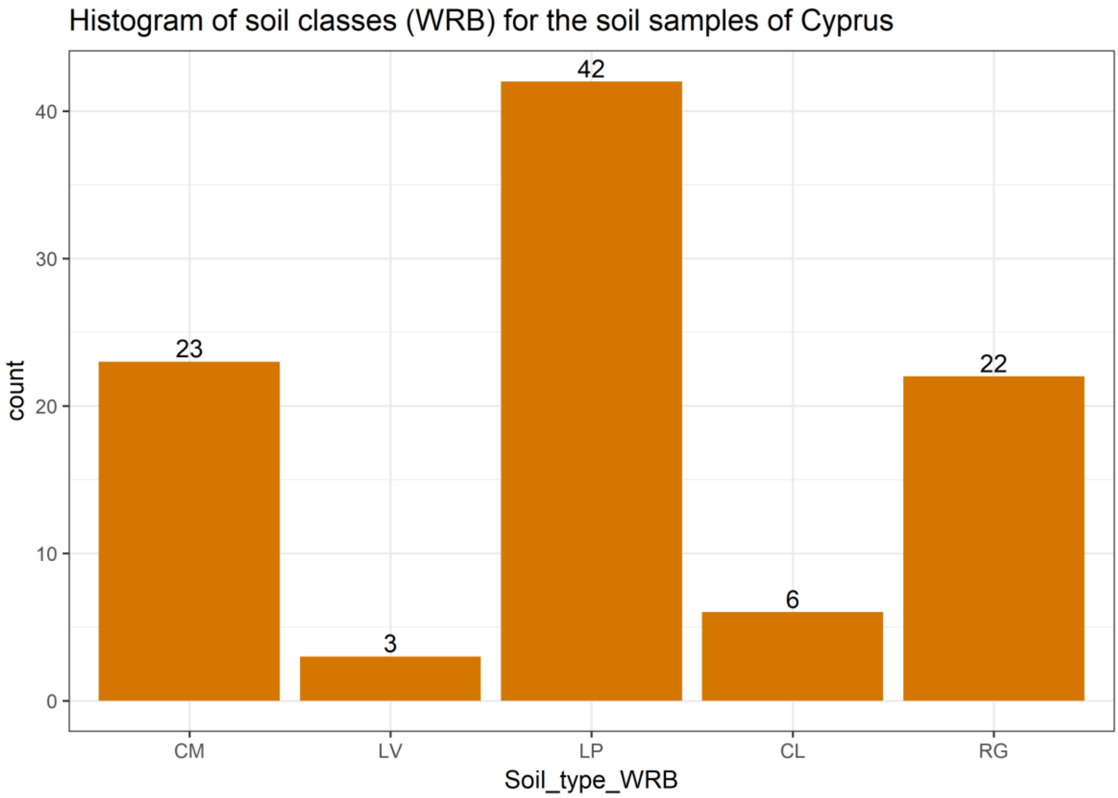

Figure 2-23. Histogram of soil classes of the soil samples of Cyprus

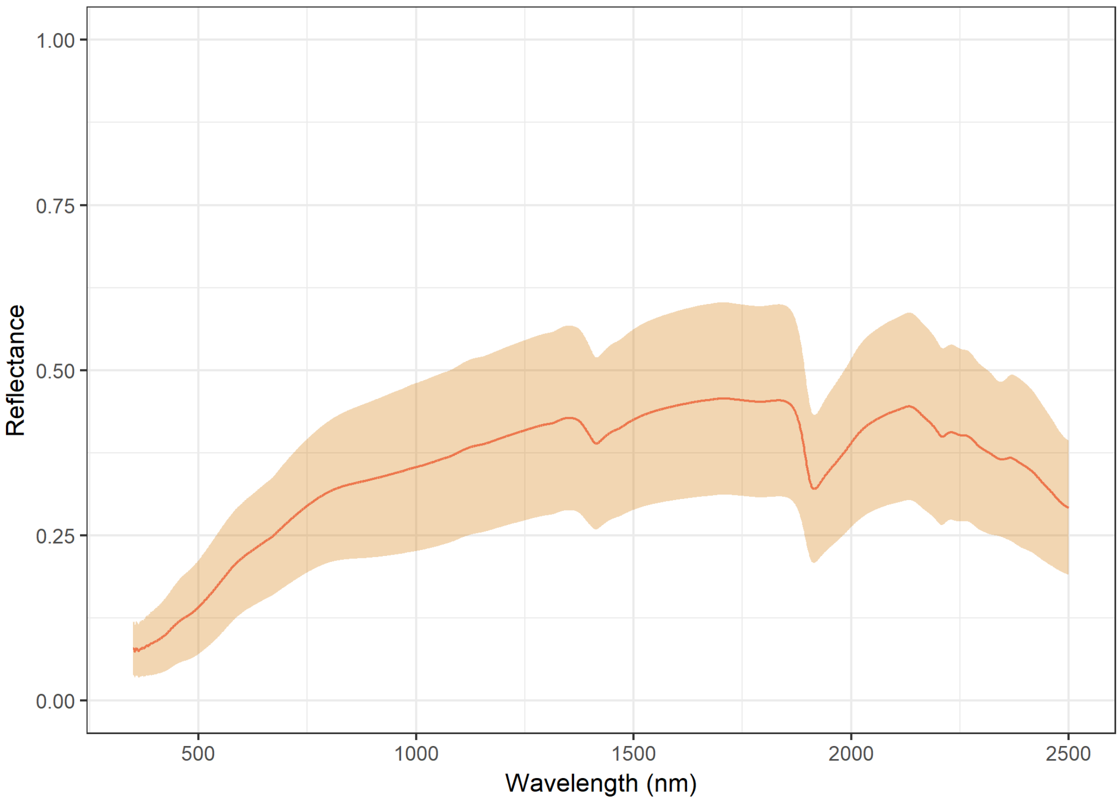

Figure 2-24. The standardized reflectance spectra of Cyprus (mean ± standard deviation)

2.1.4.4. *Egypt*

Egypt, being the only country representing North Africa in the regional SSL was identified as an important asset in the regional SSL. However, the case of Egypt had increased difficulties. Despite the fact that there exist detailed soil libraries in Egypt (with physical soil samples and their assorted physical and chemical measurements), the partners of CEDARE were unable to obtain official permits to send a large number of physical soil samples with sufficient enough matter (i.e. mass and volume) to i-BEC, Greece. In light of this event, and given that CEDARE didn't possess the necessary equipment to perform standardized spectral measurements themselves, CEDARE managed to send only a small number of samples to i-BEC. The total number of samples that are in the GEO-CRADLE SSL from Egypt is 10, and their location is presented in Figure 2-25. Due to the small number of samples which cannot be representative of the whole country, no statistical information from this population was derived.

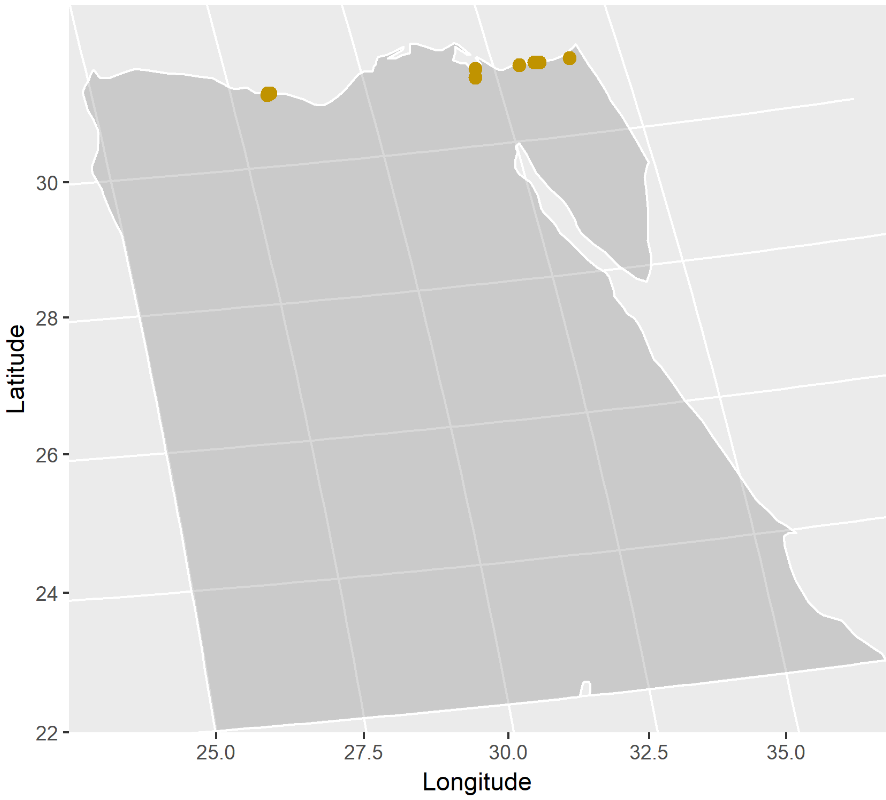

Figure 2-25. Location of the soil samples of Egypt

2.1.4.5. *FYROM*

USCM obtained 124 physical soil samples from a prior sampling campaign, conducted in 2015. They cover a large region of the country's pedological variability and are representative of the country (Figure 2-28). The chemical analysis results are given in Figure 2-27 and Table 2-5. As far as the textural parameters are concerned, only the Clay Fraction is available. Nevertheless, the soil class of the samples (Figure 2-28) which indicate that the samples are mostly Fluvisols, and their respective spectral signatures (Figure 2-29) indicate that the samples are mostly silty, with limited clay content.

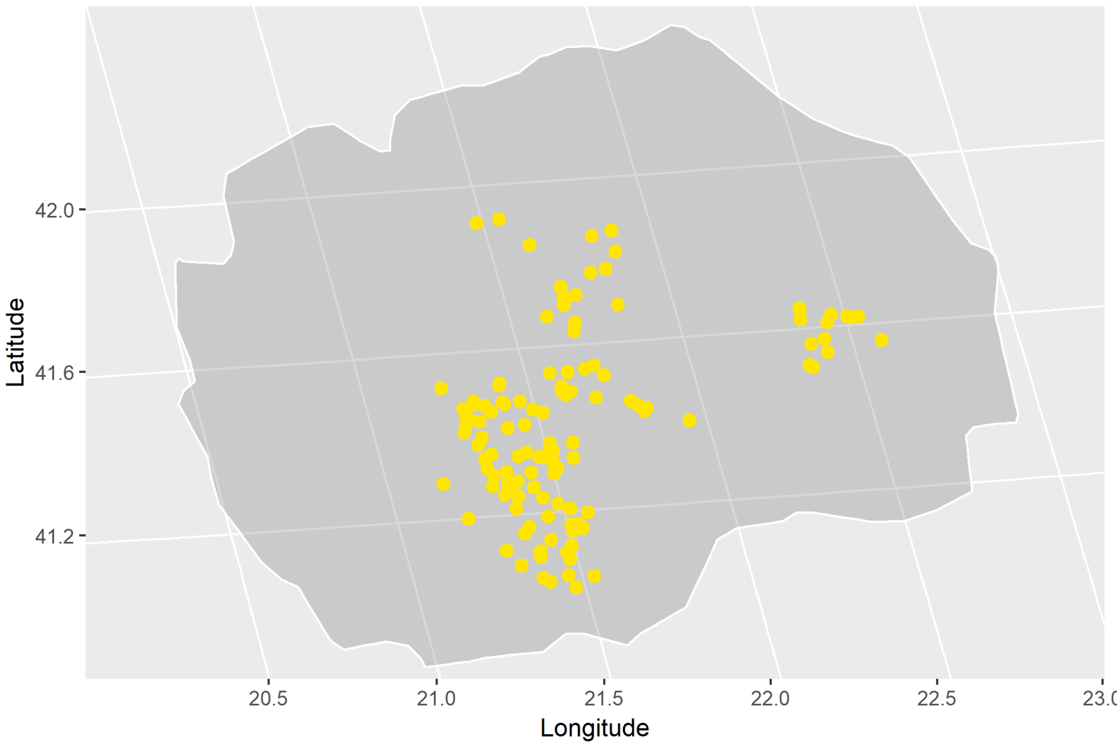

Figure 2-26. Location of the soil samples of FYROM

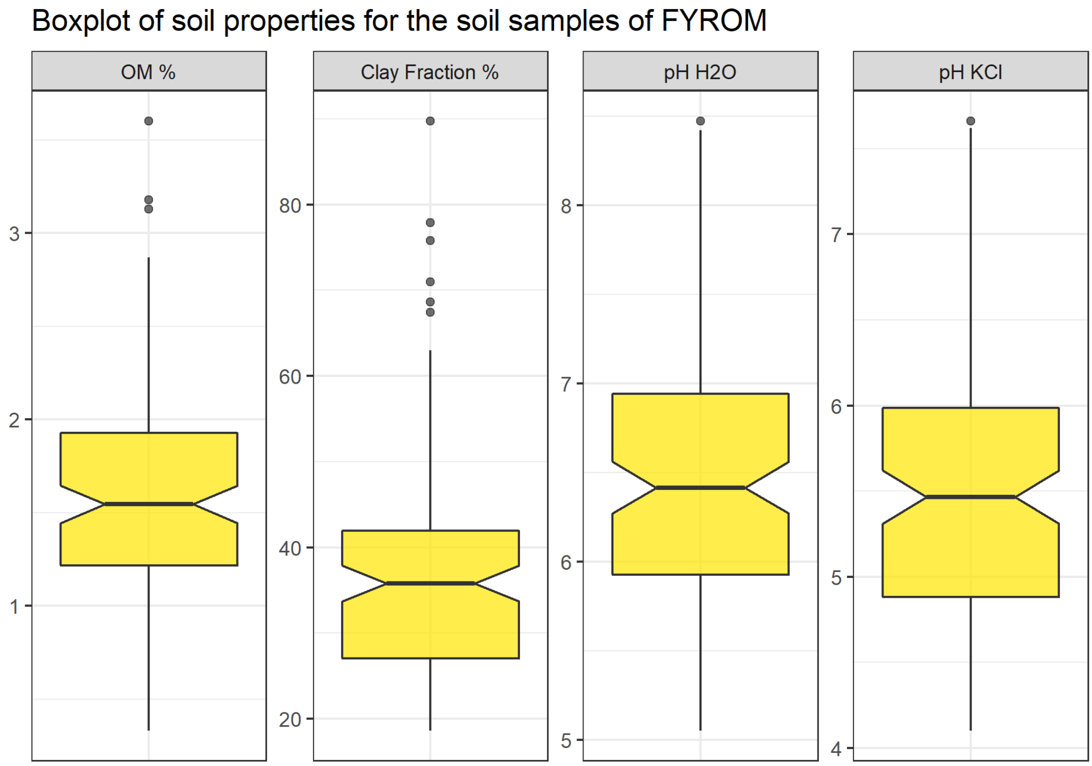

Figure 2-27. Boxplots of the measured soil properties for FYROM

Table 2-5. Major statistical moments of the measured soil properties of FYROM

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.33	1.5895	1.55	3.60	0.5548	0.7159	1.1995	124
Clay (%)	18.60	37.3032	35.80	89.70	13.0390	1.2824	2.1375	124
pH (H20)	5.05	6.5098	6.42	8.47	0.8536	0.5096	-0.3976	124
pH (KCl)	4.10	5.5619	5.47	7.66	0.8941	0.6175	-0.3575	124

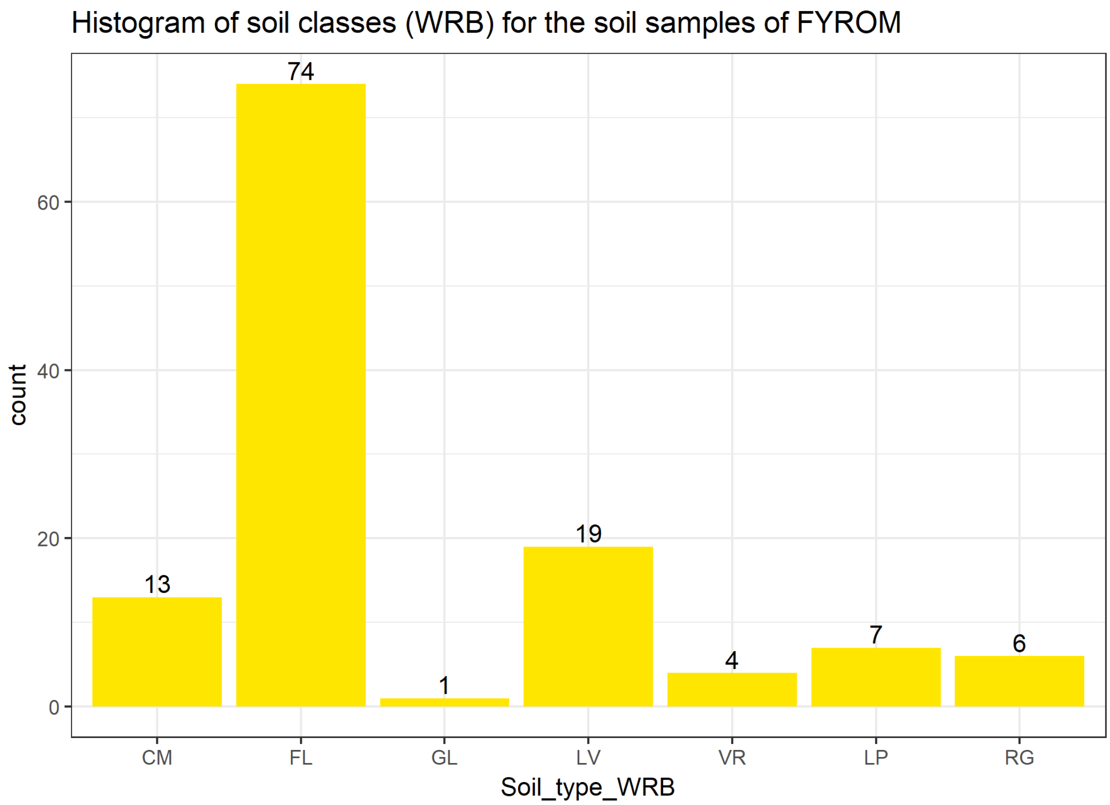

Figure 2-28. Histogram of soil classes of the soil samples of FYROM

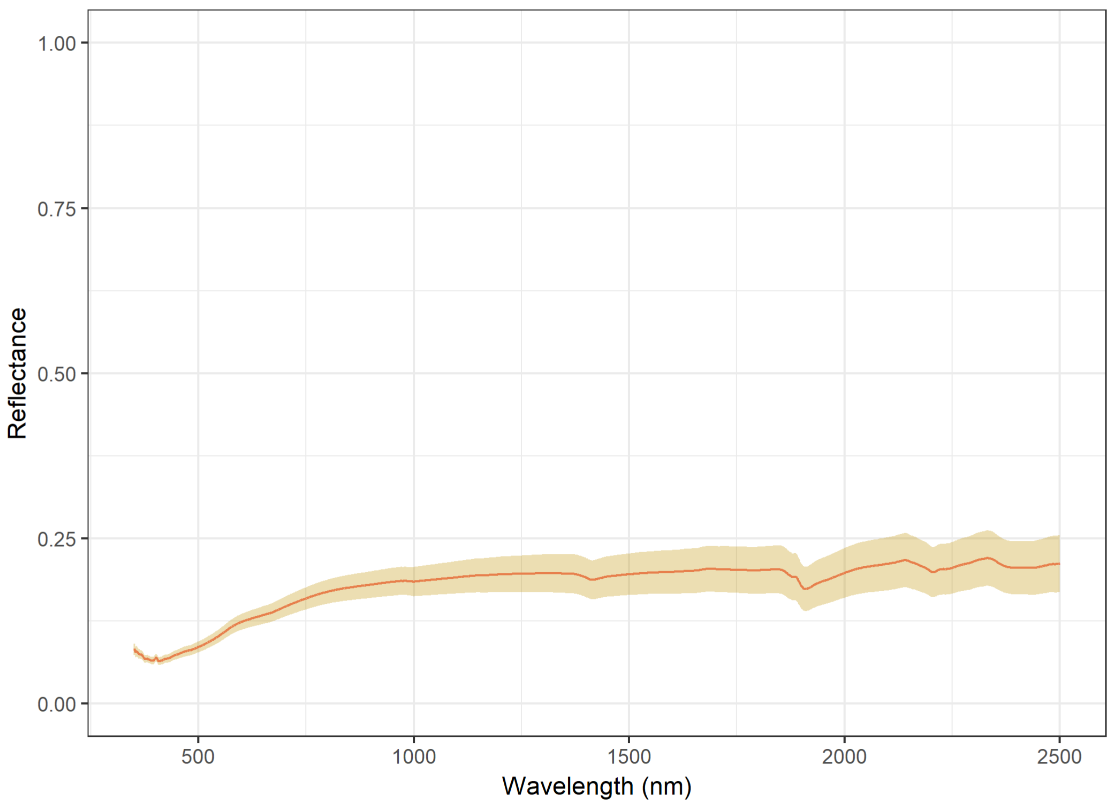

Figure 2-29. The standardized reflectance spectra of FYROM (mean ± standard deviation)

2.1.4.6. *Greece*

The soil samples of Greece originate from one of the thirteen administrative regions in Greece, namely the region of Eastern Macedonia and Thrace, located in North-Eastern Greece (Figure 2-30). The area of interest is the agricultural lands surrounding the Nestos River, covering an area of approximately 600 square kilometres (Figure 2-31). The river Nestos is one of the largest in Greece, stems from Bulgaria and flows into the Thracian sea. The geology of this area is characterized primarily by the alluvial deposits of the river. It is a relatively flat area, with an altitude between 15-90m, while the inclinations are less than 4%. The most cultivated crops in these lands are kiwi fruits, plum trees, peach trees, nectarine trees, asparagus crops, and apricot trees.

Figure 2-30. Relative location of the Eastern Macedonia and Thrace region in northern Greece

The soil samples were collected between April and June of 2015 within the frame of the European Project “AGROLESS – Joint reference strategies for rural activities of reduced inputs” (B3.11.02) of the European Territorial Cooperation Programme “Greece-Bulgaria 2007-2013”. To identify the sampling locations, a stratified sampling was conducted using as strata the Great Group of the USDA soil taxonomy system. The soils were sampled from three different layers, denoted as layer A (0-30 cm), Layer B (30-60 cm), and Layer C (60-90 cm).

Table 2-6 depicts significant statistical moments of the measured soil properties, while Figure 2-33 illustrates them in a form of a Notch boxplot. The soil textural class of the samples may be found in Figure 2-35. From these, we can infer that the soils in this area are mostly sandy, and somewhat moderate in terms of organic matter. The lower values of organic matter are due to the fact that approximately a third of the samples were collected from the deeper layer C. The soil class distribution of the soil samples is presented in Figure 2-35. The most predominant classes are the Fluvisols and Leptosols. The reflectance spectra after standardization are given in Figure 2-36. The high overall observed reflectance may be attributed to the large presence of sand, which shifts the reflectance upwards due to the particle size effect. There exists a significant rise in the visible part of the spectrum, and the small absorption band around 900 nm may be attributed to the presence of iron.

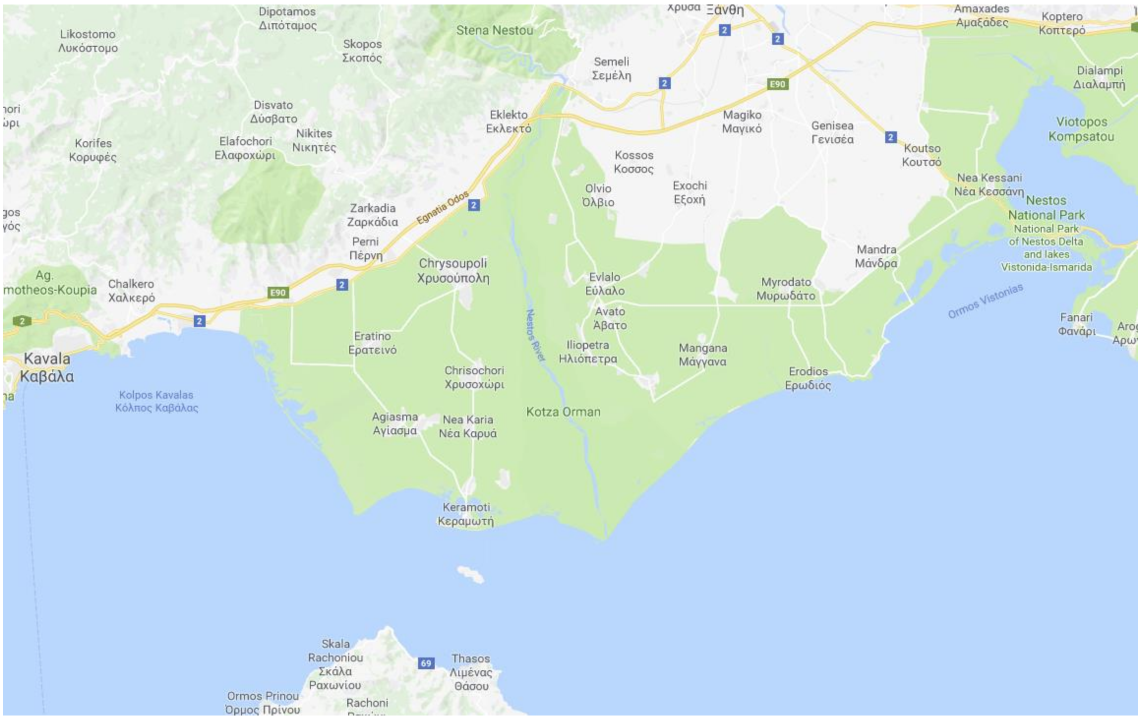

Figure 2-31. The Nestos river delta, and the surrounding agricultural lands

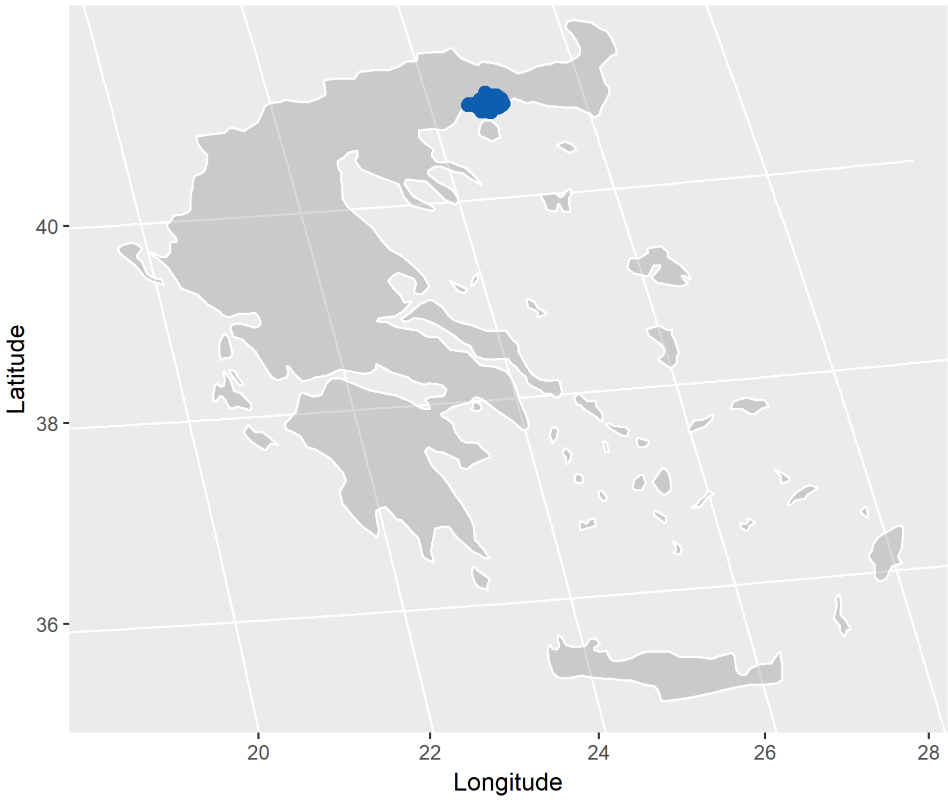

Figure 2-32. Location of the soil samples of Greece

Table 2-6. Major statistical moments of the measured soil properties of Greece

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0	0.9401	0.86	4.18	0.6287	1.0880	2.0493	928
Sand (%)	2	59.0043	59.00	99.00	20.4710	0.0945	-0.6216	928
Silt (%)	0	26.1272	26.00	68.00	14.7009	0.0858	-0.8567	928
Clay (%)	0	14.9321	13.00	91.00	11.1773	1.8031	5.5072	928
NO_3 ppm	0	17.7938	5.60	661.20	38.9528	7.4106	92.3701	928
$CaCO_3$ (%)	0	0.5033	0.00	40.30	2.1806	11.4630	172.7943	928

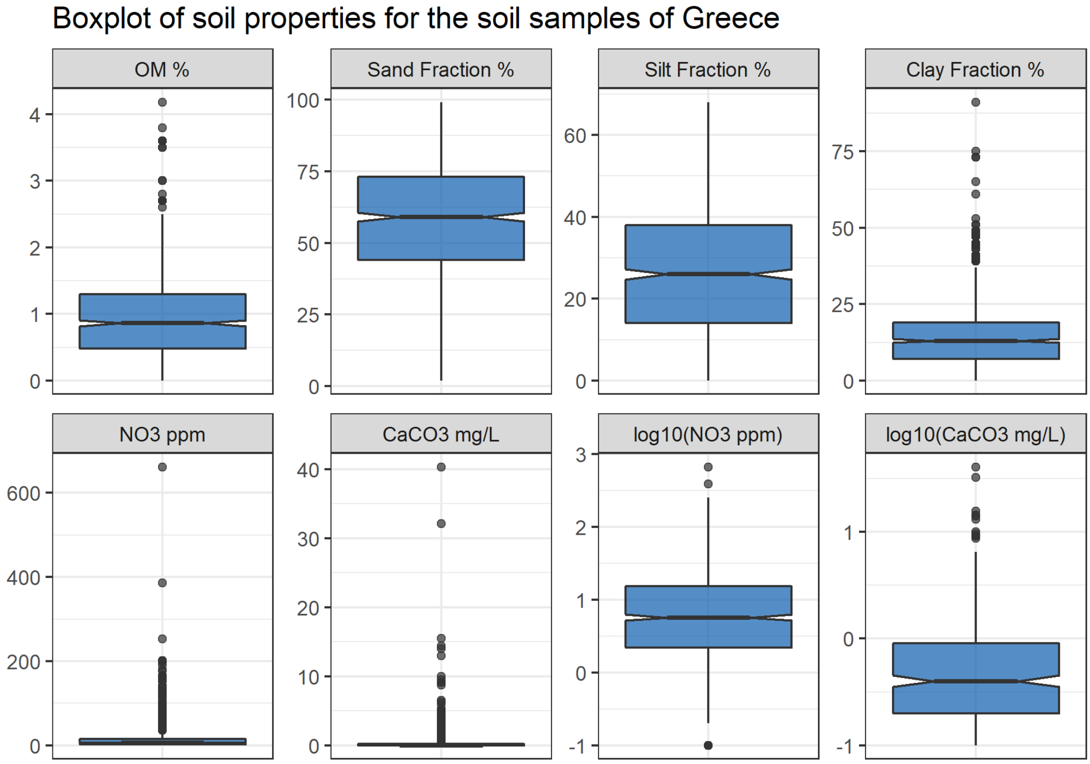

Figure 2-33. Boxplots of the measured soil properties for Greece

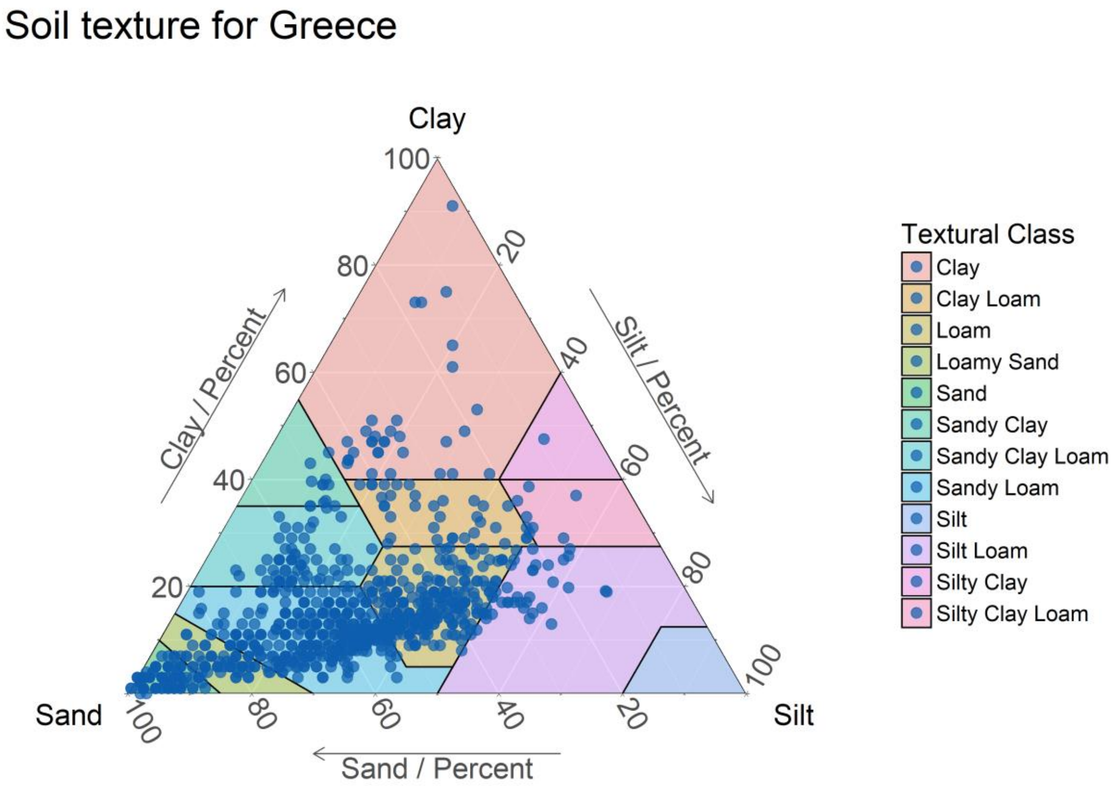

Figure 2-34. Distribution of soil texture class across all soil samples for Greece

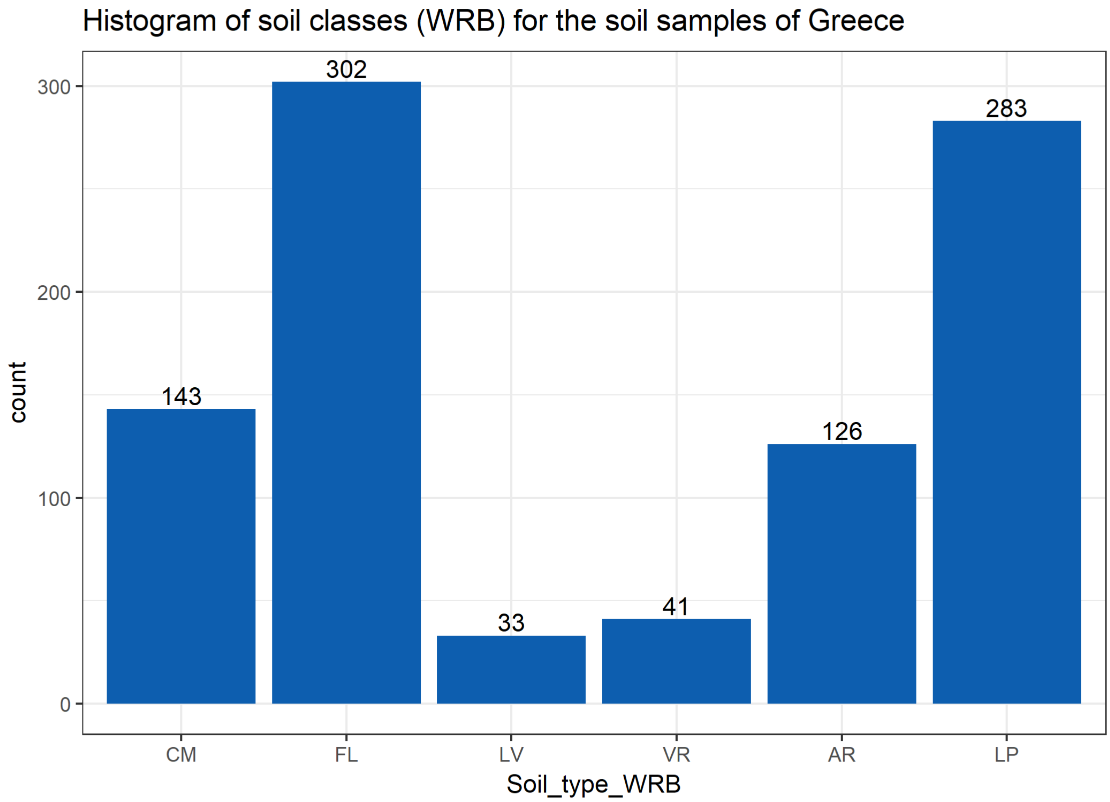

Figure 2-35. Histogram of soil classes of the soil samples of Greece

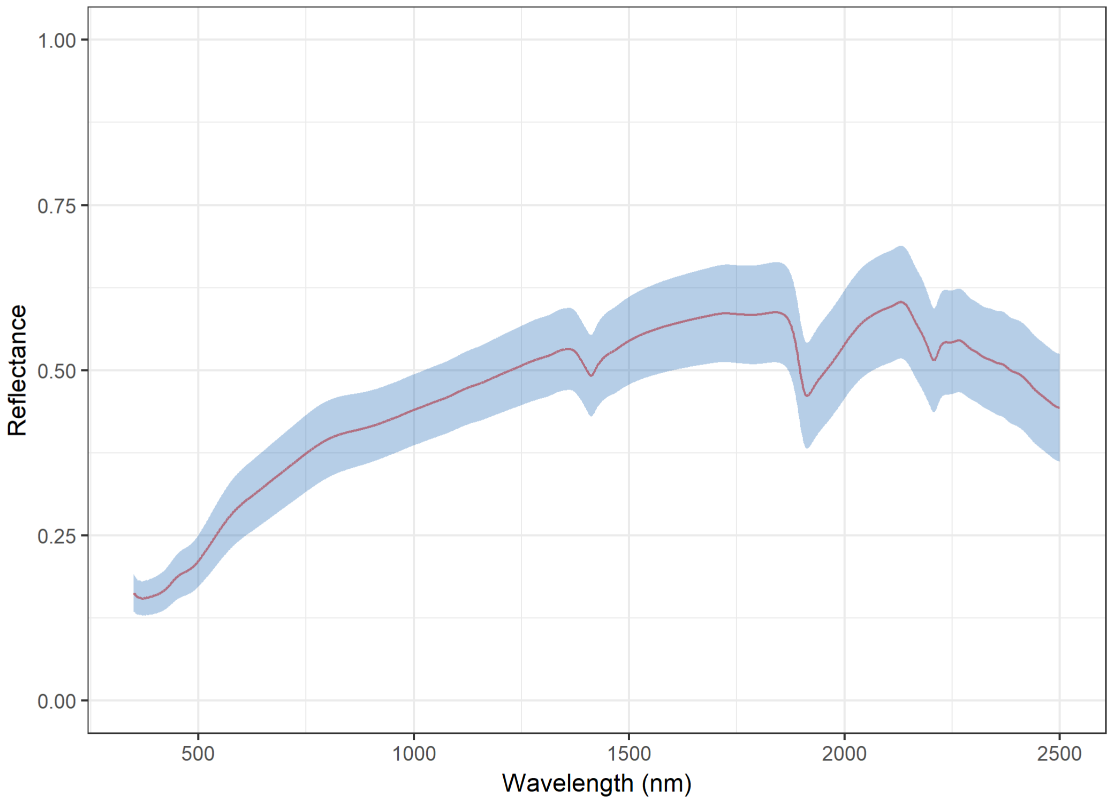

Figure 2-36. The standardized reflectance spectra for Greece (mean ± standard deviation)

2.1.4.7. *Israel*

The soil database of Israel contains 221 soil samples which were collected from various geographic locations and depths. The chemical-physical properties which were included in the dataset and the number of samples from each property (in addition to major statistical moments) are to be found in Figure 2-38 and Table 2-7.

The dataset contains soil samples that were collected in 1970, 1989 and 2017. The 1970 samples were collected by the Ministry of Agriculture and Rural Development as part of a national soil survey project. The 1989 samples were collected and analysed by Professor Eyal Ben-Dor, as part of his Ph.D. degree. The most recent samples were collected by Ms. Nurit Ben Hagai from Sherut sade Zemach (http://www.zemach-nisyonot.co.il) as part of a local soil survey project.

To identify the sampling locations, a stratified sampling was conducted using as strata the subgroup of the USDA soil taxonomy system. Soils were sampled according the diagnostic horizons exist in each profile. Therefore, some profiles contain 2 samples while others 6 samples.

Table 2-7 shows the soil property measures of the database and Figure 2-38 present it in a form of a boxplot. The low values of OM in the dataset are due to the location of the sampling points. Most of the samples were collected in semi-arid and arid climate zones (the southern and eastern part of Israel) or from the lower part of the soil profiles. The arid climates together with the limestone bedrock in the central mountain range also affect the high percentages of calcium carbonate and the higher clay contents (Figure 2-39) in the samples. In contrast, the coastal areas are dominated mostly by the sandy soils texture whose continuity is interrupted by alluvial sedimentation.

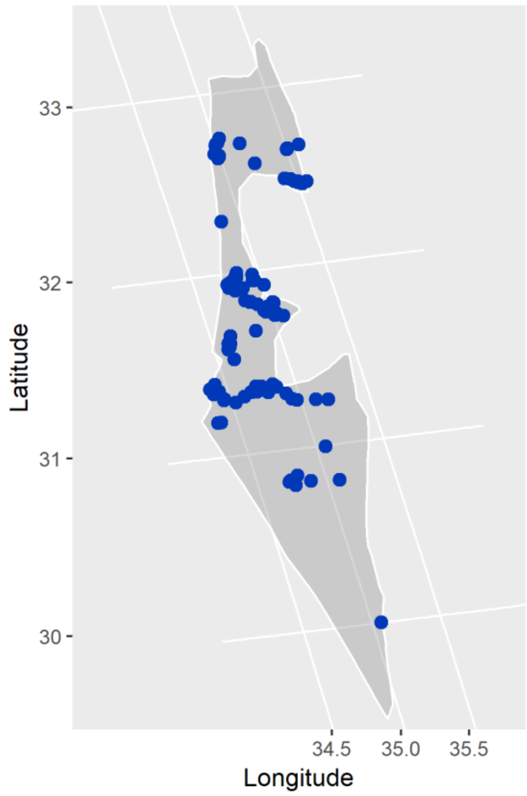

Figure 2-37. Location of the soil samples of Israel

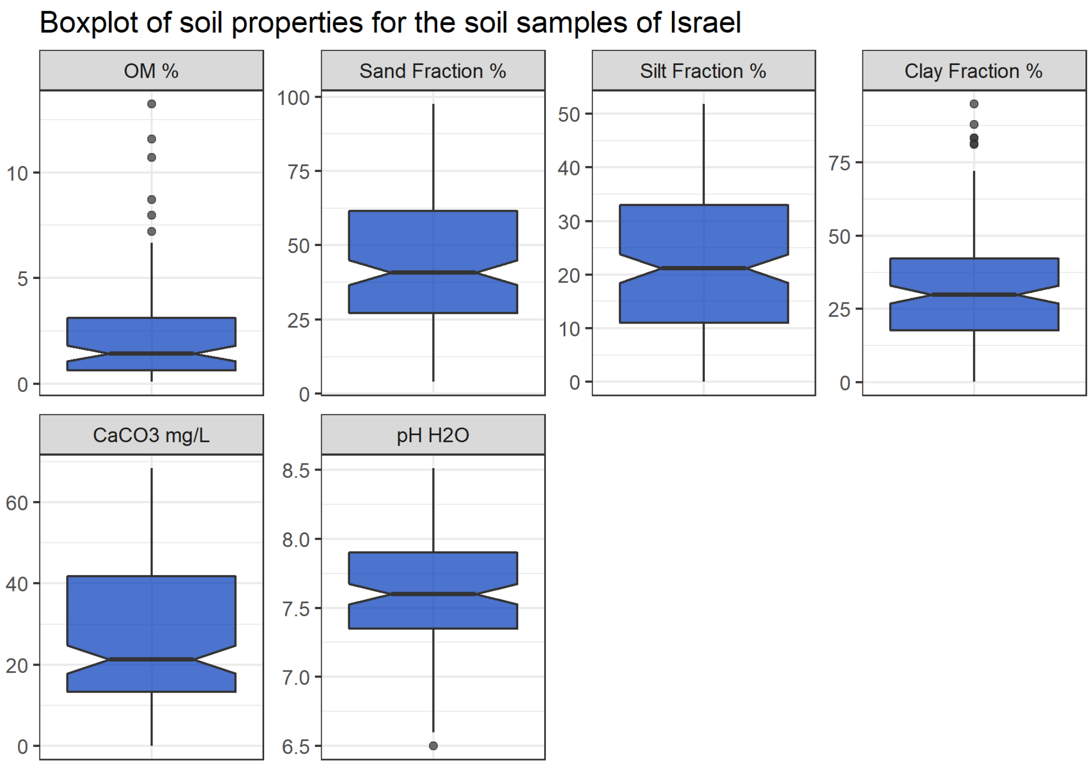

Figure 2-38. Boxplots of the measured soil properties for Israel

Table 2-7. Major statistical moments of the measured soil properties of Israel

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.09	2.2858	1.43	13.23	2.4412	2.1144	8.1727	113
Sand (%)	4.03	45.7762	40.80	97.50	24.4758	0.3489	2.1616	168
Silt (%)	0.00	21.8301	21.15	51.80	13.1211	0.0242	1.9686	168
Clay (%)	0.20	32.3180	29.95	94.80	19.4955	0.8566	3.4469	168
$CaCO_3$ (%)	0.00	26.3482	21.30	68.34	18.5582	0.5502	2.3537	169
pH (H20)	6.50	7.6060	7.60	8.51	0.3914	-0.0965	2.7976	136
EC (µS)	0.07	1.2049	0.67	7.97	10.5286	0.5385	8.64260	141

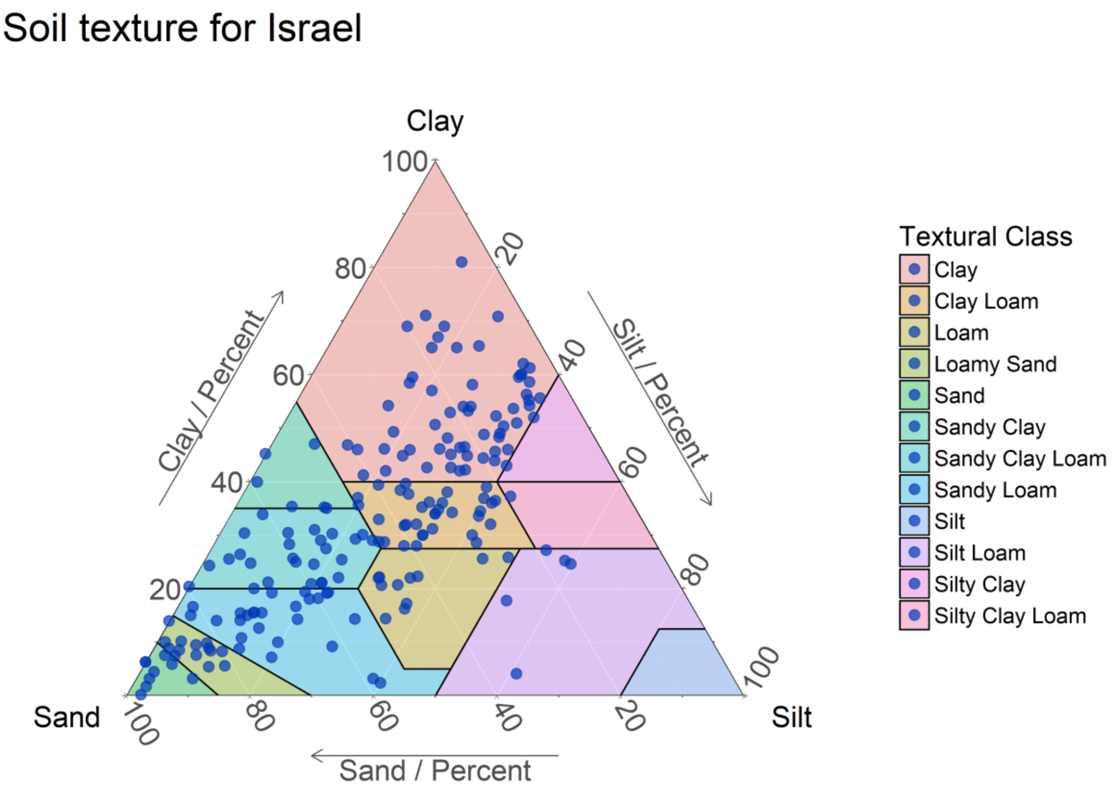

Figure 2-39. Distribution of soil texture class across all soil samples for Israel

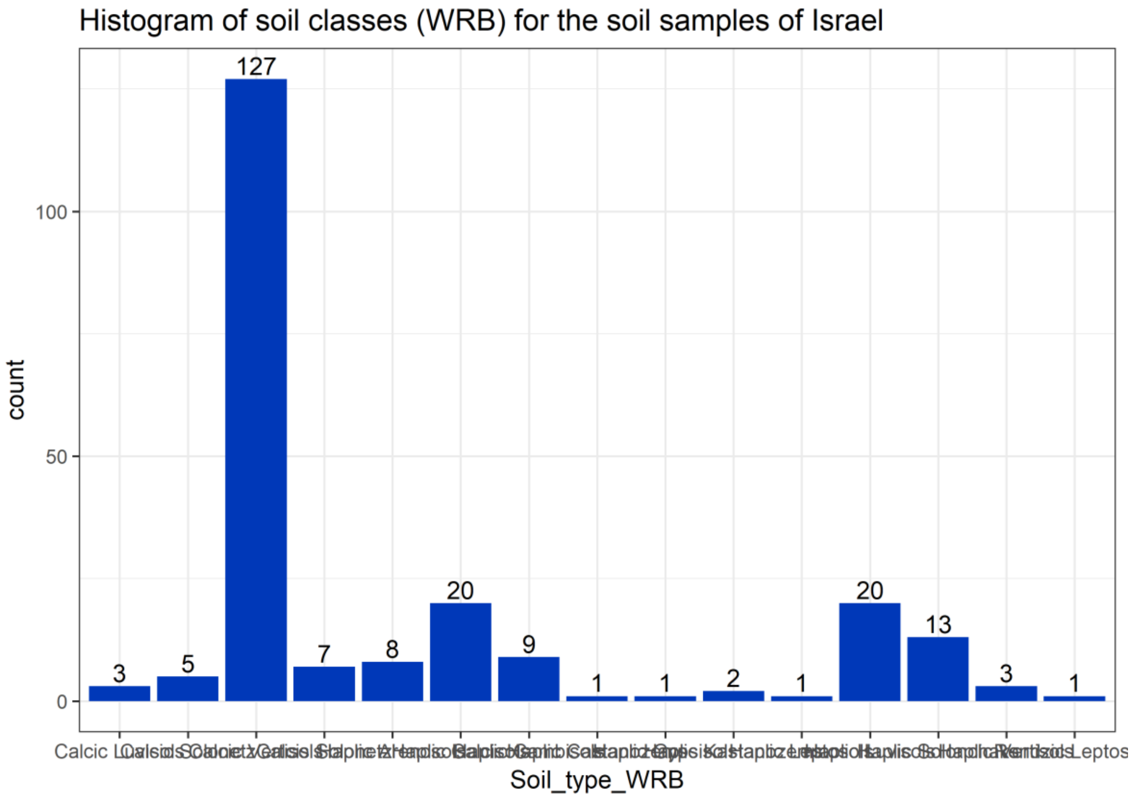

Figure 2-40. Histogram of soil classes of the soil samples of Israel

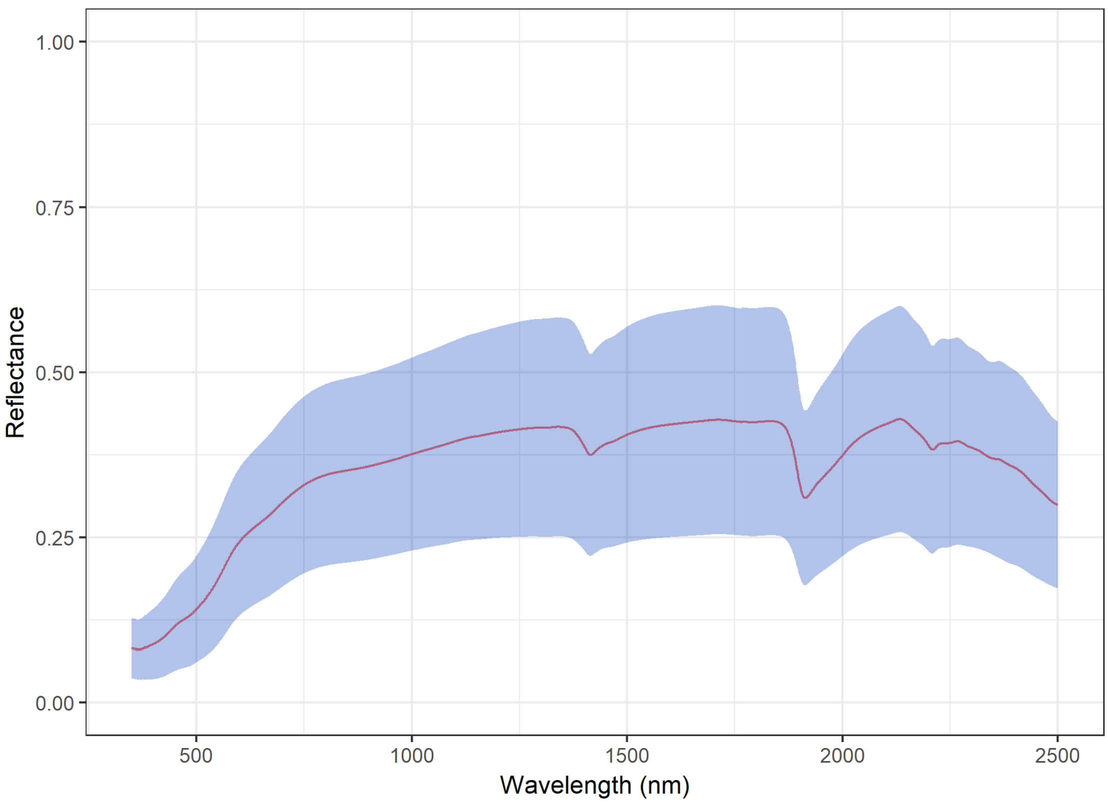

Figure 2-41. The standardized reflectance spectra of Israel (mean ± standard deviation)

2.1.4.8. *Serbia*

Serbia is a country which prior to the GEO-CRADLE project had not contributed in any effort to establish a soil spectral library. In other words, the dataset made available through GEO-CRADLE is the first effort to create an open SSL for the country, adhering to strict standards. Although no SSL existed, some soil libraries and soil sampling campaigns were conducted, primarily for the development of soil classification maps and soil thematic maps of the area. The IPB was able to secure a number of such soil samples which were sampled in previous years, and no new soil sampling campaigns were conducted. The location of these soil samples is depicted in Figure 2-42; it covers a number of regions in the country. The results of the chemical analysis may be found in Table 2-8 and Figure 2-43.

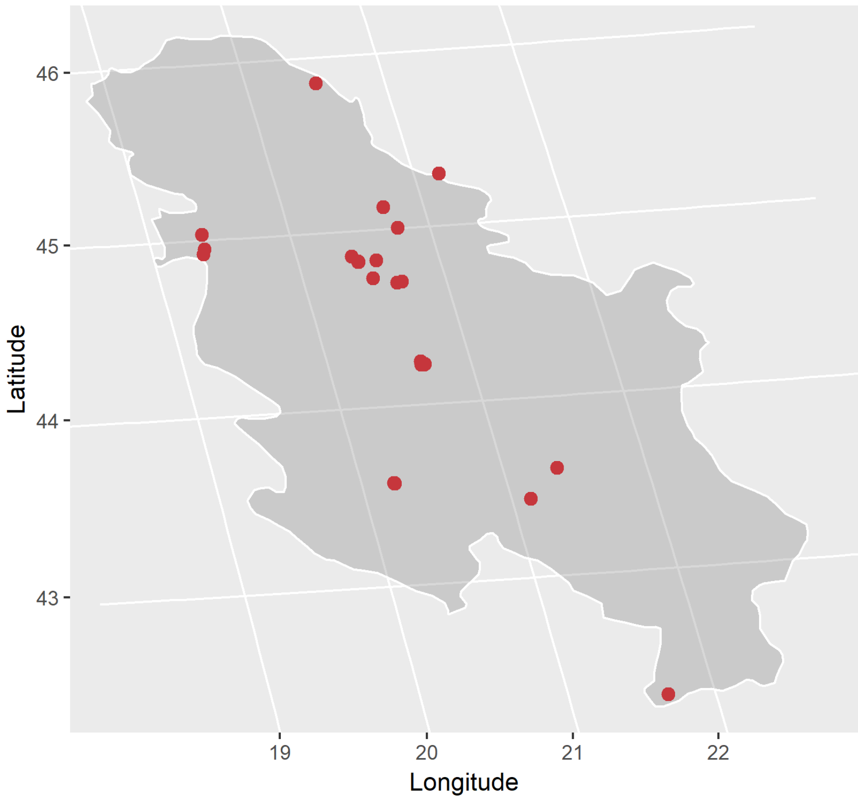

Figure 2-42. Location of the soil samples of Serbia

Table 2-8. Major statistical moments of the measured soil properties of Serbia

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.11	1.3148	0.91	4.76	0.9638	1.5854	5.1155	63
Clay (%)	1.32	22.9100	21.93	54.21	11.7444	0.2715	2.6786	63
NO_3	0.05	0.6995	0.33	4.72	0.9389	2.6163	9.7011	63

$CaCO_3$ (%)	0.01	0.7803	0.72	2.05	0.5285	0.5275	2.4241	63
pH (H_20)	5.41	7.3182	7.50	8.57	0.7760	-1.0270	3.2552	63
pH ($CaCl_2$)	4.67	6.4619	6.47	7.71	0.7167	-0.6402	3.1820	63
pH (KCl)	4.05	6.0348	6.05	7.31	0.7929	-0.7093	2.9159	63

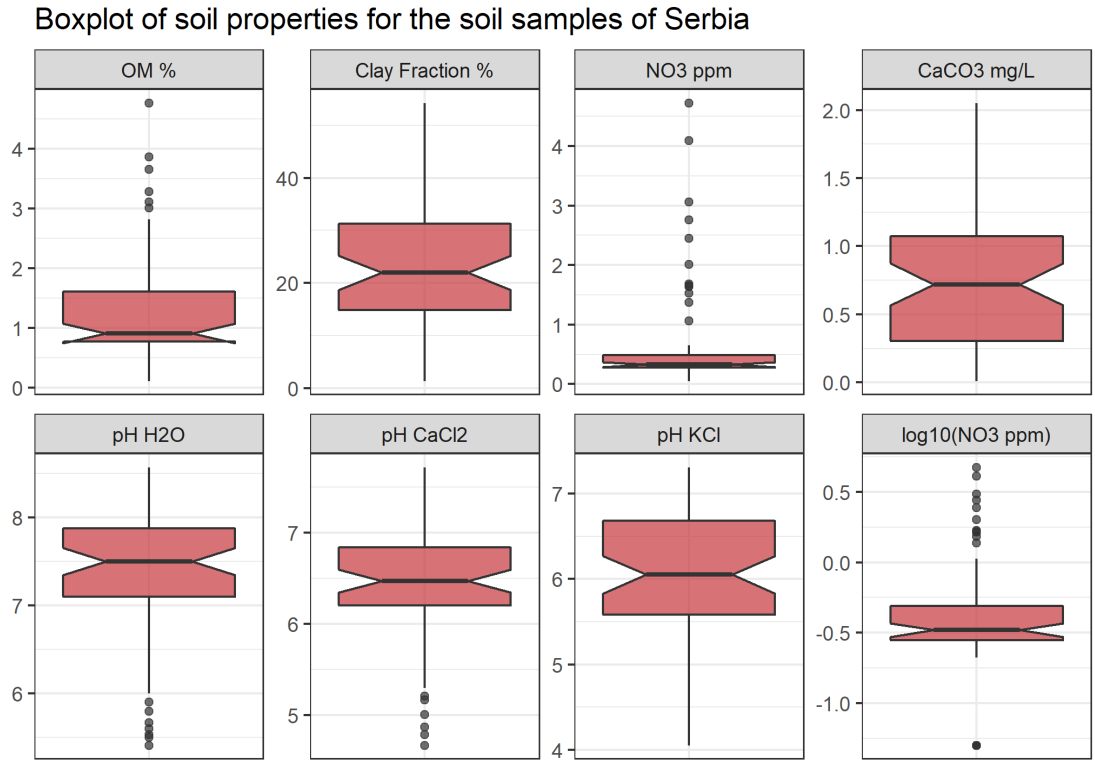

Figure 2-43. Boxplots of the measured soil properties for Serbia

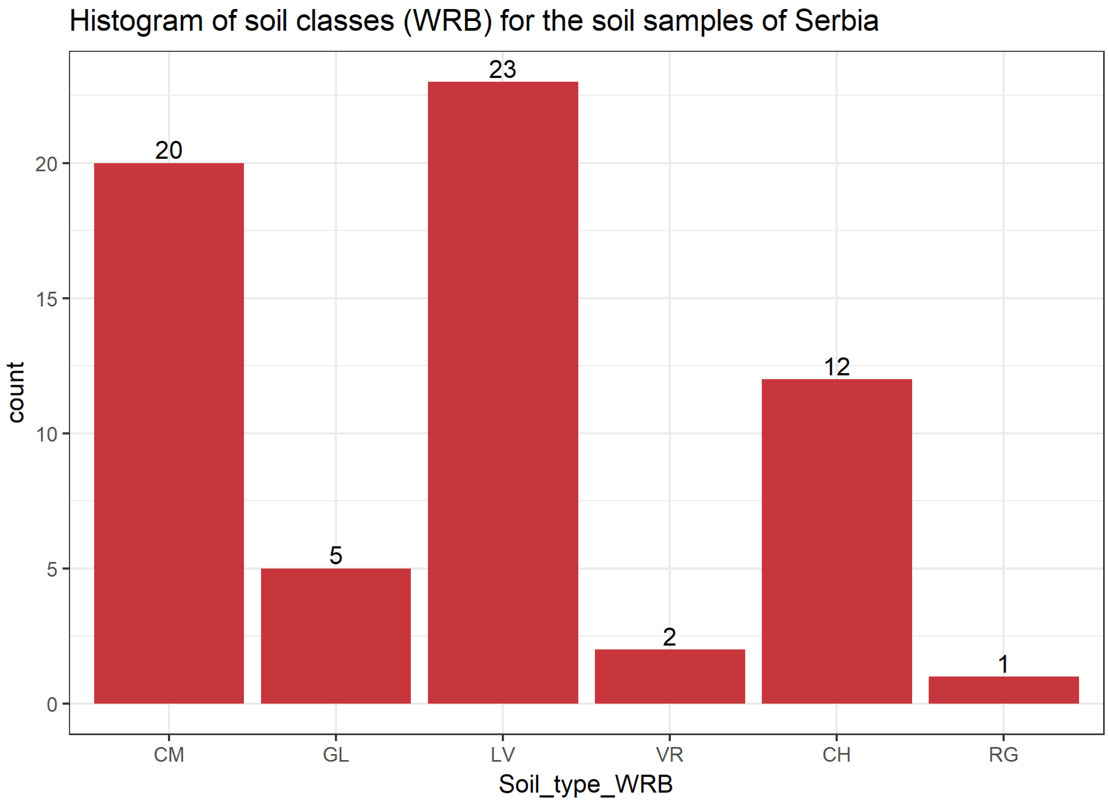

Figure 2-44. Histogram of soil classes of the soil samples of Serbia

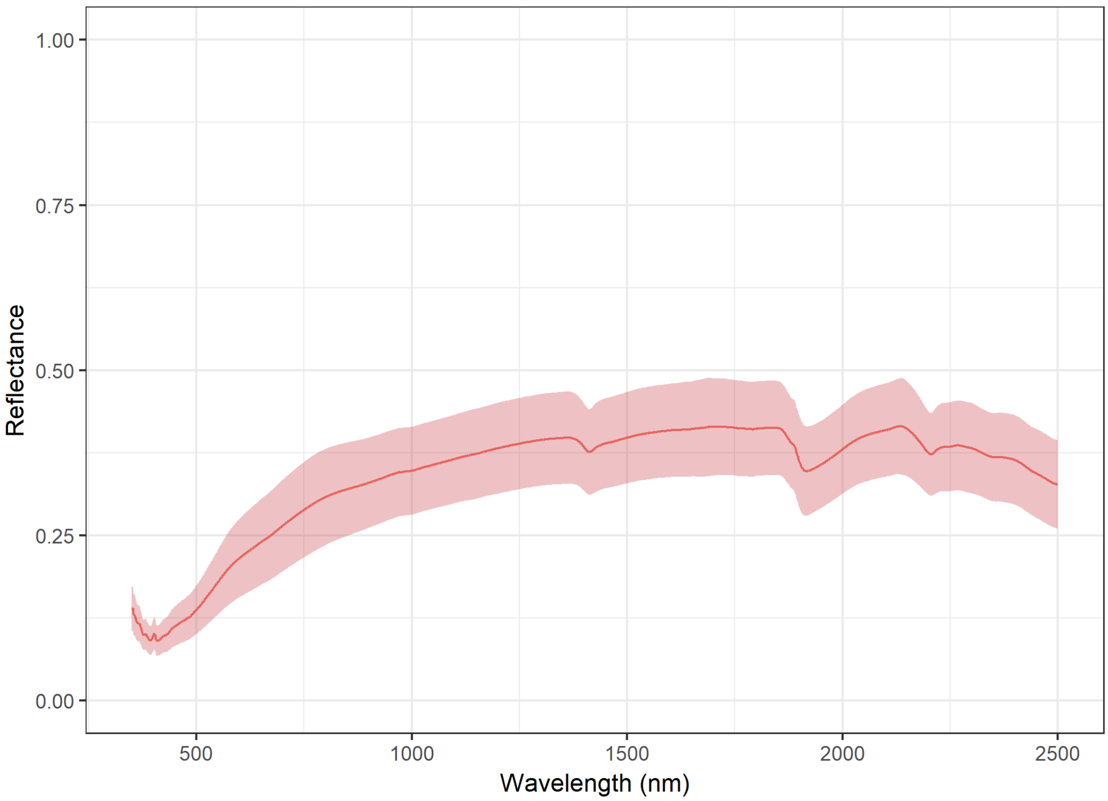

Figure 2-45. The standardized reflectance spectra of Serbia (mean ± standard deviation)

2.1.4.9. *Turkey*

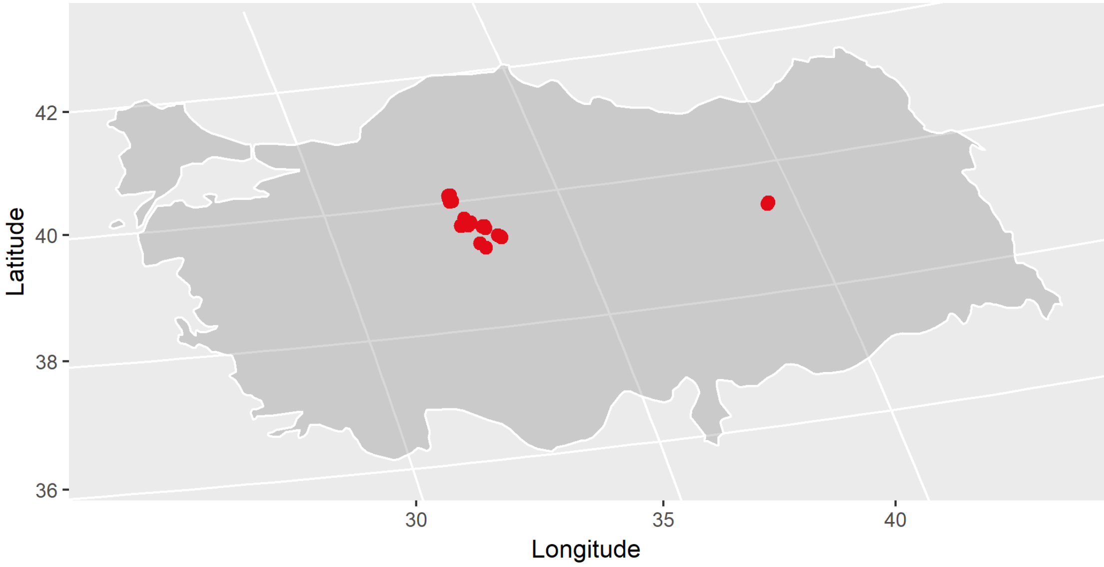

Figure 2-46. Location of the soil samples of Turkey

Table 2-9. Major statistical moments of the measured soil properties of Turkey

Property	Min	Mean	Median	Max	SD	Skew	Kurtosis	N
OM (%)	0.00	1.4545	1.26	5.09	1.13	1.002	0.735	94
Sand (%)	11.95	48.9943	50.57	86.20	19.64	-0.006	-1.122	98
Silt (%)	2.09	21.4671	19.90	47.78	9.10	0.881	0.502	98
Clay (%)	5.07	29.5386	25.78	76.46	15.98	0.644	-0.252	98
$CaCO_3$(%)	0.58	21.2726	18.48	89.99	17.86	1.568	2.989	100
pH (H20)	5.75	8.1471	8.17	9.76	0.58	-0.722	2.917	100
EC (μS)	0.02	1.6778	1.42	9.86	1.16	4.259	28.309	100

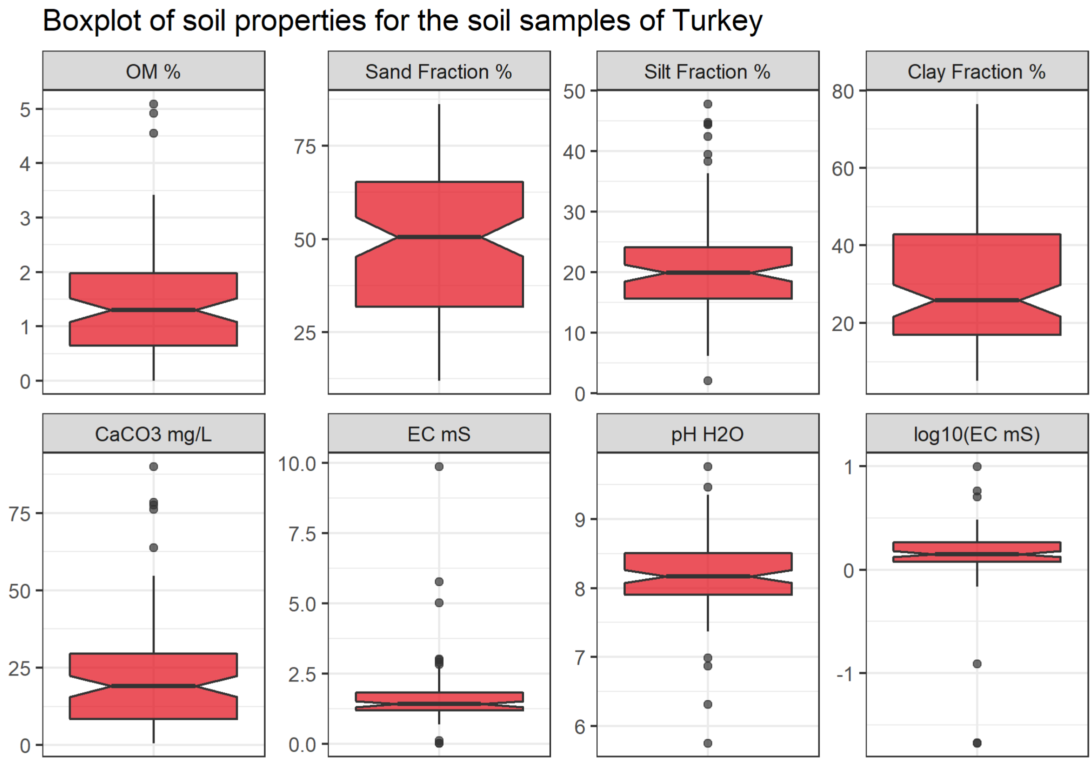

Figure 2-47. Boxplots of the measured soil properties for Turkey

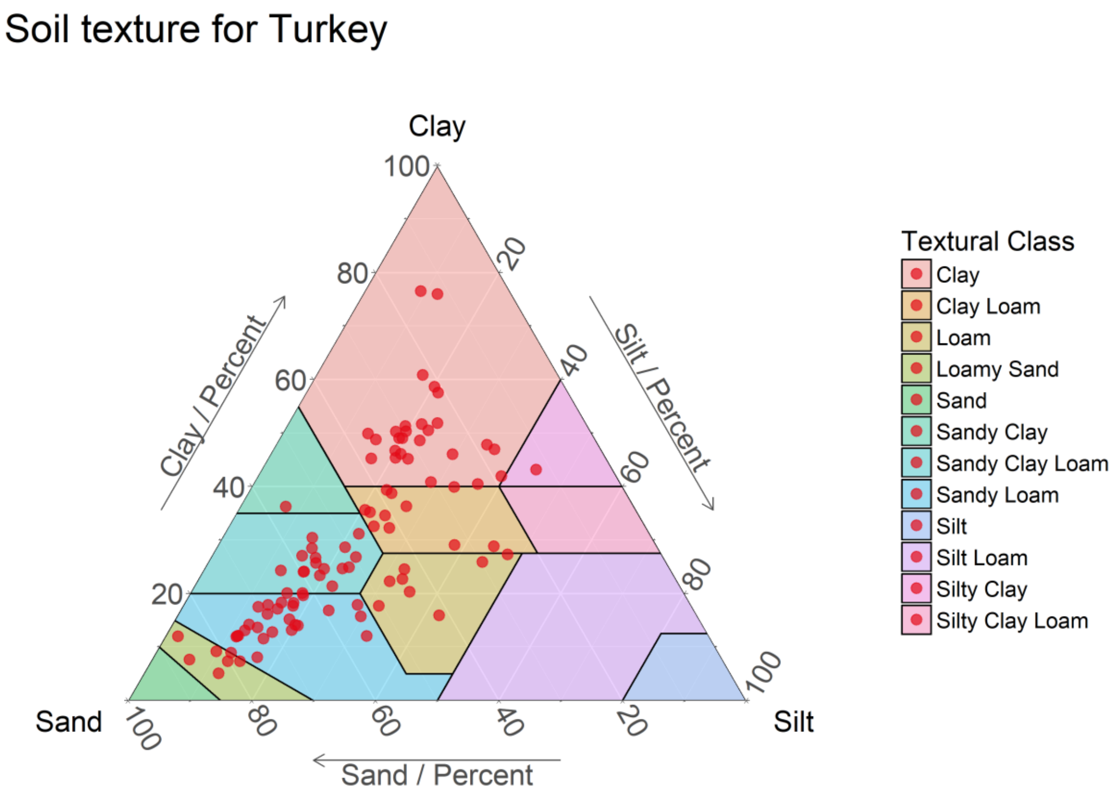

Figure 2-48. Distribution of soil texture class across all soil samples for Turkey

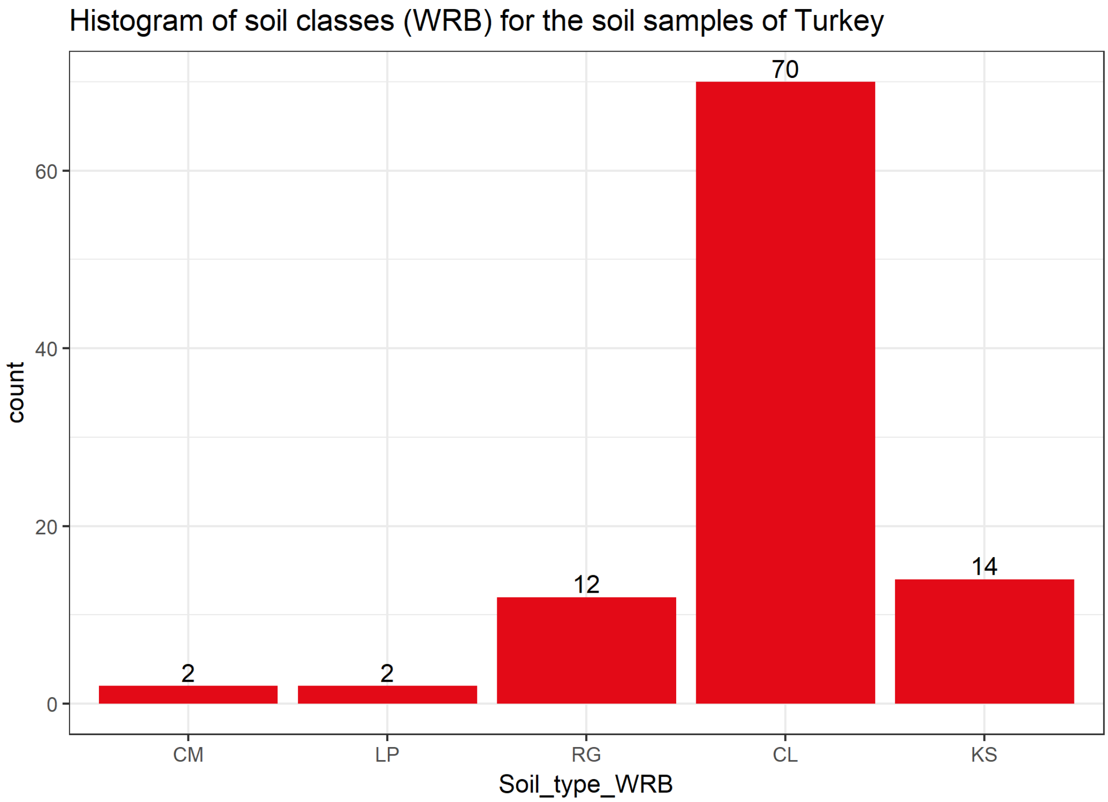

Figure 2-49. Histogram of soil classes of the soil samples of Turkey

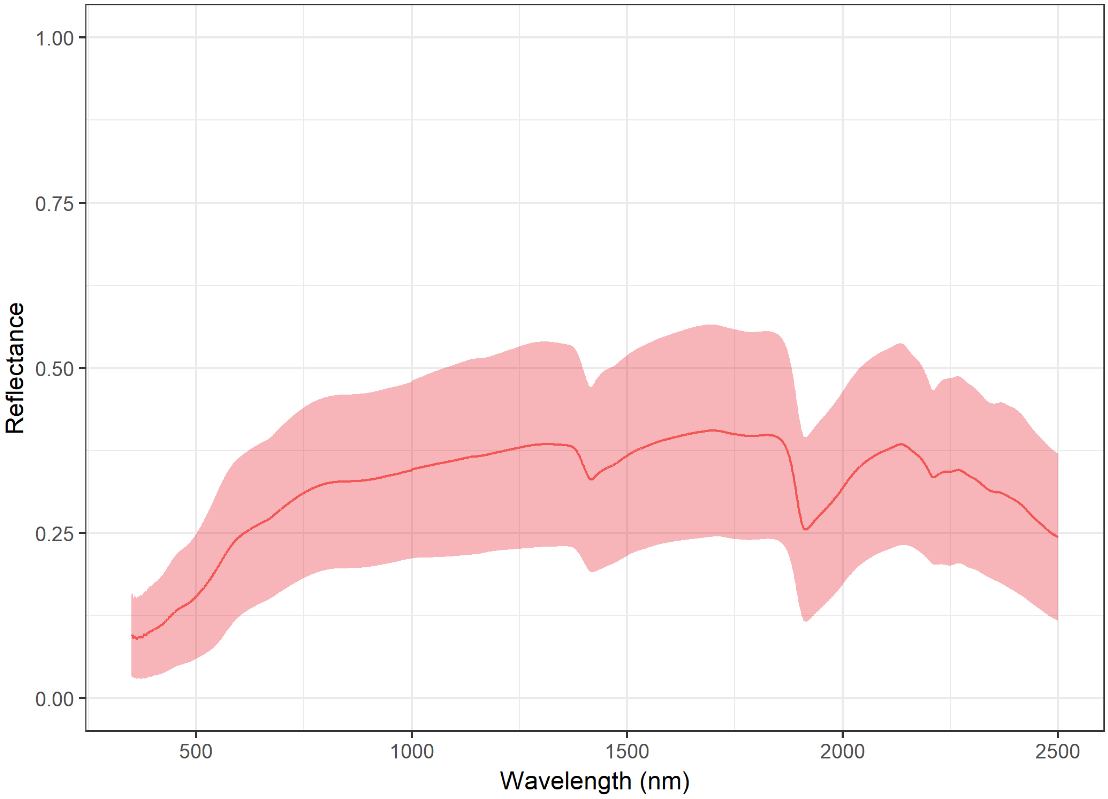

Figure 2-50. The standardized reflectance spectra of Turkey (mean ± standard deviation)

2.1.5. Integration with the DataHub

During the project meeting in Limassol, in addition to the soil sampling protocol given to all participants in this pilot activity, a template was decided and agreed upon w.r.t to the format of the file storing the soil spectral libraries. For each of the countries participating in the pilot, a Comma Separated Value (CSV) file was prepared, containing the same header names. The metadata for each sample are (bold letters denote the header name in the CSV file):

- Sample **ID**; a 15-character unique identifier of the format XX-YY-ZZZ-WWWWW, with:
  - XX being the country code
  - YY being the soil sample's class in WRB
  - ZZZ being the depth from which the sample was collected
  - WWWWW a unique id number for each sample – this number is zero-padded to ensure 5 characters exist for each sample
- GPS coordinated (**Latitude** / **Longitude**), **Elevation** (in meters), **Depth** of sample (in centimetres), **Sampling_date**
- Soil class in WRB and potentially in USDA (**Soil_type_WRB**, **Soil_type_USDA**)
- The climate of the area using the Köppen climate classification (**Climate_Koeppen**)
- Physical and chemical properties as follows: **OM, $CaCO_3$, Sand_Fraction, Silt_Fraction, Clay_Fraction, NO_3, EC**
- Standardized reflectance spectra (**X350-X2500**)

To assist and enable future researchers to use the GEO-CRADLE SSL, the resulting datasets were uploaded in the datahub in the following way:

- Per-country CSV files, where for each country its corresponding SSL as developed in the GEO-CRADLE project is contained; and
- A CSV file containing the complete GEO-CRADLE SSL (i.e. all the per-country csv files combined).

All the above may be found in the regional datahub at the following page (Figure 2-51):

http://datahub.geocradle.eu/dataset/regional-soil-spectral-library

Figure 2-51. The main page of the datahub concerning the GEO-CRADLE SSL

2.1.6. **Overview of the regional soil spectral library**

The regional SSL consists of a grand total of 1754 soil samples with standardized reflectance spectra, their analytical measurements for their physical and chemical properties, and their associated metadata. All samples do not have the same properties measured; therefore, there exists some inherent sparsity in the data. This is demonstrated in Table 2-10, which depicts for each soil property the number of samples analysed per each country. The most predominant ones are the soil texture and SOM, both equally important for agriculture and the management of water extremes.

This regional SSL was developed in a region of the world that was underrepresented in other contemporary SSLs. In particular, taking into consideration the closed Global SSL, and the open LUCAS SSL, the former covers very sparsely the region while the latter covers only Greece (Figure 2-52). In terms of absolute numbers, the impact of this SSL is to increase by roughly 400% the openly available soil spectra data for the region. Moreover, because this SSL was built using a standardization protocol, it is readily expandable and research organizations from the region may easily contribute to this database.

Notch boxplots of the soil properties may be found in Figure 2-53. This Figure suggests that there is large diversity among the soils contained in the SSL, allowing models to be developed which can adequately generalize. A pairwise correlation plot among all physical and chemical soil properties present in the SSL is presented in Figure 2-54 which can be used to draw a first-order quality control of the data against pedological criteria, although given the sparsity of the data it is precarious to extrapolate further. In particular, within the SSL there are soils with different texture and of different soil type (Figure 2-56). A great variety of climatic, topographical and geological conditions, together with the diverse anthropogenic influences has resulted in a diverse soil cover in the RoI. A testament to this is that the samples represent 18 out of the 32 reference soil groups of WRB. The most predominant soil classes are (in decreasing frequency): a) the Fluvisols, b) the Leptosols, c) the Cambisols, d) the Luvisols, e) the Vertisols, and f) the Arenosols. This is in compliance with the most prominent soil classes which can be found in the region. The textural classes are also well represented (Figure 2-55), albeit no pure Silt samples exist, and there exists a slight trend towards sandy soils (because of the large presence of Sandy soils from Greece).

As far as the spectral signatures are concerned (Figure 2-57), indicate a large soil mineralogical variation in the SSL because of the diverse soil-forming environments of the sample sites. Moreover, the near infra-red region exhibits well-defined absorption features close to 1414 and 1915 nm, which are assigned to OH-soil hygroscopic water in clay minerals. The main spectral differences may be ascribed to absorption bands in the visual range related to iron oxides, and SOM or carbon.

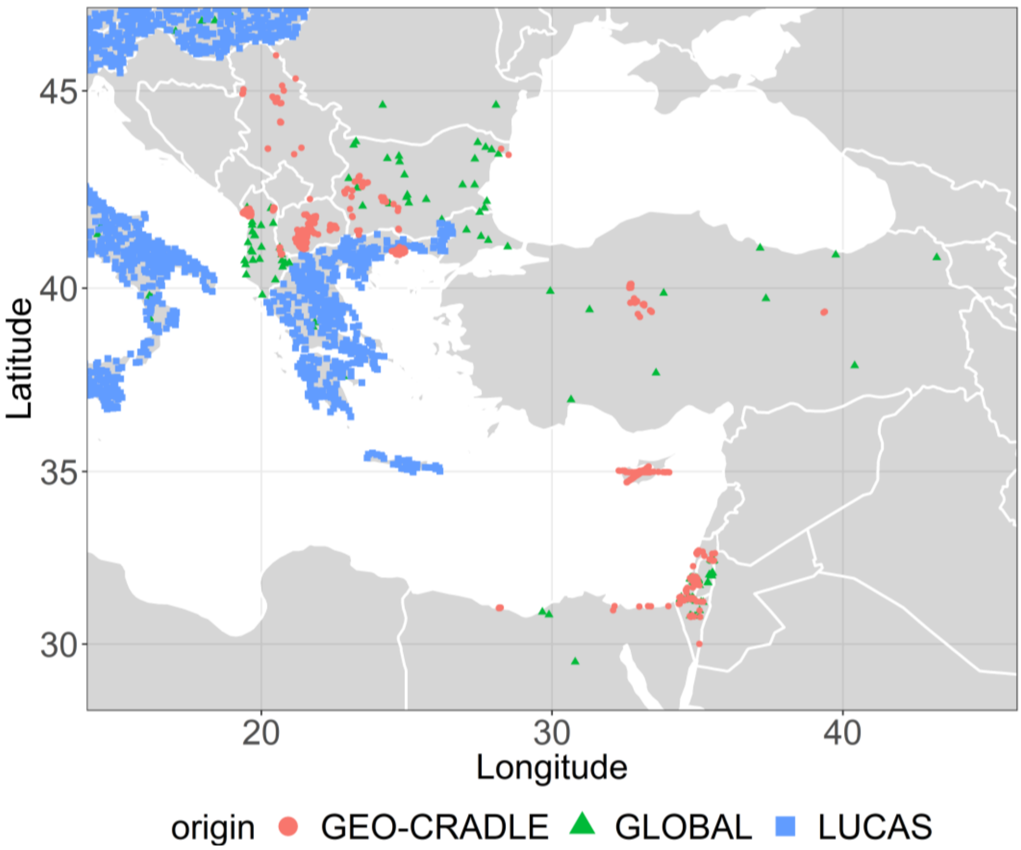

Figure 2-52. Comparison of the samples' location with the Global SSL and the LUCAS SSL

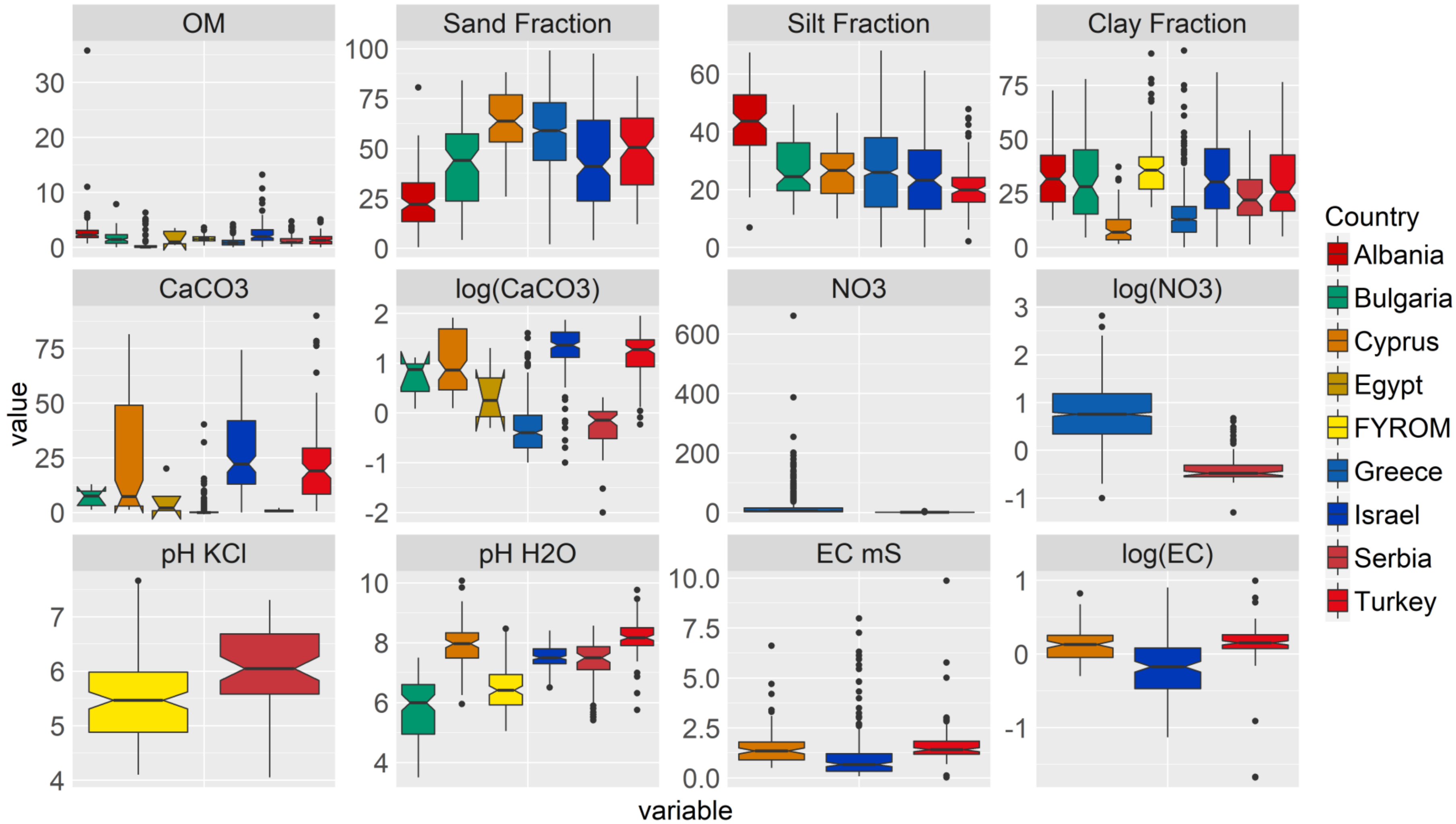

Figure 2-53. Boxplot of all soil properties per each country in the GEO-CRADLE SSL

Table 2-10. Overview of the analysed samples in the regional SSL per country

Country	Samples	OM	Texture	$CaCO_3$	pH	NO_3	EC	CEC
Albania	107	107	107	0	0	0	0	0
Bulgaria	105	105	105	0	105	0	0	105
Cyprus	96	96	94	96	96	0	93	0
Egypt	10	6	0	4	6	0	6	0
FYROM	124	124	124	0	124	0	0	0
Greece	928	928	928	928	0	928	0	0
Israel	221	113	168	169	136	0	141	0
Serbia	63	63	63	63	63	63	0	0
Turkey	100	94	98	100	100	0	100	0
All	1754	1636	1687	1360	630	991	340	105

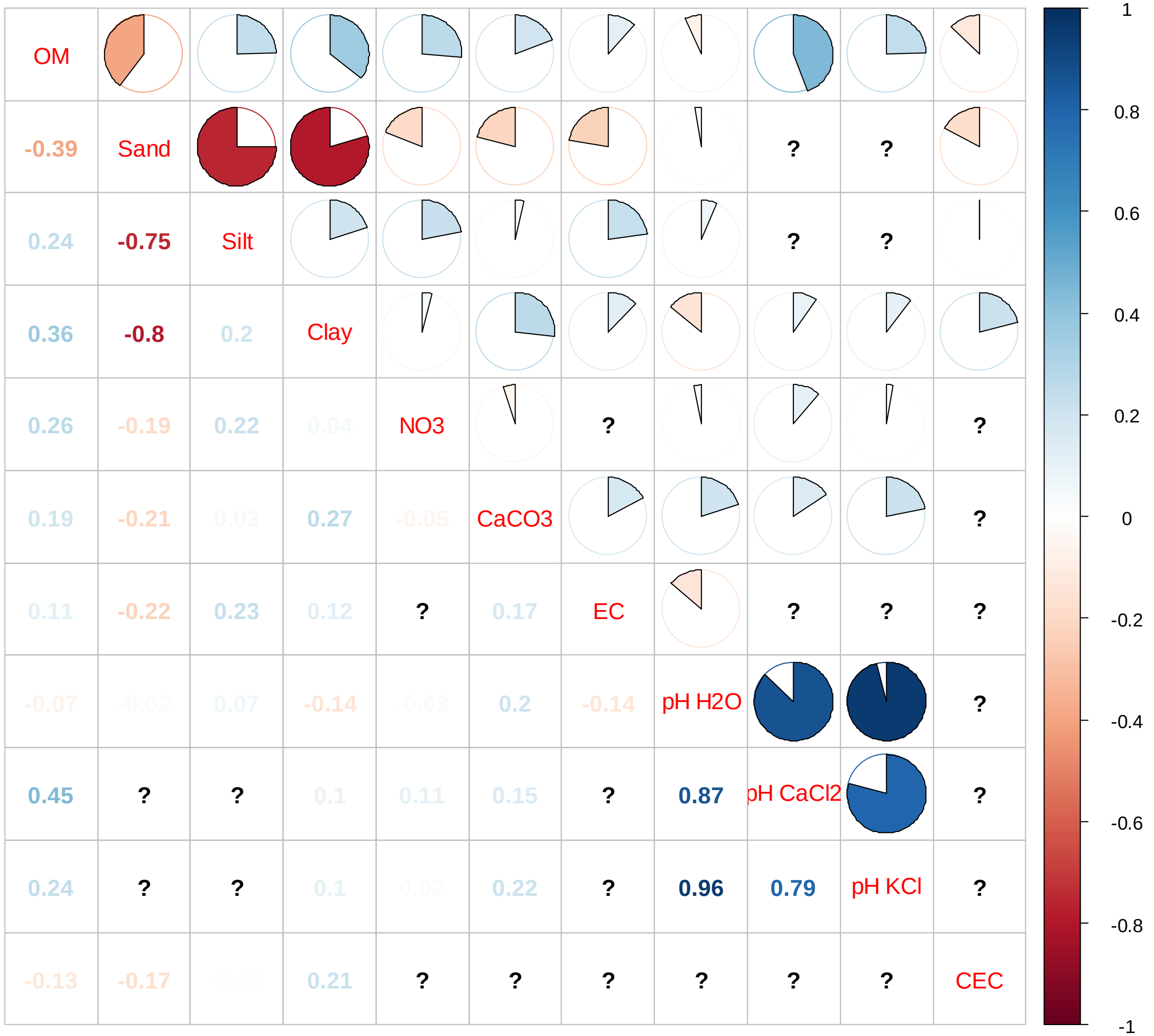

Figure 2-54. Pair-wise correlation plot of the soil properties within the GEO-CRADLE SSL

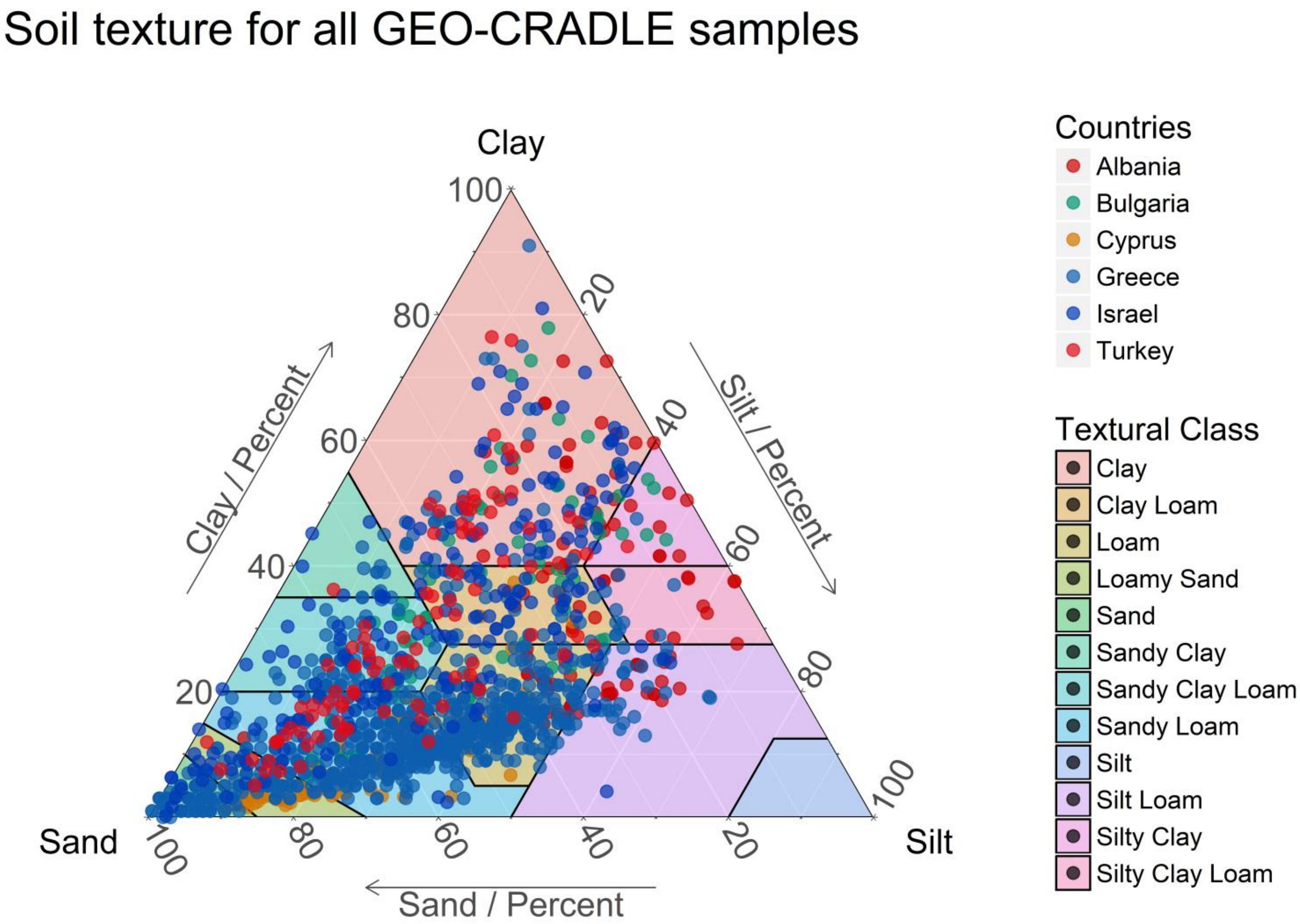

Figure 2-55. The soil textural class distribution for all the samples in the GEO-CRADLE SSL

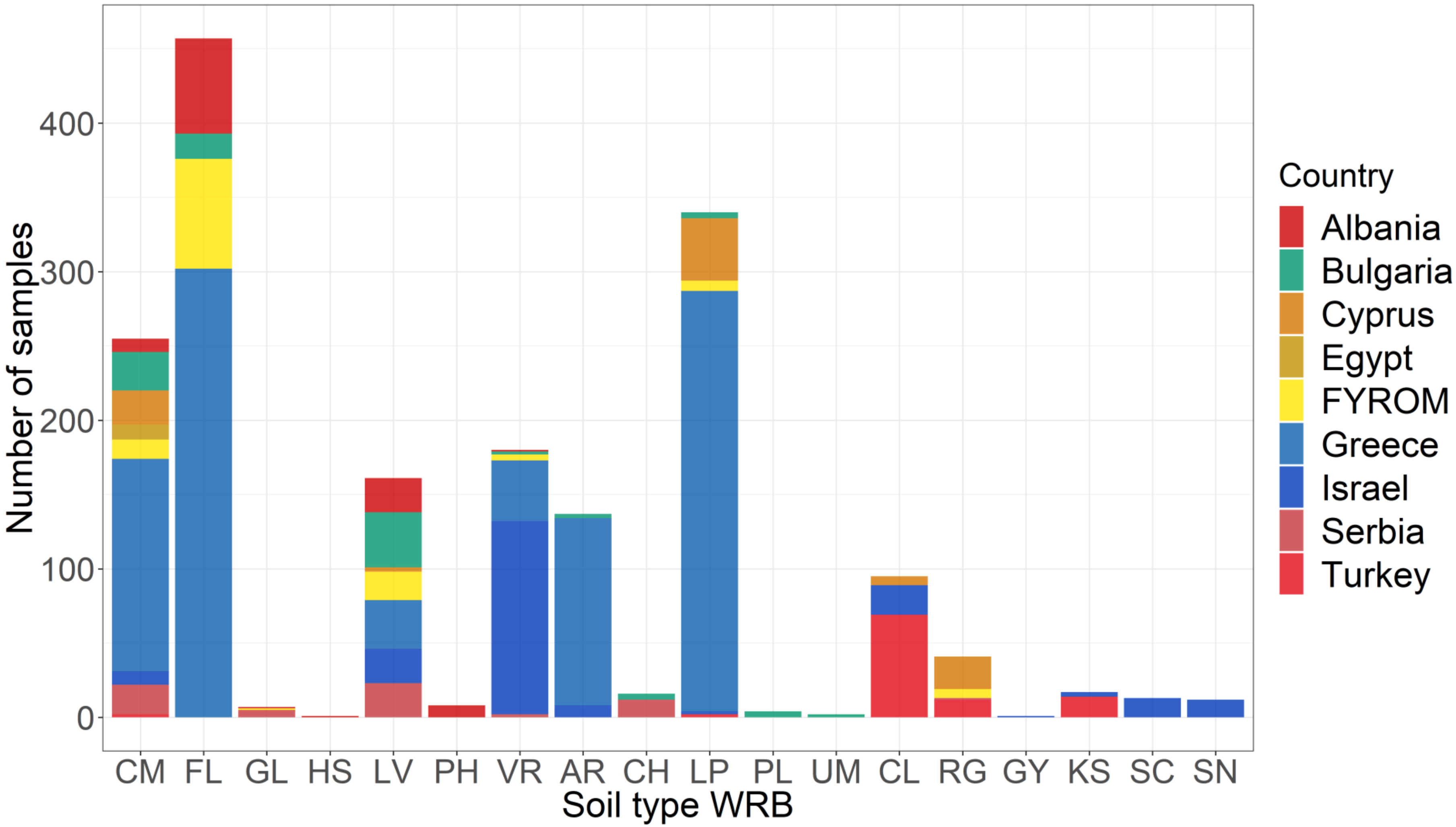

Figure 2-56. Histogram of soil classes of the soil samples present in the GEO-CRADLE SSL

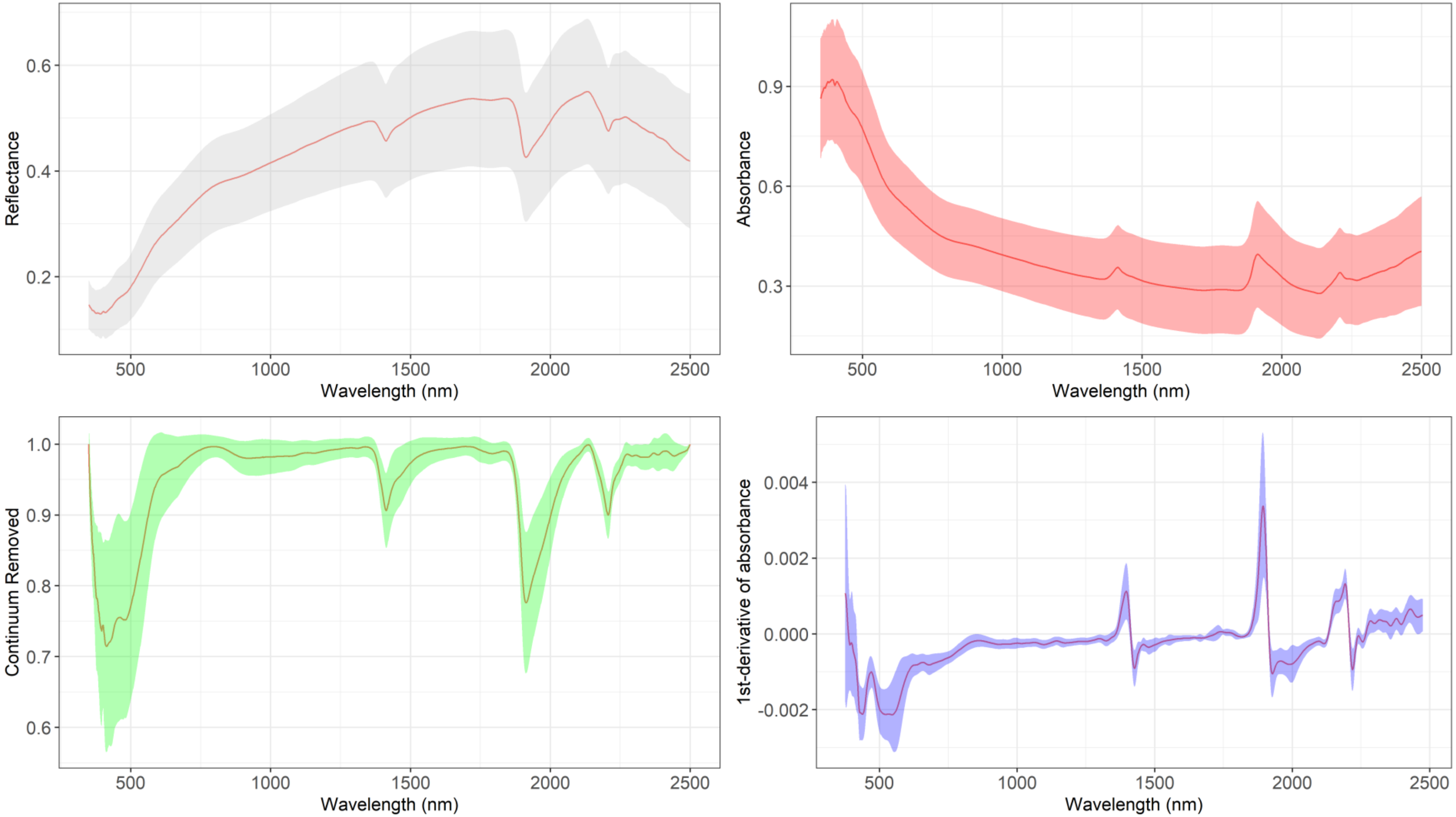

Figure 2-57. Spectral signatures of the soil samples in the GEO-CRADLE SSL (mean ± standard deviation)

2.2. *Implementation of the myDEWETRA platform*

2.2.1. **The myDEWETRA platform**

The MyDEWETRA platform implementation in the IFS-WEM pilot intends to support the main issue of the same pilot; that is, to improve agricultural management and assist the decision process providing efficient instruments for water extreme management.

On this line, MyDEWETRA makes in action a collection, systematization, and visualization of various kinds of heterogenous data and model outputs, either automatically or manually recorded, in the MyDEWETRA GEO server, allowing their combination and display on the same web based interface producing added value.

Two levels of MyDEWETRA implementation are carried out for the IFS-WEM pilot.

The **first level** is a Regional scale implementation proposed for the entire ROI. It contains a set of open source data, which form the minimum dataset to perform demonstrative predictions of risk of flooding and drought scenarios.

In particular, the data available on MyDEWETRA are related to the weather forecast model outputs global scale (e.g. GFS), land use/land cover maps, satellite based rainfall observation (e.g. GPM, TRMM), global scale flood risk hazard (e.g. GAR2015 hazard maps) and global scale drought index (SPI, SPEI).

The **second level** of MyDEWETRA implementation is centered on the Basin scale. The area selected as "test-case" has been the Drin-Buna river basin in Albania (sse Section 2.5).

All data on MyDEWETRA are accessible through the platform and the GEO-CRADLE Data HUB.

2.2.2. **Data infrastructure of MyDEWETRA platform**

MyDEWETRA is a web-portal of the Italian Civil Protection Department developed by CIMA Research Foundation based on the existing Dewetra platform, which consists of an integrated real-time system for hydro-meteorological and wildfire risk forecasting, monitoring and prevention based on the rapid availability of geospatial data among multiple relevant institutional stakeholders. It improves the accessibility and comparability of hazard, exposure and risk information and data at multiple levels.

The platform MyDEWETRA is a web-portal aimed at data visualization from different sources, ensuring interoperability with already existing webservices, and complying with main relevant international standards. As a web-portal, it provides access possibly to several applications that can be easily extended. In other words, it provides interfaces for both human-to-machine as well as machine-to-machine communication.

One of the main issues of MyDEWETRA platform is to allow to the data from forecast models, remote and ground observations to be integrated with data from the territorial elements (exposures) as well as data from operators and citizens to analyse the situation in real time and deferred time to foresee possible future scenarios.

The main application is ***Dewetra*** (***current version 2.0***), a web-GIS platform aimed at multi-risk mapping and hydrological forecasting and monitoring; it is an integrated system for real-time monitoring, prediction and prevention of natural disasters worldwide. Owned by the Italian Department of Civil Protection, the system has been developed by CIMA Research Foundation with Open Source technologies. It is compliant with the most common and widespread European and International standards, and this portal is currently developed only for the management of static information.

Since 2012, Dewetra is promoted by the Commission of Hydrology of the World Meteorological Organization as a system for improving flood forecasting and warning.

One of the key requirements for facilitating data sharing in a national or regional context is to provide solutions that ensure accessibility to information without the physical transfer from data owners/producers.

The portal has been developed using Open Source technologies and following the OGC and Inspire standard for the sharing of data. The highest-level architecture based on the latest technologies can be represented by the scheme reported in *Figure 2-58***Error! Reference source not found.**.

The server side uses components written in Java and Python language. The services are exposed as APIs Rest and provide the functionality to navigate the application and access the data. On the client side are used HTML 5.0, Angular JS and Bootstrap to allow the user to interact with the REST services of the Server. On the server side, one of the most interesting features is the management of the push messages sent by the server to connect clients: that virtually cancel the latency between the time when the data are available in MyDEWETRA and when can be displayed on the client side (*Figure 2-59*).

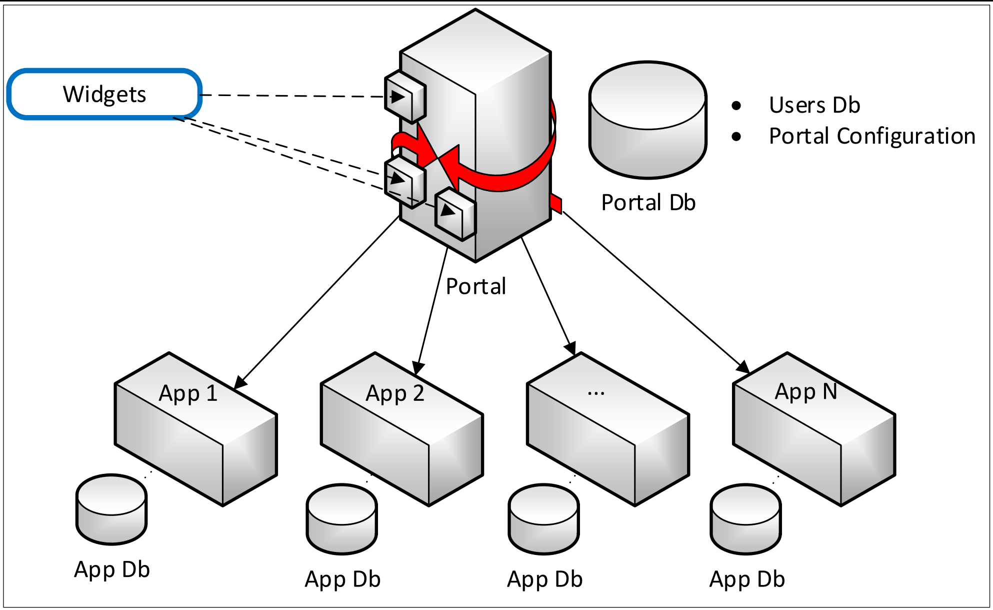

Figure 2-58. MyDEWETRA Portal Schema.

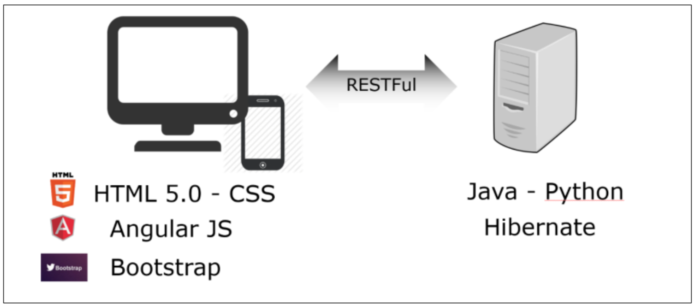

Figure 2-59. MyDEWETRA System architecture.

The share of data and information among institutions and agencies can be achieved with different configurations of data infrastructures following several data sharing policies (*Figure 2-60***Error! Reference source not found.**). With a centralized infrastructure, data are physically transferred to a central data server where MyDEWETRA platform is installed, to allow the visualization of data and products to end-users. The physical transfer of data can be a barrier for an effective data sharing policy, because institutions are much more inclined to allow a regulated access to data rather than transfer them. A distributed data infrastructure is based on the principle of accessibility to data and products (INSPIRE Directive): data owners and/or producers store, update and ensure accessibility to data without physically transferring them to third parts. In a distributed infrastructure, MyDEWETRA is composed by several nodes (DEWETRA Data Servers - DDS) installed in each institution; the portal allows end-users to visualize and process data in a unique environment.

Due to the few amounts of data provided by the partners, it has been decided to realize a centralized infrastructure version of MyDEWETRA for the WP4.2.

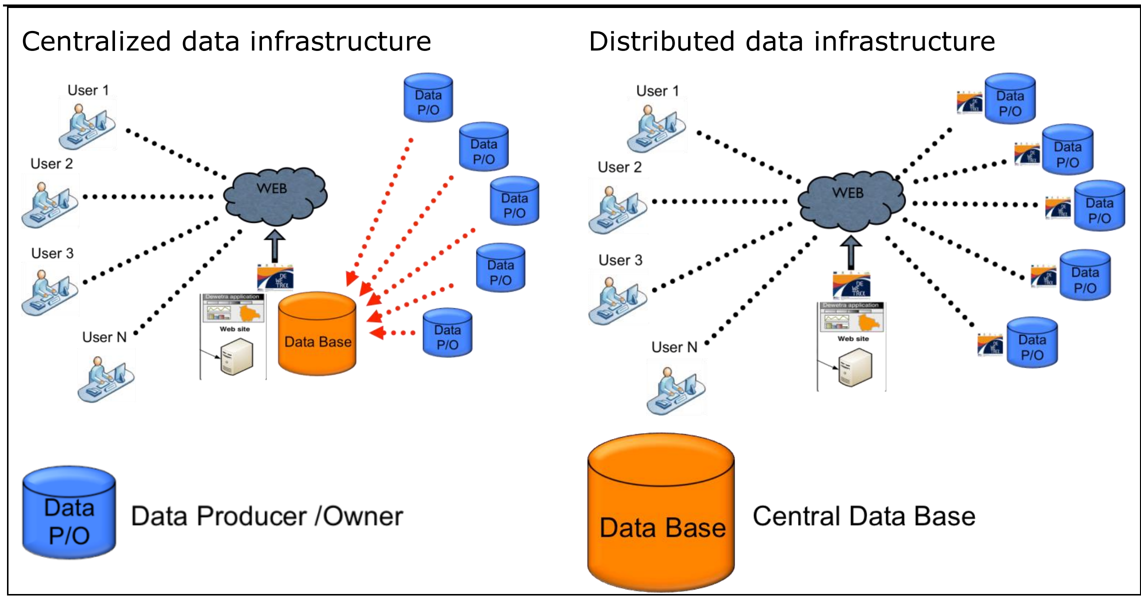

Figure 2-60. MyDEWETRA centralized and decentralized data infrastructure.

2.2.3. **Content of MyDEWETRA**

MyDEWETRA has two main components:
1. MyDEWETRA web application
2. MyDEWETRA data server

The web application is the core component of the system and ensures the accessibility to real time data and maps. In the case of the MyDEWETRA decentralized data infrastructure, the data servers are installed in each institution that is engaged in sharing data among the network. In this case the data servers used is provided from CIMA.

End users can access the web application from any personal computer connected to the Internet: to access the platform users may visit the url:

http://geocradle.mydewetra.org

and then press "Free Entrance". *Figure 2-61* shows a screenshot of the main page (panel A) as currently structured, with the application logo that stands as the background and the credits of the same reported in Italian and English. On the top of the main page the link to the ***MANUAL*** of MyDEWETRA is available together with a brief description of GEO-CRADLE project (***DESCRIPTION***-panel B), the objectives (***OBJECTIVES***-panel C) and the involved partners (***TEAM***- panel C).

Once the access through the button "***FREE ENTRANCE***" is accomplished, the user is in the "***DASHBOARD***" (*Figure 2-62*) which provides a quick check of some of the data available in ***Dewetra 2.0***. More precisely, the dashboard shows some snapshots on the two main categories of data treated into WP4.2, i.e. Floods and Drought:

- in the first row are visualized 4 panels which show the GAR 2015 - Flood Hazard Maps on two different area of interest: two are centred on the Balkans region and two on the MENA and all 4 show hazard maps are referred to return periods t=50 and 100 years;

- in the second row are visualized 4 panels which refer to Standardized Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) with two different accumulation periods (6 and 12 months).

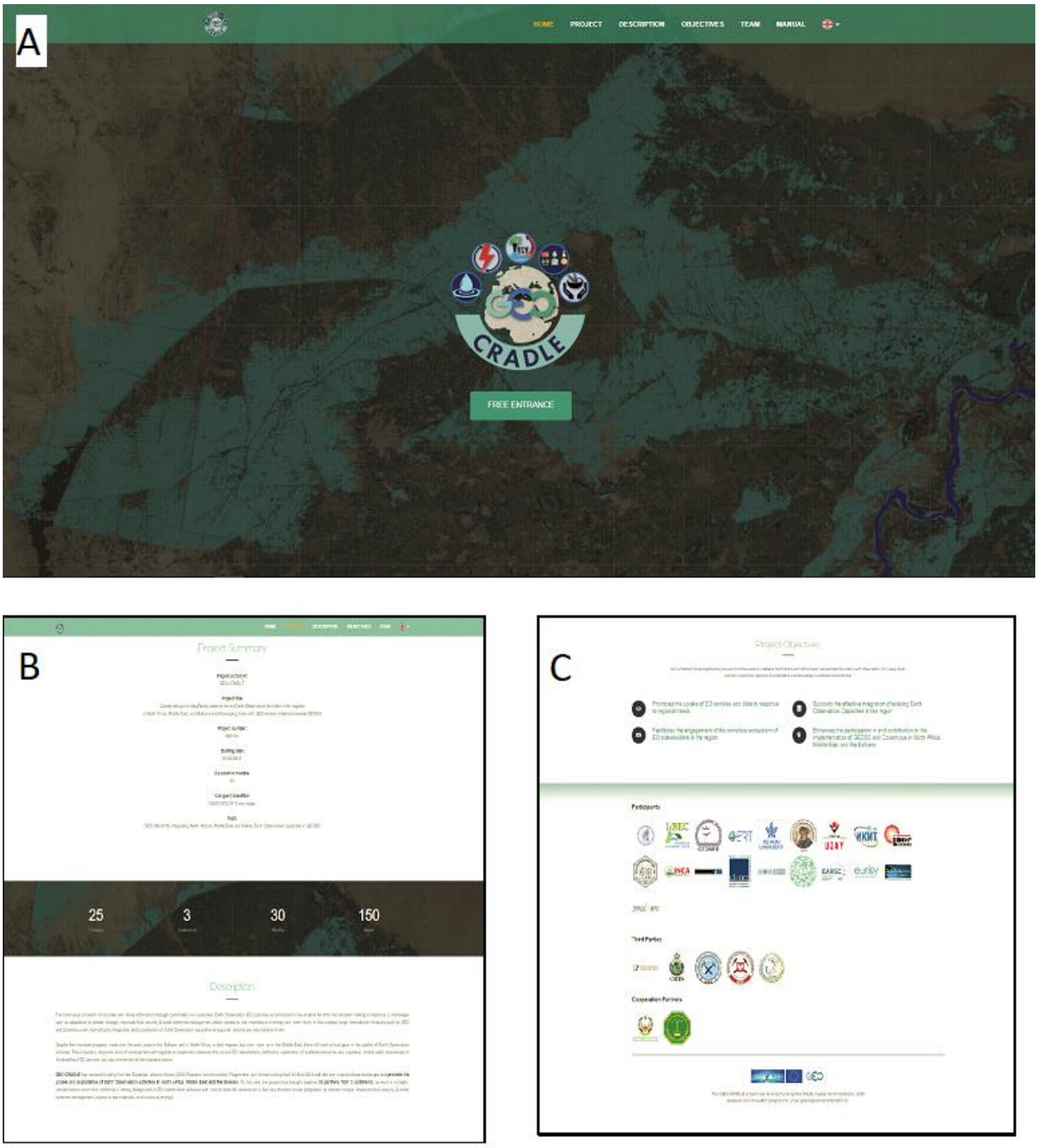

Figure 2-61. MyDEWETRA main access page and the informative public section.

From the"***DASHBOARD***" (*Figure 2-62*), to enter ***Dewetra 2.0*** users have two possibilities. They may click either:

1) on the icon in the displayed widgets regarding some of the available ***Dewetra 2.0*** layers (e.g., FLOODS or DROUGHT maps)

2) on the icon in the sidebar - upper left of the screen.

The Web-GIS platform included into MyDEWETRA uptakes all the functionalities of Dewetra involved with the static themes, geospatial layers are classified in compliancy with INSPIRE Directive and EU Floods Directive. The interface has been specifically designed to enhance navigation and research of information.

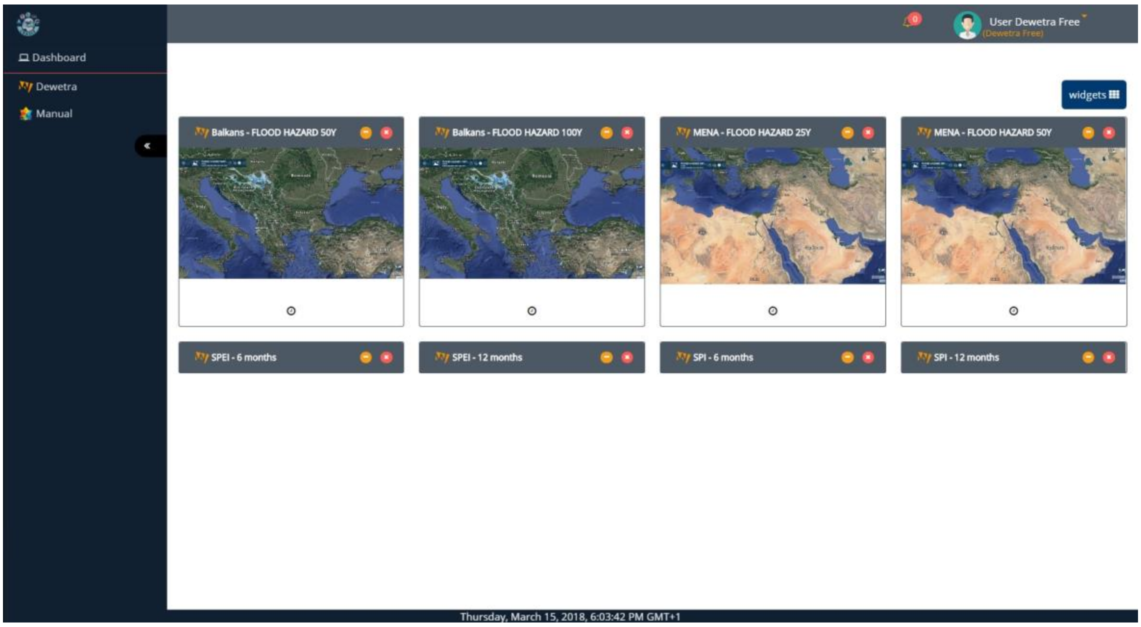

Figure 2-62. MyDEWETRA DASHBOARD with widgets representing some of the FLOODS and DROUGHT maps available in DEWETRA 2.0

2.2.4. **MyDEWETRA User Interface**

Dewetra 2.0 is the evolution of the former integrated system Dewetra, operational at the Central Functional Centre of the National Civil Protection Department (DPC) /o Prime Minister's Office since 2008, for the forecasting, monitoring and real-time surveillance of all the environmental risks. The application provides, through a graphical interface, high-resolution and continuously updated information, allowing the user to monitor -for example- weather events, to build detailed risk scenarios and evaluate the potential impact of the phenomena on communities and infrastructure.

One time the user clicks on the icon "**W**" on the orange background in the sidebar-upper left of the screen, the ***Control Map*** is opened. The ***Control Map*** is instantiated as the system is started, using the Google Hybrid map provided by Google-Maps services as the background layer. The ***Control Map*** is one of the components of the ***User Interface*** together with:

- ***Toolbar***
- ***Display***

- ***Layer List***
- ***Additional Tools***

as shown below in *Figure 2-63*.

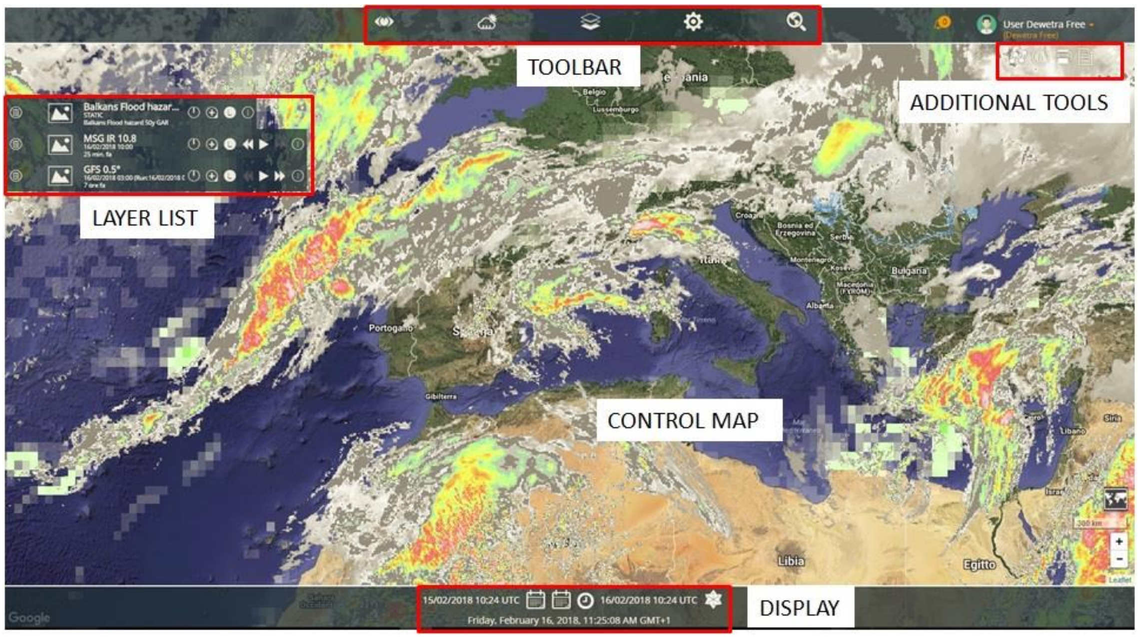

Figure 2-63. MyDEWETRA User Interface

Data are generally organized into three main categories: observations, forecast models and static layers. Each category is further structured in **tags** (a thematic classification, i.e. rain, thunderstorms, soil moisture etc.) and/or **folders** (by means of which the data are stored separately depending on their source: radar, satellite, weather stations etc.).

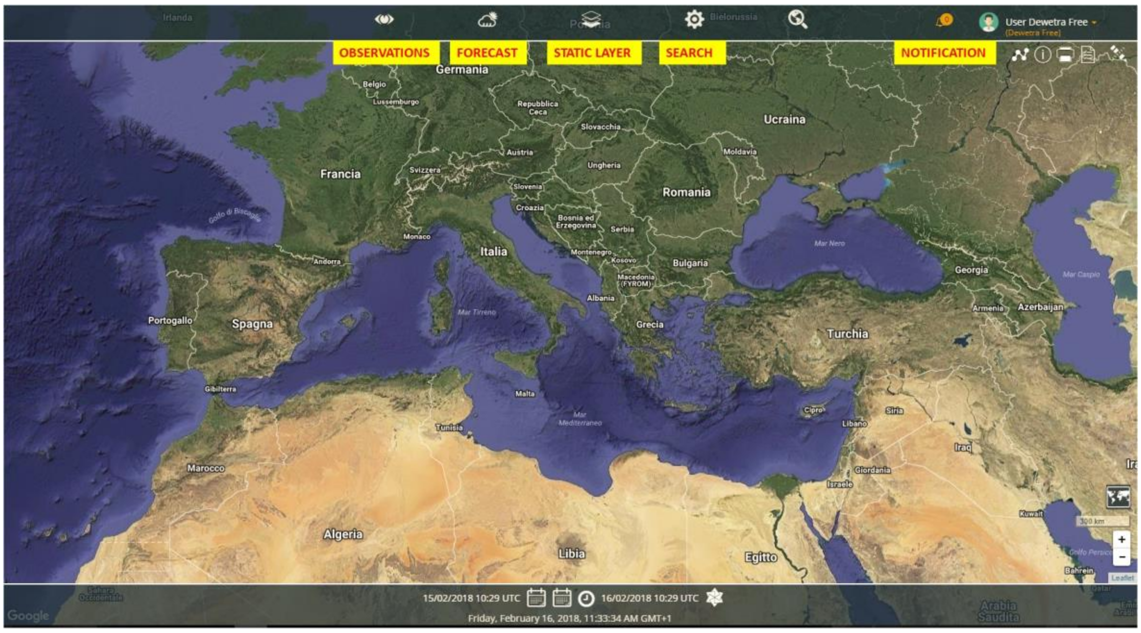

Figure 2-64. Control Map and Toolbar action buttons

On the top of the ***Control Map***, the ***Toolbar*** (*Figure 2-64*) contains the following action buttons:

- *Observations*: is the section dedicated to observational data and diagnostic models related to a time series. The data in this section can come from weather station, radar, satellite imagery but also webcams and radio gps. This data can be represented in their native resolution, but it also possible some elaboration to make the information more useful for an observer.
- *Forecast Models*: lists all the available outputs from forecast systems (numerical weather prediction models, hydrological models, landslide susceptibility models, fire models, etc.);
- *Static Layers*: provides all the information needed to design a comprehensive risk scenario such as the exposures or the hazard maps. Static Layer are made of data from the territory collected to know as many useful information from the territory as possible. They are defined "static" because they don't change frequently in time but need only to be checked and updated sometimes.
- *Tools*: enables some ancillary functions such as Export, and Report Scenarios
- *Search*: is the tool allowing the users to search for any element visualized by the platform such as weather stations, toponyms, etc.
- *Notifications*: this area is dedicated to display the messages delivered by other users of the portal.

In the next pages are reported in detail the data implemented in the sections named Observations, Forecast Models and Static Layers in the GEO-CRADLE MyDEWETRA release.

Every time a layer is pulled on, the application displays it on the ***Control Map*** and the Layer List (top left of the screen- *Figure 2-65*) allowing the user to visualize

- name of the layer
- layer description
- reference date
- initialization time of the run (if the chosen layer is a model's output)
- Spatial aggregation (if enabled)

Figure 2-65. Layer lists example

Users are enabled to manage the layer they pulled on by selecting one of the buttons available on the layer list:

- **Turn on / Turn off** visualizes / deletes a previously loaded layer
- **Zoom to Layer** allows the user to bring the zoom back to the default level for that layer
- **Legend** displays the pop-up window showing the legend for the selected layer

- **Scroll** offers the user the option to display a set of successive time steps of a given variable within the time range set in the Display)
- **Expand** allows displaying the buttons that activate some additional functions such as:
  - the slider adjusts the transparency / opacity of each layer
  - the download button saves to user's own disk any chosen layer, in many formats
  - the refresh button updates the layer, by loading the latest version from the available
  - the metadata button that allows the user to view/download the metadata file associated to the layer

2.2.4.1. *Observed data in MyDEWETRA platform*

The first action button of the Toolbar is the Observations menu, which provides access to all the observational datasets to the user. Once the user clicks on the menu, two different views are offered; namely, the Tag and Folder mode.

In the framework of the WP4.2, the Tag mode in the release of MyDEWETRA shows the following observational data (Figure 2-66. Open observation data available in MyDEWETRAFigure 2-66) as organized by thematic criteria:

- RAINFALL
- SOIL SAMPLES
- CLOUD COVER
- DROUGHT

In the table below the layers of this category are shown according to the assigned tags.

Table 2-11. List of observed data layers available in the MyDEWETRA platform

TAG	LAYER NAME	SOURCE
RAINFALL	GSMaP	NASA-JAXA
	IMERG 3h ACC.	GPM
	PR OBS 5 -H05 (HSAF)	Italian Civil Protection Department (HSAF)
SOIL SAMPLES	SOIL ANALYSIS	GEO-CRADLE partners
CLOUD COVER	MSG IR 10.8	Italian Civil Protection Department
DROUGHT	SPI_IRI_3M	U. S. Climate Prediction Center (U. S. Climate Prediction Center Gauge - OLR Blended daily precipitation analysis)
	SPI_IRI_6M	
	SPI_IRI_9M	
	SPI_IRI_12M	

	SPEI_3M SPEI_6M SPEI_9M SPEI_12M	Instituto Pirenaico de Ecología and Estación Experimental de Aula Dei —Zaragoza, Spain

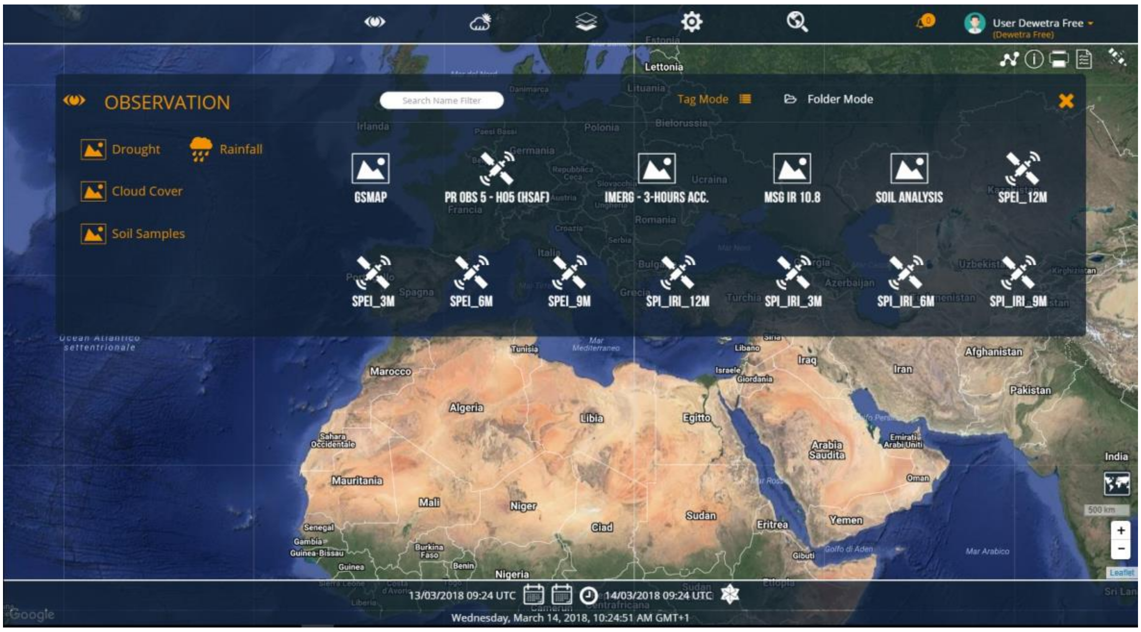

Figure 2-66. Open observation data available in MyDEWETRA

Every time a layer is pulled on, the application uploads it in the ***Control Map.*** If the cursor is left on the name of the layer in the ***Layer List*** (top left of the screen) the user enables the tooltip function to open two windows:

- the first one at the top right of the screen (Figure 2-67 - panel B) which shows the metadata of the layers, such as:
  - the name of the layer
  - the Layer description
  - the reference date
  - the initialization time of the run (if the selected layer is a model's output)
  - the Spatial aggregation (if enabled)
  - the validity interval (if it is a combined variable)
- the second one is placed immediately to the right of the ***Layer* List** (Figure 2-67 - panel A) and shows the name of the layer, the date of initialization of the run (if the case) and/or the reference date.

In the given example in Figure 2-67, the tooltip displays the available information about one of the available SPI maps.

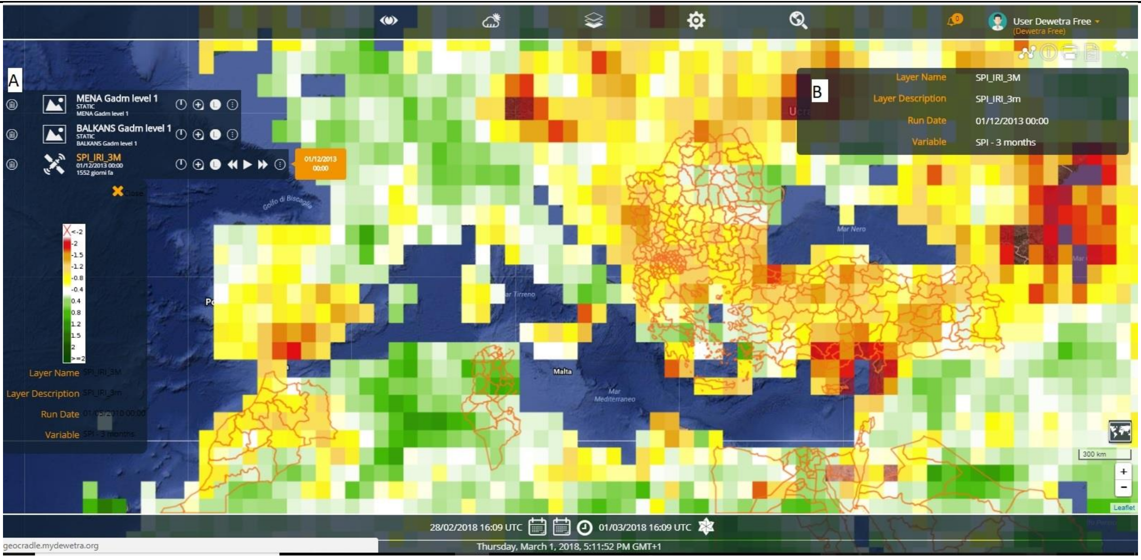

Figure 2-67. Example of one of the observational data available: 3-Month Standardized Precipitation Index (SPI_IRI_3M) is plotted together with the administrative limits for Balkans and MENA regions

SOIL SAMPLES

The results of the analysis realized on the data collected by the soil samples in the Balkans and the MENA regions described in detail in 2.1 have been implemented in the section **Observations**- **Soil Samples** under the name **Soil Analysis** (Figure 2-68 – panel A).

Clicking on **Soil Analysis**, a default set (dot points in Figure 2-68 – panel B) of the soil samples appear on the **Control Map**. If the user moves the cursor on the left on the name of the layer (Figure 2-68 – label B1) in the **Layer List,** a panel with the metadata (Layer name, Depth [cm], Variable, Date) of the visualized default data appears (Figure 2-68**Error! Reference source not found.** – label B2).

If the user moves the cursor on one of the default dots (red cross Figure 2-68 - panel C) visualized on the **Control Map**, on the upper right corner a panel appears with these information (Figure 2-68 - label C1):

- Date: the date of the soil sample;
- Elevation [m]: elevation at which the soil sample has done;
- Depth [cm]: depth at which the soil sample has done;
- USDA: USDA soil taxonomy developed by United States Department of Agriculture and the National Cooperative Soil Survey provides an elaborate classification of soil types
- WRB: World Reference Base classification system for naming soils (Figure 2-68)
- CLIMATE: Köppen climate classification of the soil sample
- Country: country of the Roi where the soil sample has done
- Region: region of the Roi where the soil sample has done (Balkans or MENA)
- OM [%]: organic matter concentration in the soil sample (Figure 2-68).

Clicking on one of the default dots, two new panels are available:

- The first panel contains the soil spectrum graph of the selected sample (Figure 2-68 - panel D);
- The second panel contains the results obtained by the chemical analysis (Figure 2-68 - panel E) made on the selected sample.

The data in these two panels can be downloaded by the user using appropriate buttons (Figure 2-68 - labels D1 and E1)

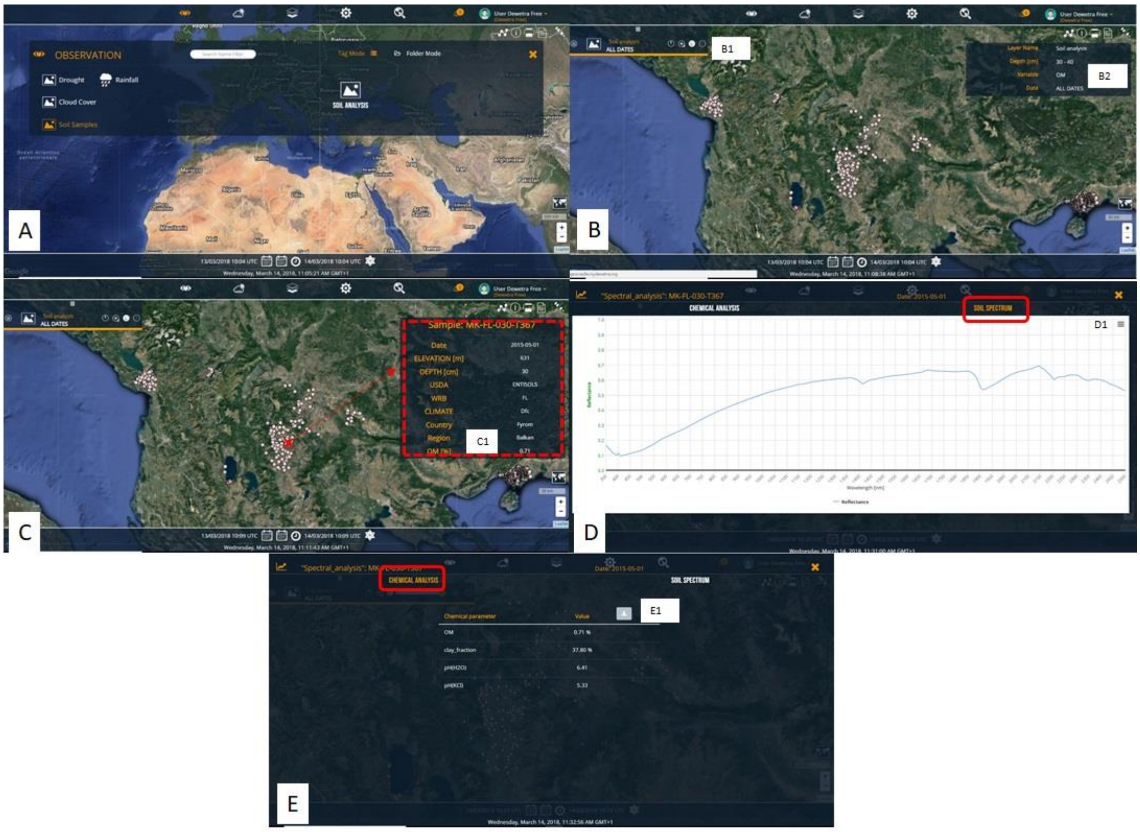

Figure 2-68. Soil Analysis layer and visualization of the results of the soil sample analysis on MyDEWTRA

To visualize all the other soil samples and the relative analysis, the user should move the cursor on the left on the name of the layer (Figure 2-68–label B1) in the **Layer List** and click on it.

At that point, a new screen pops up (Figure 2-69 – panel B) and the user has the possibility to combine different properties and consequently the soil samples which fit all the properties selected.

The properties on which makes different combinations are:

- AREA OF INTEREST: the user can decide to visualize the soil sample at COUNTRY (Figure 2-69 --panel E) or REGION (Figure 2-69 - -panel D) level;
- DEPTH [cm]: the user can visualize the soil samples at a specific level (Figure 2-69 - -panel F):

[{surface}, {10 ÷20}, {20 ÷30}, {30 ÷40}, {40 ÷50}, {50 ÷60}, {60 ÷70}, {70 ÷80}, {80 ÷90}, {90 ÷100}, {100 ÷120}, {120 ÷140}, {140 ÷160}, {160 ÷180}, {180 ÷200}, {>200}];

- SOIL TAXONOMY: the user can visualize the soil samples by using the USDA (Figure 2-69 - -panel H) or WRB (Figure 2-69 - -panel I) soil classification;
- VARIABLE: the user can visualize one of these variables for each sample (Figure 2-69 - -panel L):
  - Calcium Carbonate ($CaCo_3$);
  - organic matter (OM);
  - Sand fraction;
  - Silt fraction;
  - pH (H_20).
- DATE: list of the date of the available soil samples that satisfy all the properties selected in the fields above;

A Dynamic Palette is at disposition of the user that can ranging the range on which to visualize the specific variable choice using personal values.

When the user makes all the choices on the pop-up screen (Figure 2-70 – panel A), on the **Control Map** will be visualized the dots of soil samples that satisfy all the properties selected. If the user moves the cursor on one of the dots visualized (Figure 2-74- yellow arrow in panel B), a panel appears with these information on the upper right corner:

- Name of the sample
- Date: the date of the soil sample;
- Elevation [m]: elevation at which the soil sample has done;
- Depth [cm]: depth at which the soil sample has done;
- USDA: USDA soil taxonomy developed by United States Department of Agriculture and the National Cooperative Soil Survey provides an elaborate classification of soil types
- WRB: World Reference Base classification system for naming soils (**Error! Reference source not found.**)
- CLIMATE: Köppen climate classification of the soil sample
- Country: country of the Roi where the soil sample has done
- Region: region of the Roi where the soil sample has done (Balkans or MENA)
- Value of the selected variable.

In panel B and C of Figure 2-74, the yellow arrows indicated to different samples obtained from the selection done in the pop-up screen (Figure 2-70 – panel A): the two dotes indicated in this example show different color. This is motivated by the different value of sand fraction corresponding to the samples: the two yellow circles indicate 28% and 61% respectively.

Clicking on one of the dots, the two panels reporting the soil spectrum graph and the results obtained by the chemical analysis made on the selected sample are available to the user.

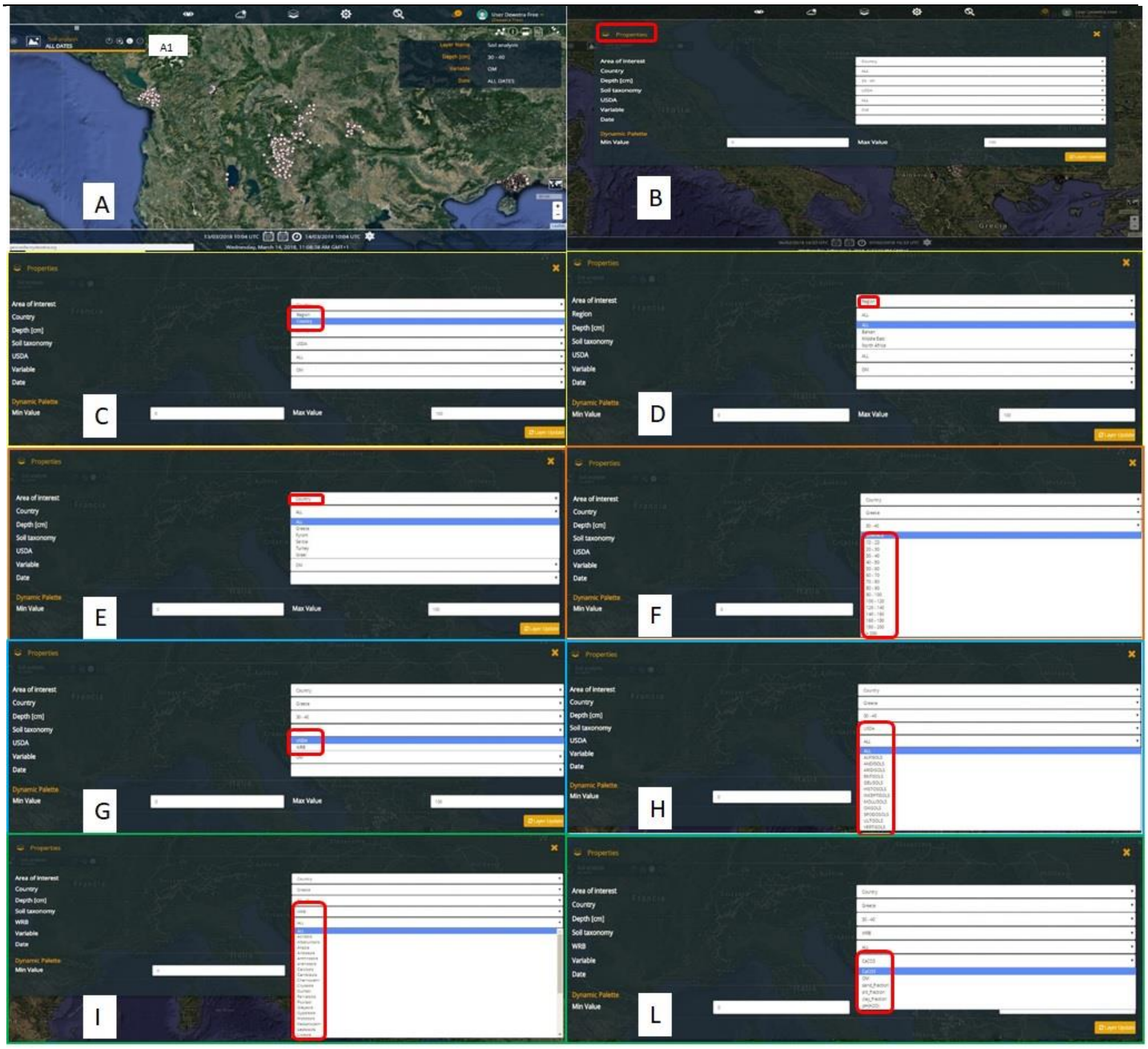

Figure 2-69. Soil Analysis layer -visualization of the steps to select the available soil samples in the RoI with specific properties.

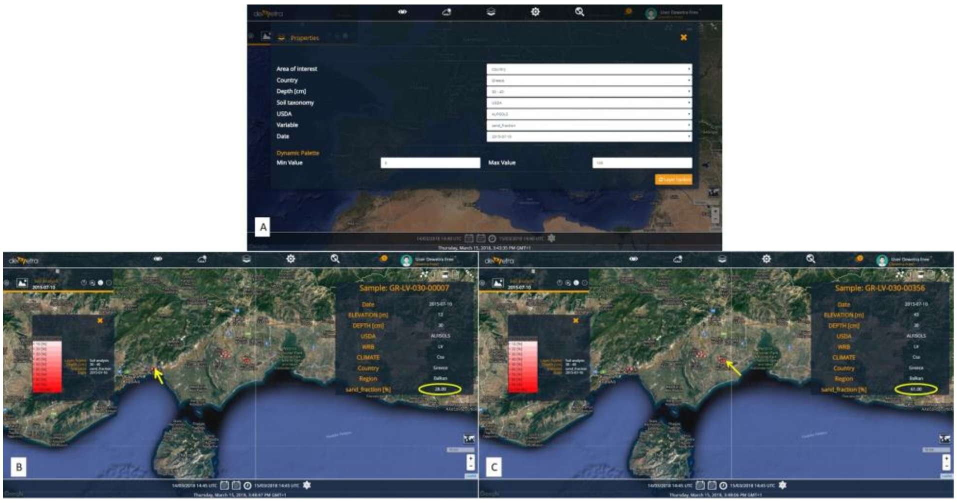

Figure 2-70. Soil Analysis layer -visualization of data samples available after executing the selection of the properties in the pop-up screen.

DROUGHT

Droughts are one of the two main thematic area of the WP4.2 together with Floods. On this line, two different indices have been implemented and now are available on MyDEWETRA.

During the "*Inter-Regional Workshop on Indices and Early Warning Systems for Drought*" happened at the University of Nebraska-Lincoln in December 2009, the drought indices in use around the world were reviewed to support the explanation of agricultural and hydrological droughts and was discussed the need of standard indices for describing different types of droughts. It was recognized the importance of drought monitoring and dissemination of early warning systems information, encouraging countries that have not already done, to take the first steps in implementing such a process.

The idea to make available some indices on MyDEWETRA platform goes in this direction: to provide a system to allow coordination between data monitoring agencies and to facilitate effective decision making. During the Workshop it was also recognized the importance of the use of the Standardized Precipitation Index (SPI) to characterize the meteorological droughts around the world. Anyway, concerns have been raised in the scientific community about the utility of the SPI as a measure of changes in drought associated with climate change, as it does not deal with changes in evapotranspiration. Thus, an alternative index that deals with evapotranspiration has been proposed, such as the Standardized Precipitation Evapotranspiration Index (SPEI).

On the base of these considerations, SPI and SPEI have been make available on MyDEWETRA under the section Observation (Figure 2-71).

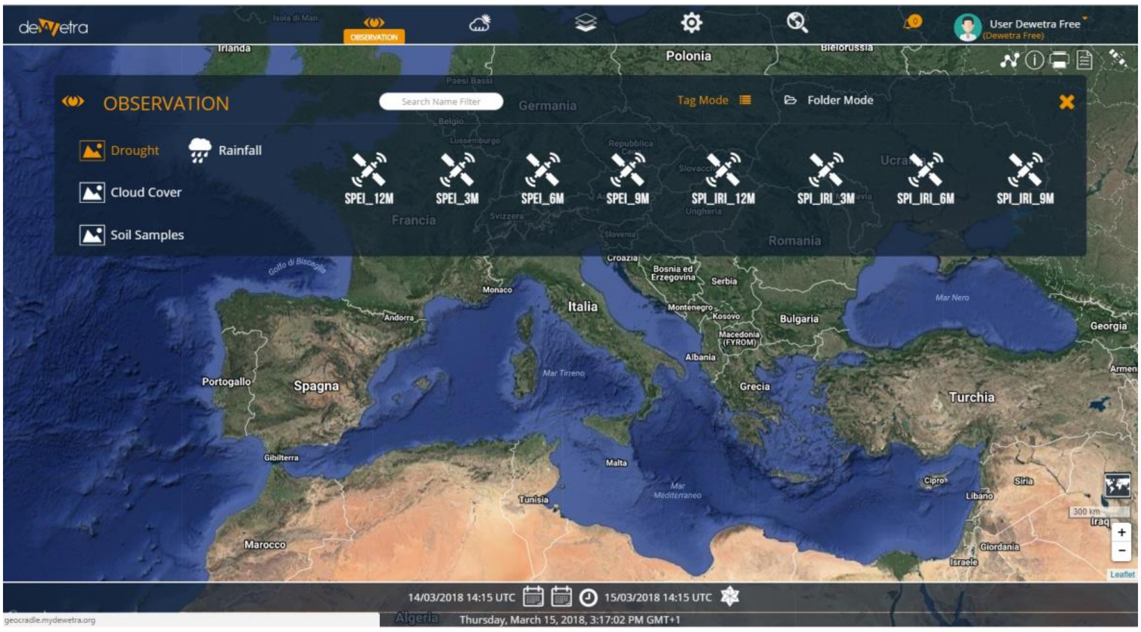

Figure 2-71. Observation section- Drought layer: SPEI and SPI on different accumulation time (3, 6, 9, 12 months) are available.

The Standardized Precipitation Index (SPI)[1] is a widely used to characterize meteorological drought on a range of timescales. On short timescales, the SPI is closely related to soil moisture, while at longer timescales, the SPI can be related to groundwater and reservoir storage. The SPI can be compared across regions with markedly different climates. From literature arisen that the raw precipitation data are typically fitted to a gamma or a Pearson Type III distribution, and then transformed to a normal distribution. The SPI values can be interpreted as the number of standard deviations by which the observed anomaly deviates from the long-term mean. The SPI can be created for differing periods of 1-to-36 months, using monthly input data.

In the contest of WP4.2 topic, analyses of observed precipitation conditions using monthly SPI are provided on MyDEWETRA. These analyses are realized by the International Research Institute (IRI) in collaboration with the NOAA Climate Prediction Center (CPC) and the University of Maine. The SPI values are generated using monthly precipitation totals at 1.0° lat/lon resolution calculated from a dataset that combines the retrospective and real-time CPC Gauge - OLR Blended (GOB) daily precipitation analysis for the globe, accumulated to monthly. The maps available on MyDEWETRA show values of the monthly SPI for 4 accumulation periods (3, 6, 9, or 12-month - *Figure 2-72*) over the globe. The maps provide an indication of short-term to longer-term wet (green, positive SPI) or dry (yellow to red, negative SPI) conditions based upon precipitation alone. The negative half of the color scale uses the same colors and thresholds of SPI corresponding to the percentiles associated with the D0 (30%tile) to D4 (2%tile) drought intensity categories used in the U. S. Drought Monitor[2].

[1] Keyantash, John & National Center for Atmospheric Research Staff (Eds). Last modified 08 Mar 2018. "The Climate Data Guide: Standardized Precipitation Index (SPI)." Retrieved from https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-index-spi.

[2] The U.S. Drought Monitor maps provides a summary of drought conditions across the United States and Puerto Rico and are updated weekly by combining a variety of data-based drought indices and indicators and local expert input into a single composite drought indicator. Ranging from D1 - D4, the map denotes four levels of drought intensity and one level of "abnormal dryness" (D0).

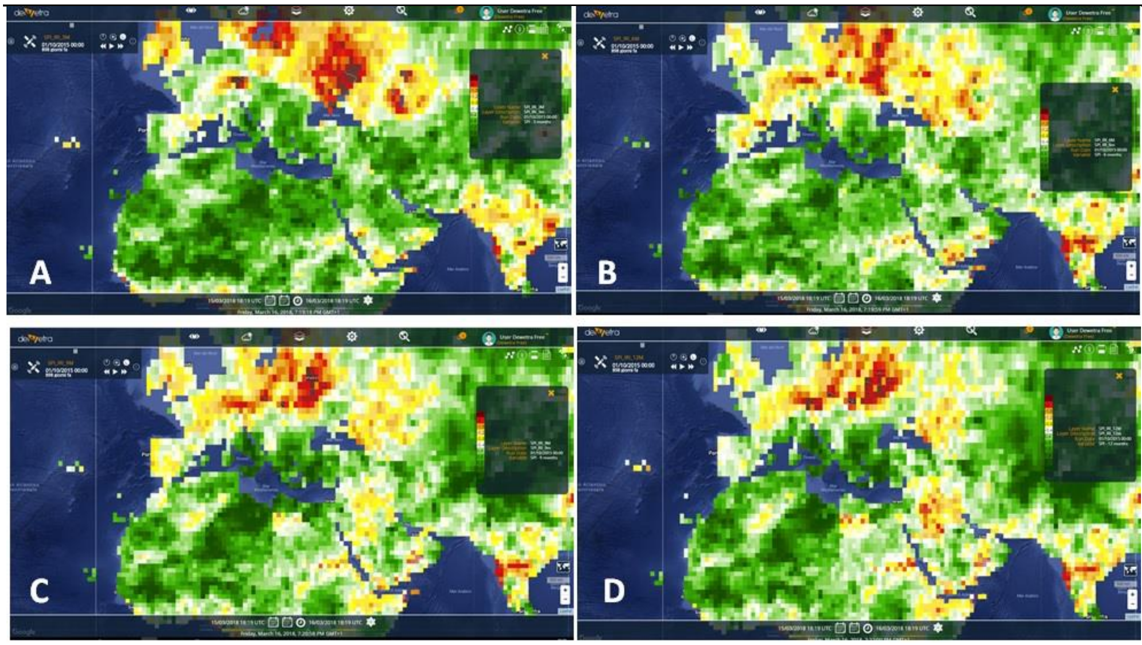

Figure 2-72. Observation section- Drought layer: SPI at accumulation time 3 (panel A), 6 (panel B), 9 (panel C), 12(panel D) months.

The Standardized Precipitation Evapotranspiration Index (SPEI)[3] is an extension of the widely used SPI. It considers both precipitation and potential evapotranspiration (PET) in determining drought. Thus, unlike the SPI, the SPEI captures the main impact of increased temperatures on water demand. Like the SPI, the SPEI can be calculated on a range of timescales from 1-48 months.

In the contest of WP4.2 topic, The Global SPEI database[4], SPEIbase, are provided on MyDEWETRA. The SPEIbase is based on monthly precipitation and potential evapotranspiration from the Climatic Research Unit of the University of East Anglia. Currently the version 3.23 of the CRU TS dataset has been used. It is usually updated as soon as new data becomes available. The data set offers long-time about drought conditions at the global scale, with a 0.5 degrees spatial resolution and a monthly time resolution. It has a multi-scale character, providing SPEI time-scales between 1 and 48 months. Currently it covers the period between January 1901 and December 2018 and the time range available on MyDEWETRA is from September 2005 up to February 2018 (Figure 2-84).

The SPEIbase consists of standardized values over the emerged land pixels. No land pixels are assigned a value of $1.0x10^{30}$. In some rare cases it was not possible to achieve a good fit to the log-logistic distribution, resulting in a NAN (not a number) value in the database. The SPEIbase is based on the FAO-56 Penman-Monteith estimation of potential evapotranspiration. This is a major difference with respect to the SPEI Global

[3] Vicente-Serrano, Sergio M. & National Center for Atmospheric Research Staff (Eds). Last modified 18 Jul 2015. "The Climate Data Guide: Standardized Precipitation Evapotranspiration Index (SPEI)." Retrieved from https://climatedataguide.ucar.edu/climate-data/standardized-precipitation-evapotranspiration-index-spei.

[4] http://spei.csic.es/database.html

Drought Monitor, that uses the Thornthwaite PET estimation. The Penman-Montheith method is considered a superior method, so the SPEIbase is recommended for most uses including long-term climatological analysis.

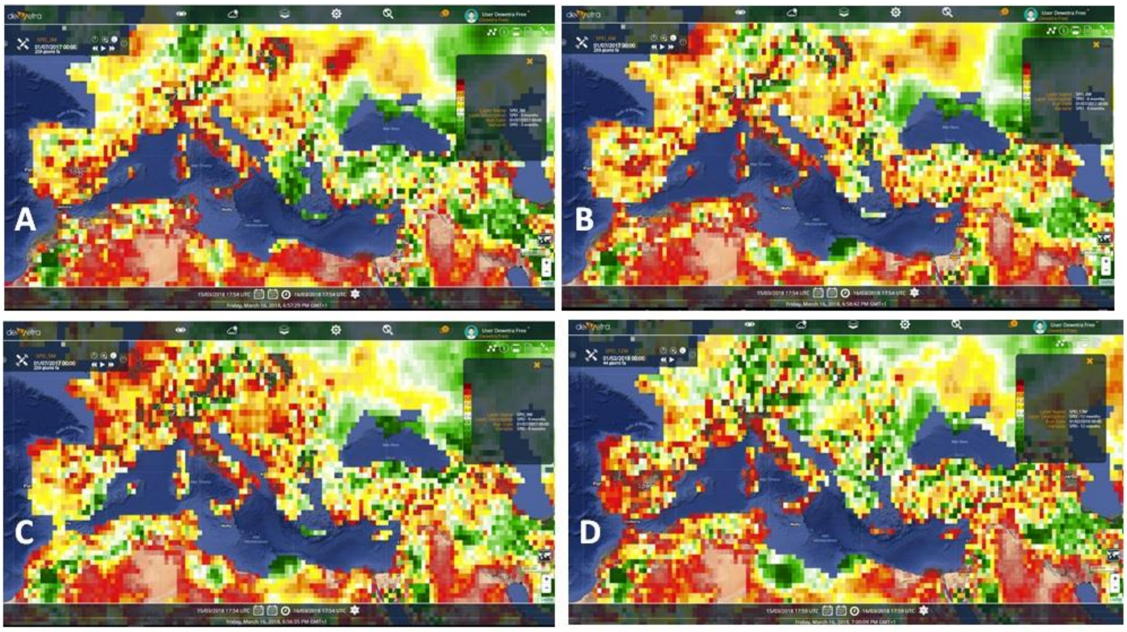

Figure 2-73. Observation section- Drought layer: SPEI at accumulation time 3 (panel A) , 6 (panel B), 9(panel C), 12(panel D) months.

For sake of clarity, both SPI and SPEI maps are implemented to the global scale to allow them to be used by any user who will learn about the presence of this platform through the GEO-CRADLE DATA HUB and not only from the GEO-CRADLE partners or some national agencies, end-users or SME in the RoI.

2.2.4.2. *Forecast models in the MyDEWETRA platform*

The second action button of the Toolbar is the Forecast menu, which provides access to all the forecast models to the user. Once the user clicks on the menu, two different views are offered: the Tag and Folder mode.

In version of MyDEWETRA for the GEO-CRADLE project (*Figure 2-74*), the Tag mode shows the following observational data as organized by thematic criteria

- HYDROLOGICAL MODELS
- METEOROLOGICAL MODELS
- FIRE MODELS

In the table below the layers of this category are shown according to the assigned tags.

Table 2-12. List of forecast models available in the MyDEWETRA platform

TAG	LAYER NAME	SOURCE
HYDROLOGICAL MODELS	ISRAEL_DETERMINISTIC_HYDRO_MODEL - WRF	Israeli Hydrological ServiceWater Authority
	FLOOD PROOFS ALBANIA – PROBABILISTIC LAMI	CIMA
	FLOOD PROOFS ALBANIA – DETERMINISTIC LAMI	CIMA
METEOROLOGICAL MODELS	GFS 0.5°	NOAA
FIRE MODELS	RISICO WORLD	CIMA
	FDI MONDO (RISICO_WORLD)	CIMA

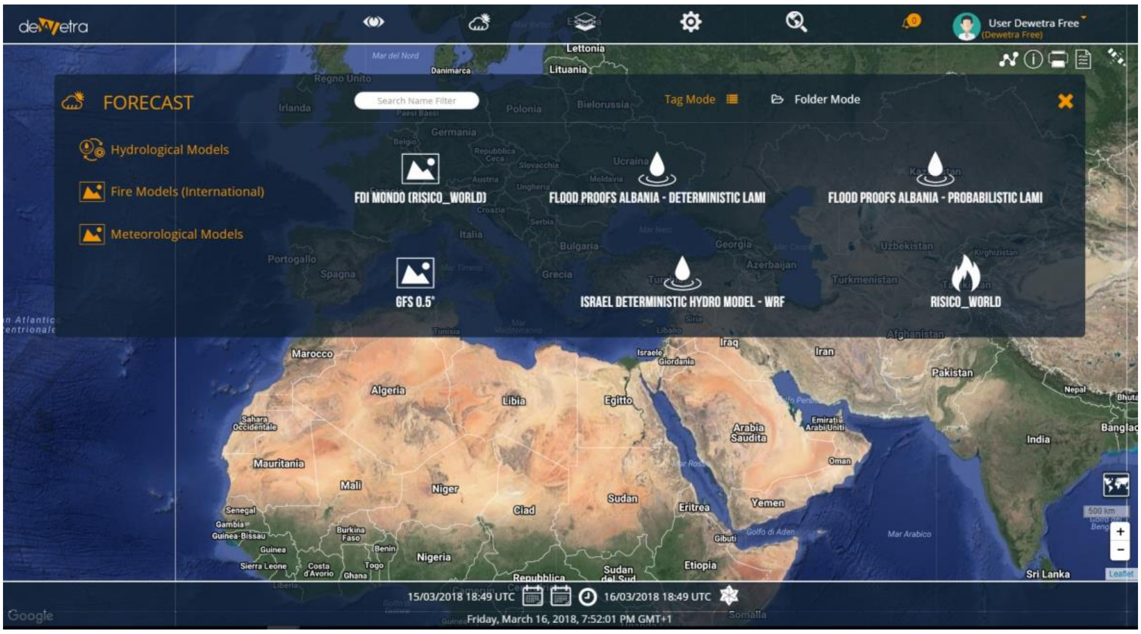

Figure 2-74. Open modelling data available in MyDEWETRA

Every time a layer is pulled on, the application uploads it in the ***Control Map***. If the cursor is left on the name of the layer in the ***Layer List*** (top left of the screen) the user enables the tooltip function to open two windows:

- the first one at the top right of the screen (Figure 2-75 - panel B) which shows the metadata of the layers, such as:
  - the name of the layer
  - the Layer description
  - the reference date
  - the initialization time of the run (if the selected layer is a model's output)
  - the Spatial aggregation (if enabled)
  - the validity interval (if it is a combined variable)
- the second one is placed immediately to the right of the **Layer List** (Figure 2-75- panel A) and shows the name of the layer, the date of initialization of the run (if the case) and/or the reference date.

In the given example in Figure 2-75, the tooltip displays the available information about one of the available variables (total precipitation) of the Global Model GFS 0.5°.

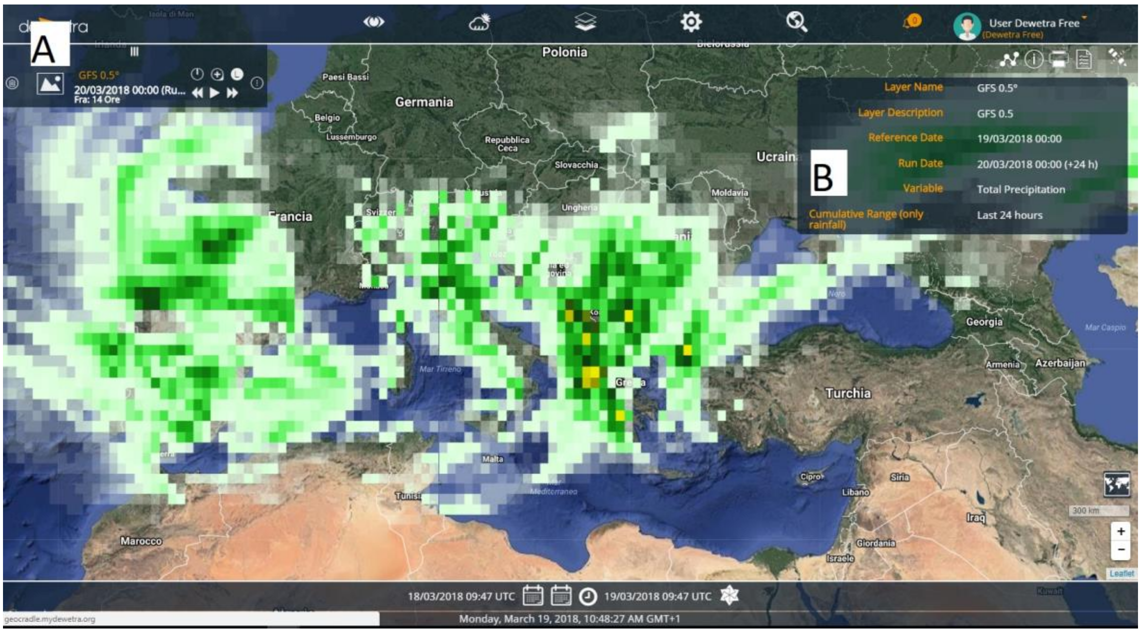

Figure 2-75. Example of one of the observational data available: 3-Month Standardized Precipitation Index (SPI_IRI_3M) is plotted together with the administrative limits for Balkans and MENA regions.

FIRE MODELS (International)

Under the thematic area related to the Water Management, CIMA provides as contribution to the WP4.2 two of the main research products developed by its researchers in the topic of modelling and detection of fire: RISICO Word and Fire Danger Index (FDI).

RISICO Word: the system RISICO provides (daily wildland fire risk forecast maps. The RISICO system has a complex software architecture based on a framework able to manage geospatial data as well as time dependent information (e.g., numerical weather prediction models, real time meteorological observations, and satellite data). Within the system semi-physical models, able to simulate in space and time the variability of the fuel moisture content, are implemented. This parameter represents the main variable related with the ignition of a fire. Based on this information and introducing information on topography and wind field the model provides the rate of spread and the linear intensity of a potential fire generated by accidental or deliberate ignition. The model takes into account the vegetation patterns, in terms of fuel load and flammability.

Territorial data used by the system are vegetation cover and topography. Meteorological data are mainly provided by numerical weather prediction models (GFS 0.5°, in this case). Meteorological data provided in real time by a meteorological network are also used by the model as well as satellite data (e.g., vegetation index, snow cover). The output information is provided on a web-gis based system according with the OGC-INSPIRE standard.

The fine fuel moisture model is derived from the FFMC (Fine Fuel Moisture Code) of the CFFDRS (Canadian Forest Fire Danger Rating System). In addition, a different nominal rate of spread (no-wind on flat terrain) has been introduced for each different class of vegetation. The operational chain of the RISICO system runs every 3 hours making use of the available observations at each time step.

The system RISICO is able to integrate the main Fire Hazard Indexes present in literature providing a suitable tool for testing the different indexes on the same platform in many environmental and climatic conditions.

RISICO represents an operational approach to forest fires management both during the prevention and firefighting phases. Prevention phase is mainly devoted to early warning: fire risk bulletins are issued by Civil Protection authorities and dispatched to all operative bodies employed both in firefighting and civil protection activities. During the firefighting activities RISICO supports decision makers to define the best strategies to cope with fires.

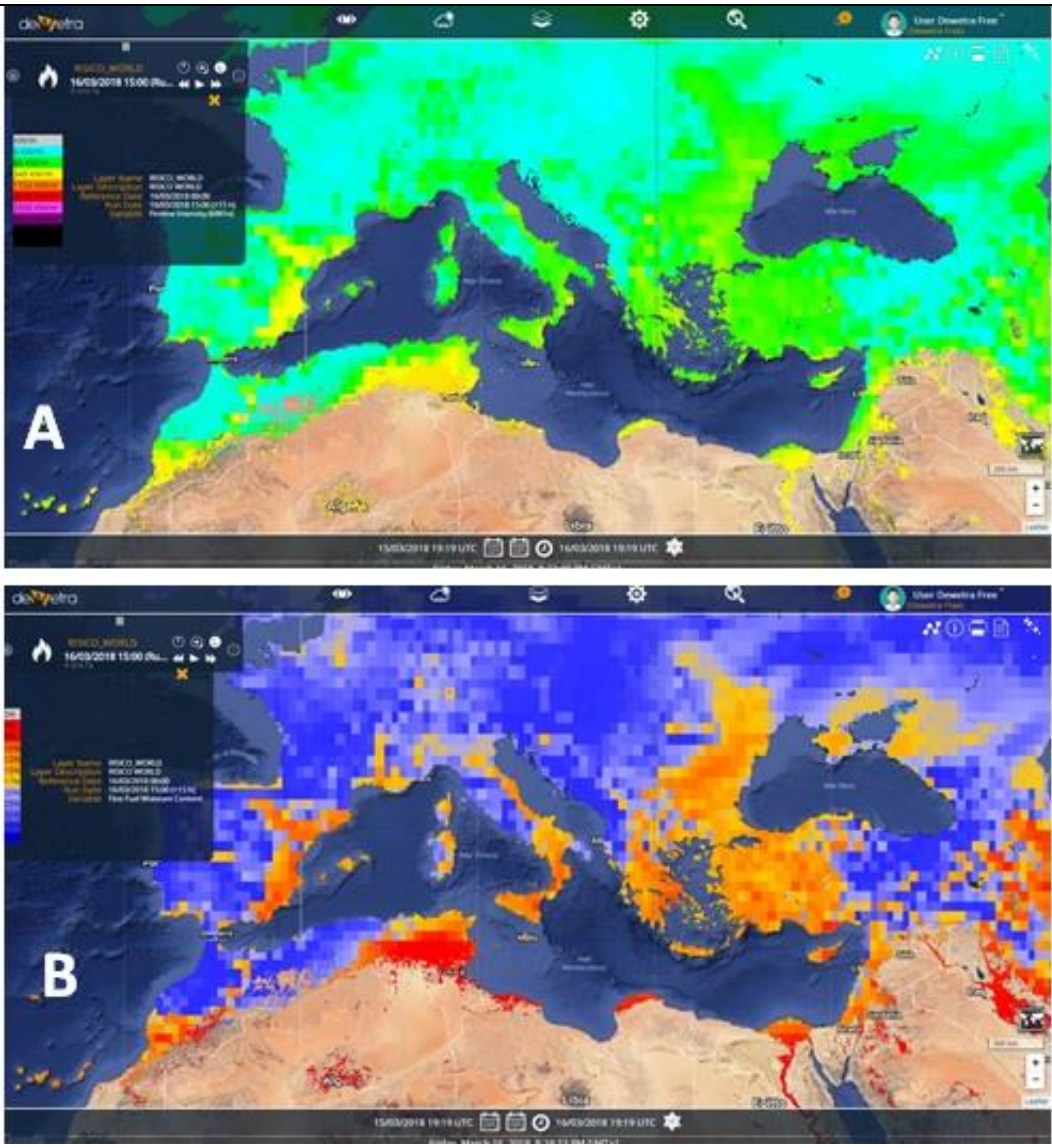

Figure 2-76. Example of some variables available from the daily run of RISICO Word: Fireline Intensity (panel A) and Fine Fuel Moisture Content (panel B)

Fire Danger Index (FDI): is the mean of the 75th percentile of the propagation velocity (PPF) values computed by RISICO World. The index calculated on the 75th percentile highlights the persistence of conditions of severe danger in a 24-hour timespan on at least a quarter of the territory considered or alternatively on the whole territory considered limited to two tri-hour intervals. This index has been defined on the basis of the variable Rate of Spread (PPF) available among the variables provided by RISICO World, obtained as the product of the Rate of Spread (velocity of propagation) for the seasonal danger map (PPF Probability of Fire Propagation), appropriately normalized between 0 and 1. The FDI Mondo is computed using the meteorological input provided by the NWP GFS 0.5°.

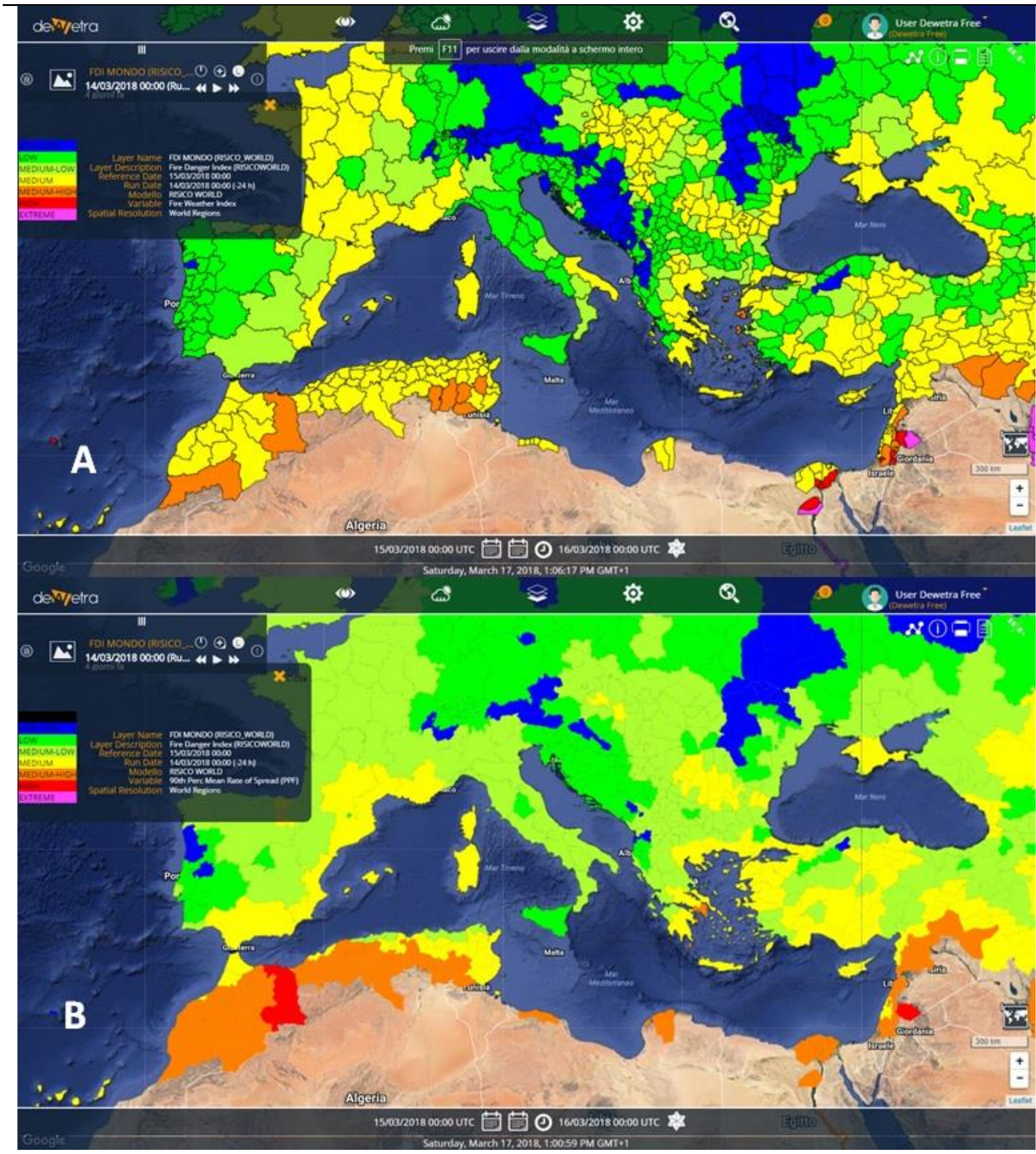

Figure 2-77. Example of some variables available from the daily run of FDI: Fire Danger Index (panel A) and 90th Perc Mean Rate of Spread (PPF) (panel B).

HYDROLOGICAL MODELS

Under the thematic area related to the Water Management, the following hydrological model contribution is presented:

Flood-PROOFS: CIMA provides as contribution to the WP4.2 one of the main research products developed by its researchers in the topic of operational floods forecasting named **Flood-PROOFS**. More precisely, it is available in MyDEWETRA a version implemented on the Drini-Buna basin in Albania. The data of runs that the users can visualize are referred to the period over which *Soil Moisture* and *Clay Content Maps* available in the ***Toolbar*** named ***Static*** (Section 2.2.4.3) are derived from data fusion techniques involving Satellite Images and Soil Sampling data (see Section 2.5 for more details). The field campaign has been scheduled from 15/07/2017 to 4/08/2017 and 107 samples were acquired for the creation of a regional soil spectral library. The hydrological runs from Flood-PROOFS model are from 15/07/2017 to 5/08/2017.

The Flood-PROOFS layer available on MyDEWETRA is a customization for Albania of the Flood-PROOFS (Flood-PRObabilistic Operational Forecasting System) system designed for operational floods forecasting by CIMA. Flood-PROOFS supports decision makers through the forecast phase and furnishes a quantitative evaluation of ground effects in term of discharge and peak flow. It consists of the following elements:

1. a precipitation downscaling module for the generation of fine-time-scale precipitation scenarios: RainFarm [21], [22]
2. a fully distributed and continuous hydrological model: Continuum [23]

RainFARM is a rainfall downscaling algorithm that can produce an ensemble of rainfall scenarios (10 in this precise case) that maintain some characteristics of the rainfall prediction derived by a meteorological model run; it can mimic the small-scale variability of precipitation needed to correctly force the hydrological model. The rainfall scenarios built with RainFARM are used to feed Continuum model which is a distributed hydrological model based on a geomorphologic approach and combines semi-empirical and physically based modules to describe the processes. It can work in a continuous way; all the main physical processes that compose the hydrological cycle are modelled.

The hydrological forecast is carried out both in a deterministic and a probabilistic way. The deterministic module of the system uses as inputs for hydrological modelling the forecasts made by the COSMO-I7 limited area meteorological model. The probabilistic runs are based on the downscaling model of RainFarm precipitation. Since observations from ground-based meteorological sensors are not available, the Continuum hydrological model at the base of this system is fed exclusively with weather forecasts. Obviously, this increases the uncertainty of the hydrological forecast since it is not possible to estimate at best the current state of the basin (soil moisture, snow cover, etc.).

Meteorological precipitation forecast (COSMO I-7) and the unpredictability of precipitation patterns at small space-time scales, are both considered. The system has a complex software architecture based on a framework able to manage geospatial data as well as time dependent information. In 2011, within the framework of the Italian Cooperation Program, the system has been implemented on the Drini-Buna basin in Albania.

Flood-PROOFS provides flood forecasts for the Drin-Buna catchment. The basin response is modelled by a fully distributed hydrological model able to models also the dams and the lake system. The basin is discretized on a regular grid of 300x300 meters. Static and dynamic input data feed the system:

- Static: topographic (DEM), soil type, soil use and vegetation cover data (GIS)
- Dynamic: meteorological forecast and observed data (if available), manoeuvres performed on the hydraulic structures (if available), satellite data (MODIS)

The DRIN-Buna catchment extends from Northeast of Albania toward the centre on the West border of Albania. It covers areas from Montenegro, Kosovo to FYROM. The whole basin is about 17.000 km^2, with five dams (Spilje, Komani, Mavrovo, Vau Dejes, Fierza dams) and two lakes (Skutari Lake and Ohrid Lake). Starting from the available Quantitative Precipitation Forecast, the main outputs produced by the computational chain Flood-PROOFS are two for each selected section. The discharge forecast is currently modelled for the critical hydraulic sections displayed in **Error! Reference source not found.** (panel A and B).

Figure 2-78 shows an example of probabilistic and deterministic outputs. Panel E shows deterministic outputs while panel C and D show the probabilistic ones for the run on 25/7/2017. The probabilistic output is composed by two graphs: the first one is the typical discharge time series (panel C) while in the second one the discharge is associated to its occurrence probability (panel D).

Figure 2-78. DRINI and BUNA river basin with FloodPROOFS river sections on MyDEWETRA.

2.2.4.3. *Static data in the MyDEWETRA platform*

The third action button of the Toolbar is the Static Layers menu, which provides access to all the time-independent data to the user. Once the user clicks on the menu, two different views are offered: the Tag and Folder mode.

In version of MyDEWETRA for the GEO-CRADLE project (Figure 2-79) the Tag mode shows the following observational data as organized by thematic criteria

- Administrative Units
- Natural risk zones
- Soil moisture
- Clay Content
- Drin Buna
- Land Cover
- Hydrography

In the table below the layers of this category are shown according to the assigned tags.

Table 2-13. List of static data available in the MyDEWETRA platform

TAG	LAYER NAME	SOURCE
ADMINISTRATIVE UNITS	BALKANS Gadm level 1	GADM
	MENA Gadm level 1	
HYDROGRAPHY	Hydroshed USGS	
NATURAL RISK ZONES	Balkans Flood hazard 100y GAR	CIMA
	Balkans Flood hazard 50y GAR	
	Balkans Flood hazard 25y GAR	
	MENA Flood hazard 100y GAR	
	MENA Flood hazard 50y GAR	
	MENA Flood hazard 25y GAR	
SOIL MOISTURE	Soil Moisture - Drin-Buna	i-BEC
CLAY CONTENT	Clay Content - Drin-Buna	i-BEC
DRIN BUNA	Clay Content - Drin-Buna	i-BEC
	Soil Moisture - Drin-Buna	

LAND COVER	Corine Land Cover BG	CLC 2012
	Imperviousness Degree BG	GMES

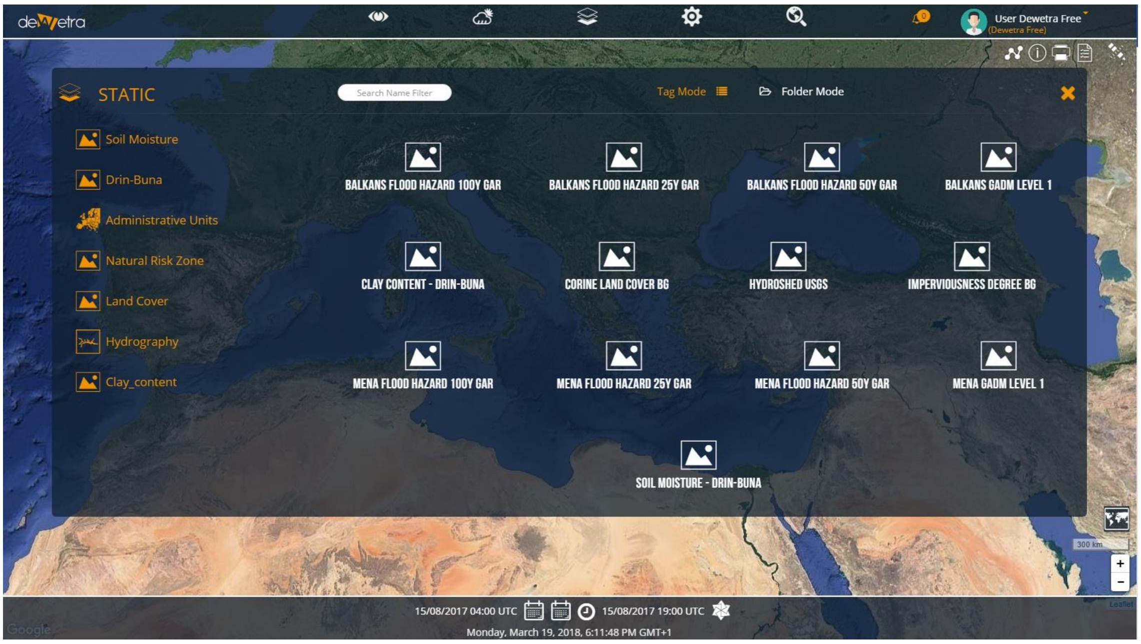

Figure 2-79. Open static data available in MyDEWETRA.

Every time a layer is pulled on, the application uploads it in the ***Control Map***. If the cursor is left on the name of the layer in the **Layer List** (top left of the screen) the user enables the tooltip function to open two windows:

- the first one at the top right of the screen (Figure 2-80- panel A2 and B2) which shows the metadata of the layers, such as:
  - the name of the layer
  - the Layer description
  - the reference date
  - the initialization time of the run (if the selected layer is a model's output)
  - the Spatial aggregation (if enabled)
  - the validity interval (if it is a combined variable)
- the second one is placed immediately to the right of the **Layer List** (Figure 2-80- panel A1 and B1) and shows the name of the layer, the date of initialization of the run (if the case) and/or the reference date.

In the given example in Figure 2-80, the tooltip displays the available information about the Soil Moisture -Drin Buna (panel A) and the Clay Content (panel B) available among the Static layer.

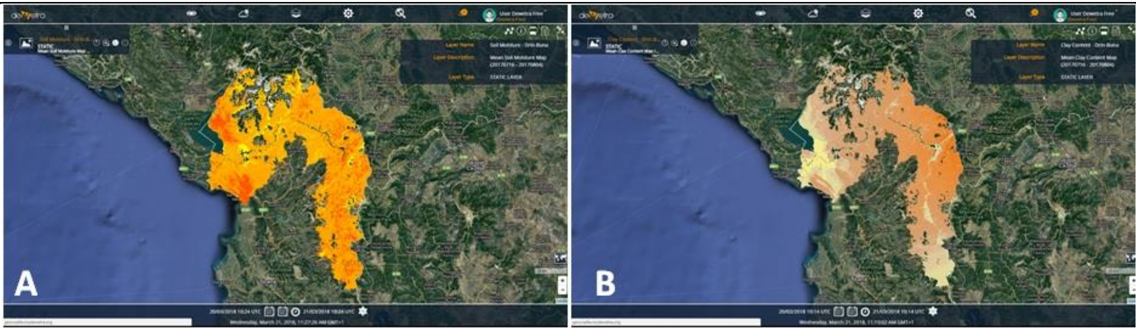

Figure 2-80. Example of STATIC data available: Soil Moisture and Clay content maps on the Drin-Buna catchment.

LAND COVER

The presence of two maps under this section is driven by some of the main issues of the pilot oriented in supporting large scale agriculture monitoring, exploding existing EO data and collecting new data. Maps are provided by one of the partners of the project (Space research and technology institute -SRTI) with the intention of paving the road towards a regional cooperation and sharing data.

In Figure 2-81 are reported the Corine Land cover 2012 (panel A-B) and Imperviousness Degree [%] (panel C-D).

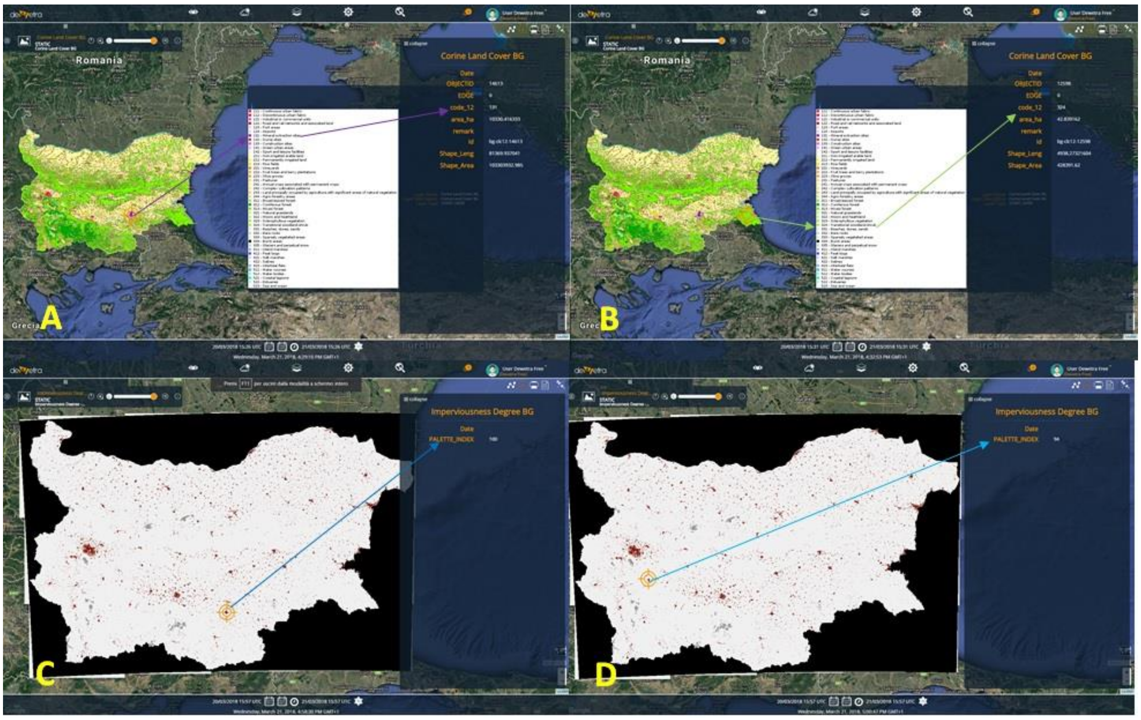

Figure 2-81. Example of values of CLC in Bulgaria.

2.3. Development of geo-statistical maps

The regional SSL can be used to derive geo-statistical maps of selected soil properties. Assuming that a region has been properly sample (i.e. the soil samples cover the underlying pedodiversity present), geo-statistical maps can be generated through extrapolation based on the data. There are two groups of interpolation techniques: deterministic and geostatistical. All methods rely on the similarity of nearby sample points to create the surface. Deterministic techniques use mathematical functions for interpolation. Geostatistics relies on both statistical and mathematical methods, which can be used to create surfaces and assess the uncertainty of the predictions. The principal idea is that neighbouring points tend to have the same properties, and hence predictions can be made with high degree of accuracy.

One common technique is Kriging (or Gaussian process regression) which is a method of interpolation for which the interpolated values are modelled by a Gaussian process governed by prior covariances. This is in contrast to other technique such as a piecewise-polynomial spline chosen to optimize smoothness of the fitted values. Under suitable assumptions on the priors, kriging gives the best linear unbiased prediction of the intermediate values. Interpolating methods based on other criteria such as smoothness need not yield the most likely intermediate values.

Due to the spatial data sparsity existing in the GEO-CRADLE SSL, there are only limited areas where this may be used successfully. One such area is the Nestos River Delta, in Northern Greece. A generated geo-statistical map of Organic Matter using Kriging is given in Figure 2-82. However, the underlying pre-requisite of geostatistics, that of the sufficient sampling, dictates the need for large datasets, which significantly drives up the cost of sampling and analyses. A faster, less costly alternative is to use Earth Observation data and the SSL which can produce more accurate results, even in areas that were not adequately sampled. This significantly improves the operational activities of the SSL, and the methodology is detailed in the following sections.

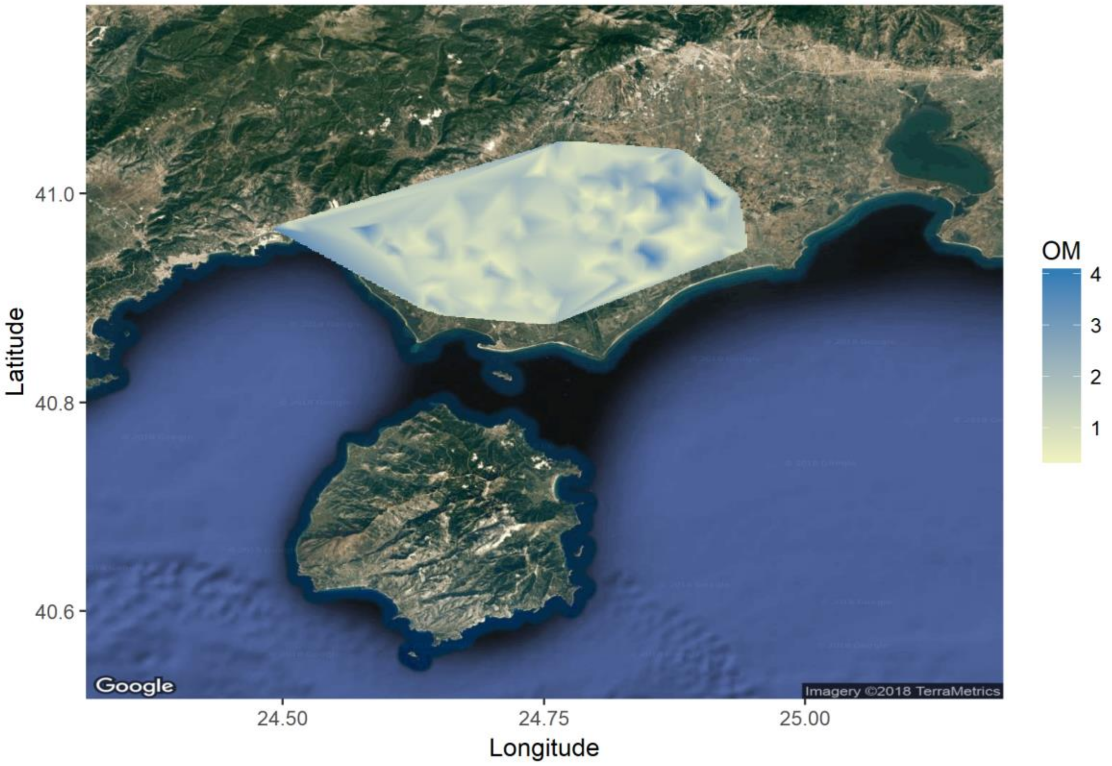

Figure 2-82. A generated map of OM in Northern Greece using geostatistical techniques

2.4. Development of laboratory soil spectroscopy models

The recorded diffuse reflectance spectra in the regional SSL cover the vis–NIR region. Many soil chemical components (chromophores) interact with the electromagnetic radiation within this region. Soil spectroscopy exploits the correlation between spectral features and chromophores in order to estimate soil variables. However, the vis-NIR region is largely nonspecific due to the significant overlapping absorption of soil constituents. In other words, it is very difficult to exactly assign a single wavelength only to one absorption band. When NIR electro-magnetic radiation interacts with a soil sample, it is the overtones and combinations of fundamental vibrations in the mid-infrared (mid-IR) region that are detected. For example, bands in the NIR and SWIR may be ascribed to specific clay minerals (e.g. Kaolin doublets at 1415 nm, Smectite Al-OH at 2230 nm). Overtones of the O–H and H–O–H stretch vibrations in the free water produce two deep absorption features at 1455 nm and 1915 nm, while organic matter has a strong relationship with electromagnetic radiation in the visible region [24]. However, generally, the NIR region is characterized by broad, superimposed, and weak vibrational modes, giving soil NIR spectra few, broad absorption features. Due to the broad and overlapping bands, vis–NIR spectra contain fewer absorptions than the mid-IR, and as a direct consequence can be more difficult to model. However, this region contains useful information on organic and inorganic materials in the soil as detailed above. This underlying information needs to be mathematically extracted from the spectra and requires powerful and robust methodologies. In the subsections that follow, the multivariate calibration approaches considered during the realization of this task are described.

2.4.1. **Overview of the machine learning algorithms used**

2.4.1.1. *Partial Least Squares Regression*

The Partial Least Squares Regression (PLSR) algorithm [25]–[28] is an extension of the simple Ordinary Least Squares (OLS) solution, and the Principal Components Regression (PCR) algorithm [29]. In the case of spectral data, the predictors (i.e. the recorded spectrum at each distinct wavelength) are highly correlated and outnumber the number of observations. In such cases, OLS fails to provide a robust solution, or even one solution at all. One simple solution is to calculate the pairwise correlations among the predictors and disregard the ones above a pre-specified threshold. This however, does not necessarily ensure that linear combinations of predictors are uncorrelated with other predictors. If this is the case, then the ordinary least squares solution will still be unstable. Therefore, this process does not guarantee a stable least squares solution. Another common solution in highly dimensional spaces is to apply the Principal Components algorithm on the predictors, which effectively transforms the input space into a new, more compact space (i.e. with fewer predictors). Each of the new predictors is a linear combination of the initial predictors; applying regression on this new input space is the PCR algorithm. Since PCA does not consider any aspects of the response when it selects its components and is only interested in covering the variability present in the predictor space, should this variability be uncorrelated to the variability of the response, then PCR will have difficulty identifying a predictive relationship when one might actually exist.

The PLS algorithm solves the aforementioned predicaments by identifying underlying, termed latent, relationships among the predictors that are highly correlated with the response variable. PLS, like PCA, finds linear combinations of the predictors and are commonly called latent variables. While the PCA linear combinations are chosen to maximally summarize the predictor space variability (i.e. to compress the input space), the PLS linear combinations of predictors are chosen to maximally summarize the covariance with the response (i.e. to compress the input space in a direction favourable to the response).

2.4.1.2. *Cubist*

Cubist is a rule-based model that is an amalgamation of several methodologies that were published in their primordial phase [30], [31] but continued evolving as a form of a commercial package. In 2011, the source code was released under an open-source licence, where the full details of the current version were passed to the realm of public property. An R package was soon introduced, which was applied in vis-NIR spectroscopy

[4] with favourable results. A detailed description of the algorithm can be found in [32]. Briefly, the algorithm defines a tree, which can be reduced to a set of rules. Each branch defines the premise part of the rules and is used to split the data using one or more predictors, while the terminal leaves contain linear regression models (i.e. the consequent of each rule). Cubist can also use a boosting-like scheme which is termed committees. Here, iterative models are created by taking into account the predictions of the previous models. Additionally, Cubist can also adjust the predictions of the rule-based model using known errors on the training set; when a prediction is made on a testing pattern, its nearest neighbours in the training set are identified and the prediction is corrected using their know errors.

2.4.1.3. *Support Vector Regression*

Support vector machine (SVM) analysis is a popular machine learning tool for classification and regression, was first identified by Vladimir N. Vapnik and Alexey Ya. Chervonenkis in 1963. In 1992, Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes [33]. A version of SVM for regression, termed Support Vector Machines for Regression, was proposed by Vapnik et al. in 1996 [34]. It is considered a nonparametric technique because it relies on kernel functions. A Gaussian radial basis function was used as a kernel function throughout this document.

2.4.1.4. *Elastic Net*

Elastic Net regularization [35] which is a general form of the ridge and lasso methods. In particular it uses both λ_1 and λ_2 penalties of the respective methods. Thus, the method extends the ordinary least squares method, by performing simultaneous regularization and variable selection. The tuning parameters are the λ_2 term and the fraction s of the L_1-norm of the coefficient vector, relative to the norm at the full least squares solution.

2.4.1.5. *K-Nearest Neighbours (KNN)*

The weighted k-nearest neighbours (kNN) algorithm is a type of instance-based learning which in order to predict the output of a testing pattern uses its k nearest neighbours (defined using a distance metric) from the calibration set and calculates the prediction as the weighted average of the neighbours' respective outputs

2.4.2. **Creating the calibration sets**

To validate the results of a model it is imperative to split the dataset into two disjoint sets. The first is termed calibration set and is used to calibrate the model. The second set is called validation and is used to validate the model – that is, when the outputs of the validation set are known they can be compared with the model's prediction.

To establish these two sets, the widely used Kennard-Stone algorithm [36] was employed. The Kennard–Stone algorithm allows to select samples with a uniform distribution over the predictor space. It starts by selecting the pair of points that are the farthest apart using a pre-defined geometrical distance metric. This pair is assigned to the calibration set and removed from the list of points. Then, an iterative procedure commences which essentially adds the next sample which is the farthest apart from its closest neighbours in the calibration set. As far as the distance metric, the Mahalanobis distance in the PC space was used in our approach.

2.4.3. **Validation measures**

After the calibration of the models, they were used to predict the know properties y of the validation sample. The model's predictions $\hat{y}$ can then be compared with the know outputs. The following validation measures where used to assess the accuracy of the developed models: The root mean square error of prediction, calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=N}(y_i - \hat{y_i})^2}{N}}$$

The coefficient of determination R^2 which is the proportion of the variance in the dependent variable that is predictable from the independent variable, which is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{i=N}(y_i - \hat{y_i})^2}{\sum_{i=1}^{i=N}(y_i - \overline{y})^2}$$

The ratio of performance to the interquartile distance, which has been shown to be a more robust performance estimator, calculated thusly [37]:

$$RPIQ = \frac{Q3 - Q1}{RMSE}$$

where Q3 – Q1 denotes the difference between the 3rd and 1st quartiles of the measured variable.

2.4.4. **Pre-processing techniques**

Spectral pre-processing (or pre-treatments) refers to the application of a data transformation to the initial recorded spectra, in order to enhance significant features. These transformations can account for non-linearities, measurement errors, as well as sample variations and noisy spectra. As noted, the chemical composition of the soil sample has an effect on its reflectance spectrum. However, its structural properties also influence the spectrum by causing non-linear light scattering effects. More concretely, the particle size distribution (i.e. its texture) affects the degree of scattering. A coarser structure increases the scatter (and as such reduces reflection) and the apparent absorbance increases as the path length increases. The pre-processing techniques aim to both enhance the more chemically significant peaks in the spectra as well as reduce unwanted effects such as baseline shifts and overall curvature. [38]

The following pre-processing techniques were examined and applied to the recorded reflectance spectra (Figure 2-83):

- The pseudo-**absorbance** transformation as $-\log_{10} R$
- The **first-derivative** transformation of the absorbance spectra, using a Savitzky Golay filter of 1st order, polynomial size of 3, and window size of 51
- The **second-derivative** transformation of the absorbance spectra, using a Savitzky Golay filter of 1st order, polynomial size of 3, and window size of 51
- The **continuum removal** calculated as the quotient of each spectrum divided by its fitted convex hull
- The **Standard Normal Variate** transformation which transforms each spectrum to have zero mean and a variance of 1
- The **de-trending** transformation which acts as a baseline correction and removes the linear trend

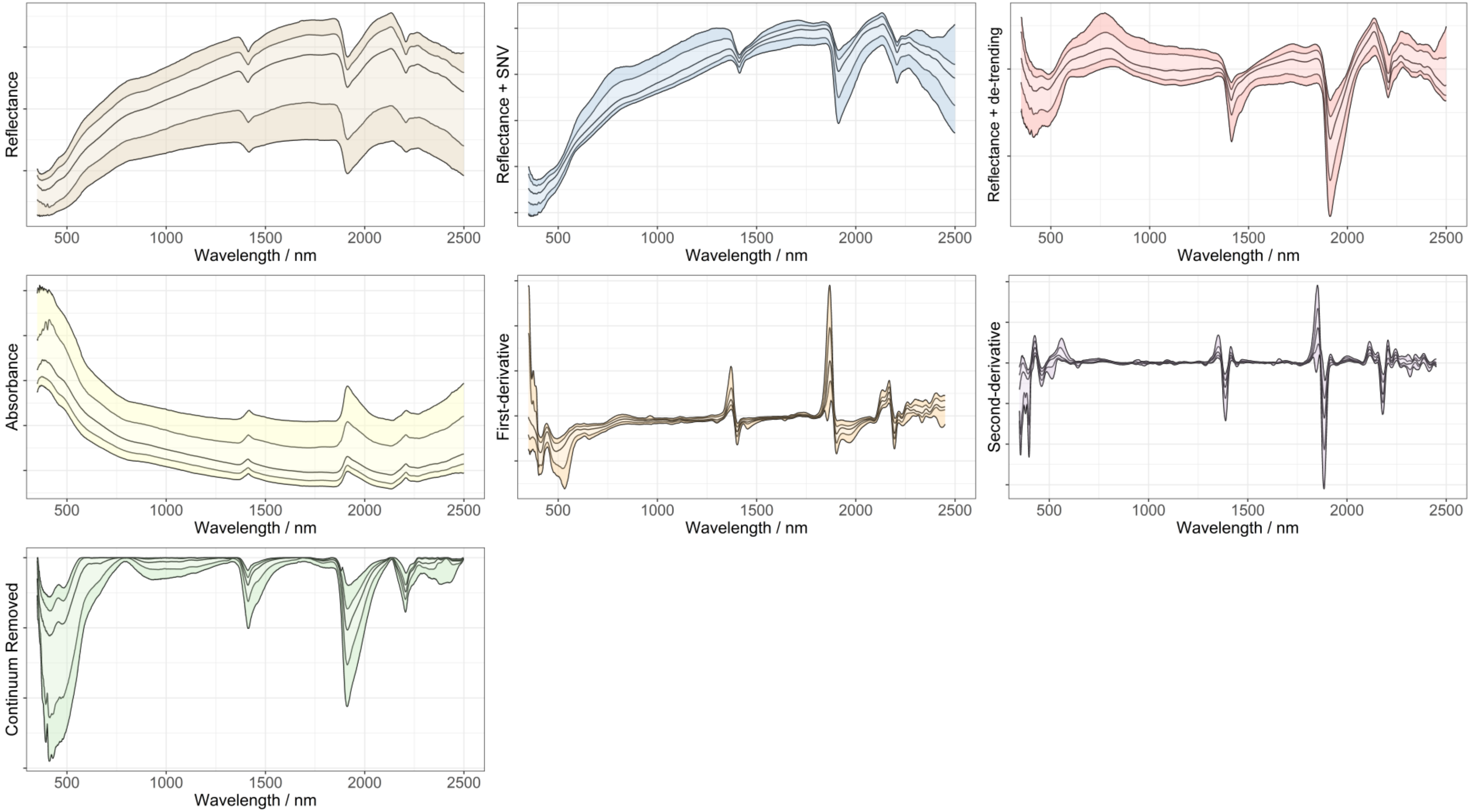

Figure 2-83. The initial reflectance spectra, and the 6 pre-treatments examined; depicted are the 5th, 16th, 50th, 84th and 95th percentiles

2.4.5. **Results of global models**

One approach is to create global models, i.e. models which will use all the samples which comprise the GEO-CRADLE SSL, in order to predict the soil properties. The idea is that irrespective of the sample's origin, a global model should be able to give a good first-order approximation of its physical and chemical properties. To generate these models, each combination of spectral pre-treatment and algorithm was tested in order to develop the most accurate models. The best models developed are given in Table 2-15 for 3 of the most ubiquitous properties. These results indicate that a global model, although it may yield somewhat moderate results for the Clay Fraction, its accuracies are diminished for other soil properties. This suggests that naïve global models are not sufficiently accurate to discern minute changes in the concentration of the constituents. In the literature there are many references which may found that conclude that global models usually fail to produce sufficient accuracy; this gave rise to a number of other local approaches which attempt to utilize the spatial distance and/or spectral similarity to create local models (e.g. [5], [39], [40]).

Table 2-14. Results of the best models for the 3 most common properties

Property	Preprocess	Algorithm	RMSE	R^2	RPIQ
Clay Fraction	Reflectances	SVM	8.90	0.76	1.19
OM	Abs + 1st der.	SVM	0.52	0.49	1.82
$CaCO_3$	Abs + 1st der.	Cubist	1.86	0.27	0.37

2.4.6. **Results of per-country models**

Another approach entails the development of specific models per each country. Although not a very sophisticate, it rests upon the assumption that the regression problem should be decoupled given that soils from one region may be significantly dissimilar to soils from another. To this end, country specific models were developed using every possible combination of pre-processing technique and soil property (Table 2-15). It must be noted that the results were not particularly improved, highlighting the difficult of the task. Although some notable examples exist (e.g. Clay Fraction for most countries, OM for Greece and Israel) some models do not produce sufficiently good results. This signals the fact that a more complex decoupling local approach must be applied and put to the test.

Table 2-15. Results of the best models per-country for the 3 most common properties

Country	Property	Preprocess	Algorithm	RMSE	R^2	RPIQ
Albania	Clay Fraction	Abs + 2nd der.	SVM	8.62	0.54	1.15

Albania	OM	CR	Cubist	0.61	0.39	1.62
Bulgaria	Clay Fraction	Ref + detrending	SVM	8.00	0.78	3.64
Bulgaria	OM	Absorbances	SVM	1.33	0.45	0.77
Cyprus	Clay Fraction	Ref + SNV	SVM	5.49	0.20	1.55
Cyprus	OM	Abs + 2nd der.	SVM	2.96	0.30	0.09
Cyprus	$CaCO_3$	Reflectances	Cubist	19.75	0.59	2.29
FYROM	Clay Fraction	Ref + SNV	PLS	10.62	0.50	1.35
FYROM	OM	Reflectances	Cubist	0.52	0.39	1.22
Greece	Clay Fraction	Ref + SNV	Cubist	6.12	0.38	1.31
Greece	OM	Abs + 1st der.	SVM	0.33	0.65	2.42
Greece	$CaCO_3$	Ref + detrending	SVM	0.67	0.15	0.53
Israel	Clay Fraction	Abs + 2nd der.	SVM	6.59	0.88	3.91
Israel	OM	Ref + detrending	SVM	2.03	0.51	1.17
Israel	$CaCO_3$	Reflectances	SVM	16.14	0.05	1.11
Serbia	Clay Fraction	CR	SVM	10.57	0.24	1.09
Serbia	OM	Abs + 2nd der.	Cubist	0.87	0.22	1.03
Serbia	$CaCO_3$	Reflectances	SVM	0.53	0.10	1.77
Turkey	Clay Fraction	Reflectances	SVM	8.69	0.76	3.30
Turkey	OM	Abs + 1st der.	SVM	0.78	0.46	1.38
Turkey	$CaCO_3$	Abs + 1st der.	Cubist	11.28	0.27	1.37

2.4.7. **The Local Gaussian Process Regression approach**

Taking into consideration all of the above, as well as the current research trends and state-of-the-art of soil spectroscopy, local models must be developed which will compensate for the poor performance results of the global models. The context of locality can have a dual meaning: a) it may refer to spatial locality, i.e. samples that are geographically neighbours, b) or to spectral locality, i.e. samples whose spectral signature is similar. Both of these are important and may be exploited to enhance the performance of the models.

To utilize both criteria of locality, a novel memory-based learning approach was considered, building upon works of the Spectrum-based Learner [39], a local PLSR model developed for the LUCAS SSL [40], and a local PLSR model developed for the Chinese SSL [5]. The basic idea of the model is to use memory-based learning to create for each testing pattern (unknown sample) a new model developed after taking into account its closest neighbours, defined using both terms of spatial and spectral locality. The underlying model utilizes Gaussian Process regression to predict the soil properties. The algorithm has been detailed extensively and was submitted for publication in a peer-reviewed scientific journal.

Therefore, the modified Local Gaussian Regression algorithm (LGR) involves the following steps:

1. Fit a set of pre-processing techniques S in the original spectra and then calculate as a weighted mean of the spectral and spatial distance, the distance vector between each prediction sample and all the reference samples. Possible combinations of weights with corresponding values of 0, 1/2 and 1 are evaluated for spectral and geographical coordinates vectors.
2. Generate datasets with the most similar local calibration sets for each prediction sample. In this line, the model search for k nearest neighbours that fulfil the criteria to minimize each di as generated in the previous step. The optimal k for nearest neighbours tested among the values 40 to 250 following a step of 15.
3. Fit a multivariate regression model with the local calibration samples (described in step 2) and select the optimal parameters that minimize the RMSE in the calibration step. Those parameters could be the weight of geographical and spectra co-variables and the selected pre-processing method in computation of distance vector.
4. Finally, a Gaussian process regression with a linear covariance function is performed for each prediction sample.

Table 2-16 indicates the results of the performance of the best prediction of soil attributes utilizing the LGR after hyper-parameter tuning. It is noted that for all the models, an equal weight (1/2) for the spectral and spatial distance produced the most predictive model.

Table 2-16. Results of the LGR algorithm – k indicates the number of selected nearest neighbours

Property	k	RMSE	R^2	RPIQ
Organic C %	190	0.49	0.64	1.85
$CaCO_3$	100	3.45	0.93	0.69
$\log_{10}(CaCO_3)$	115	0.18	0.90	2.99
pH_{water}	115	0.49	0.82	2.97
Clay %	190	6.50	0.83	2.49
Sand %	220	10.88	0.78	2.94
Silt %	190	8.54	0.66	2.49
EC	85	1.70	0.62	0.76

2.4.8. **A GA-based stacking approach**

A novel GA-based stacking algorithm for predicting soil properties was developed during the project's realization phase and applied in the GEO-CRADLE SSL. The algorithm has been detailed extensively and was submitted for publication in a peer-reviewed scientific journal. This global approach relies on the application of ensemble learning [41], [42] which effectively combines the base models developed. In this approach we elected to examine the performance of the algorithm using as base models the globally developed models (i.e. where each model uses soil samples extending to the whole RoI). Instead of selecting the best model out of the many developed ones, it is possible to combine their predictions, to create a meta-model. Model stacking refers to the process of training a learning algorithm to combine the predictions of other learning algorithms, in order to obtain results which are better than the results obtained by each constituent model. The approach involves the stacking of the different models (termed L1 models) through the creation of a meta-regressor comprised of models stemming from different pre-processing techniques and machine learning algorithms. The derived ensemble regressor is named L2 model. Taking into account that there are many different pre-processing methods, and many different machine learning algorithms (as seen in the previous sections), it is computationally expensive to search every possible combination; thus, a genetic algorithm (GA) [43] is used to automatically find the most efficient and accurate combination of models. Since this approach relies on the combination of different models, it is important to combine models that produce accurate but uncorrelated predictions (considering that combining the same predictions cannot possibly yield better results). This is the objective of the custom GA algorithm.

The necessity of GA is dictated by the sheer number of potential combinations of models. Suppose that we have P pre-processing methods and M_{L_1} machine learning algorithms, with a total of $P \times M_{L_1}$ of different L1 models. Further, granted that the minimum number of L1 models that can be combined is 2, suppose that there is an arbitrary integer k such that $2 \leq \ k \leq P \times M$ denoting the upper limit on the maximum models to be combined by the stacking algorithm. Then, the different combinations of models which can be combined using one of the M_{L_2} machine learning algorithms are:

$$c = M_{L_2} \times \sum_{i=2}^{i=k} \binom{P \times M_{L_1}}{k}$$

A visualization of the number of combinations for various values of k, P and $M_{L_1} = M_{L_2} = M$ is given in Figure 2-84.

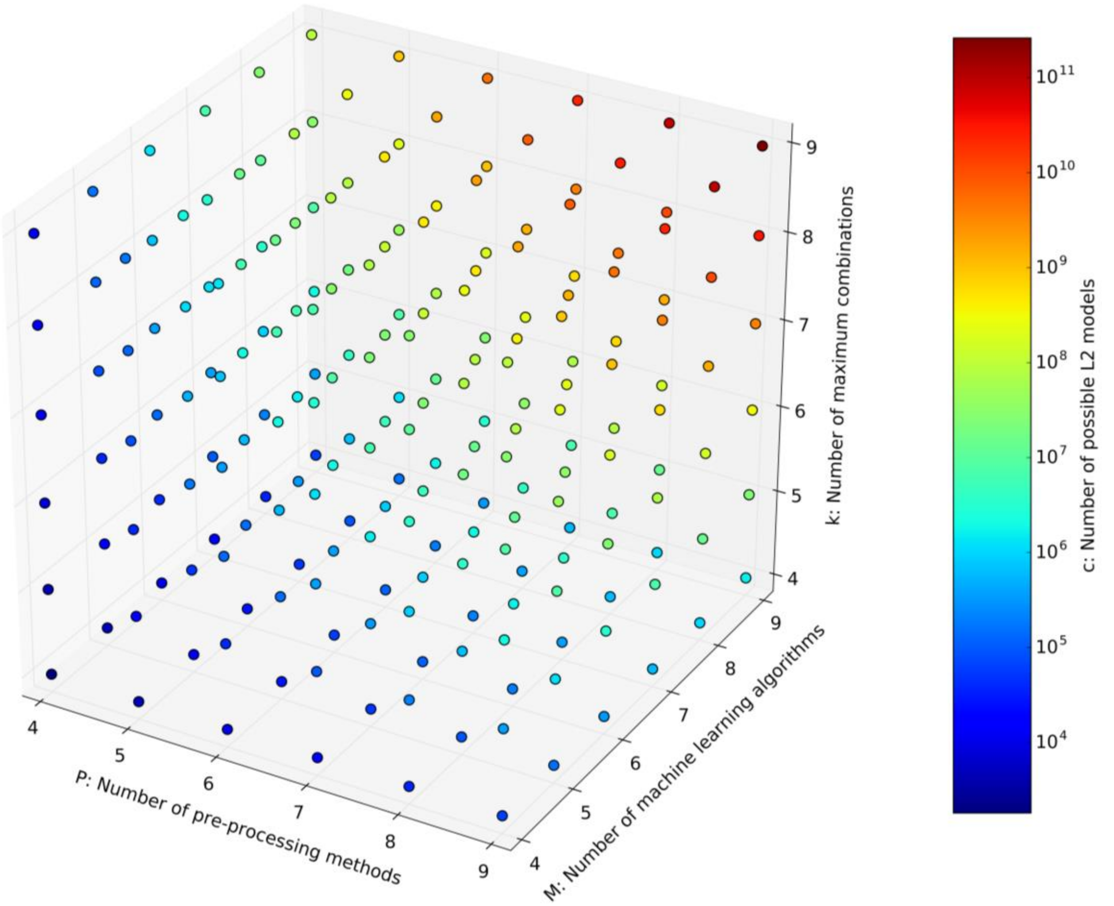

Figure 2-84. The number of possible L2 models for various values of P, M, and k.

A high-level overview of the GA-based stacking algorithm is presented below:

1. Define and apply all P pre-processing methods on the initial spectra
2. Define all M_{L_1} machine learning algorithms, and build all $P \times M_{L_1}$ L1 models using the calibration data
3. Define k as the maximum number of L1 models which can be combined by the L2 model, and all M_{L_2} machine learning algorithms.
4. Randomly generate an initial population of NP chromosomes, each fully defining a L2 model, and calculate their respective fitness
5. For each of the training generations repeat:
    a. Generate NP offspring using mutation and two-point modulo crossover while also maintaining diversity in the population
    b. Evaluate the fitness of the offspring

c. Perform an elitist selection on the joint population of parent and offspring chromosomes to create the new population of size NP

6. Output the best chromosome (L2 model) of the population and use it to predict the testing data

The advantage of the algorithm is that it combines the L1 models that are built either way by the experts, who test different combinations to identify the best model; this is the approach followed and detailed in Section 2.4.5. In other words, it does not add a significant computational cost and the methodology builds upon what is contemporarily used.

Once a L2 model configuration is established, new testing patterns are predicted per Figure 2-85, i.e. the constituent L1 models are used to produce their predictions which are served as input to the stacking model (L2 model) which creates the final estimation of the output.

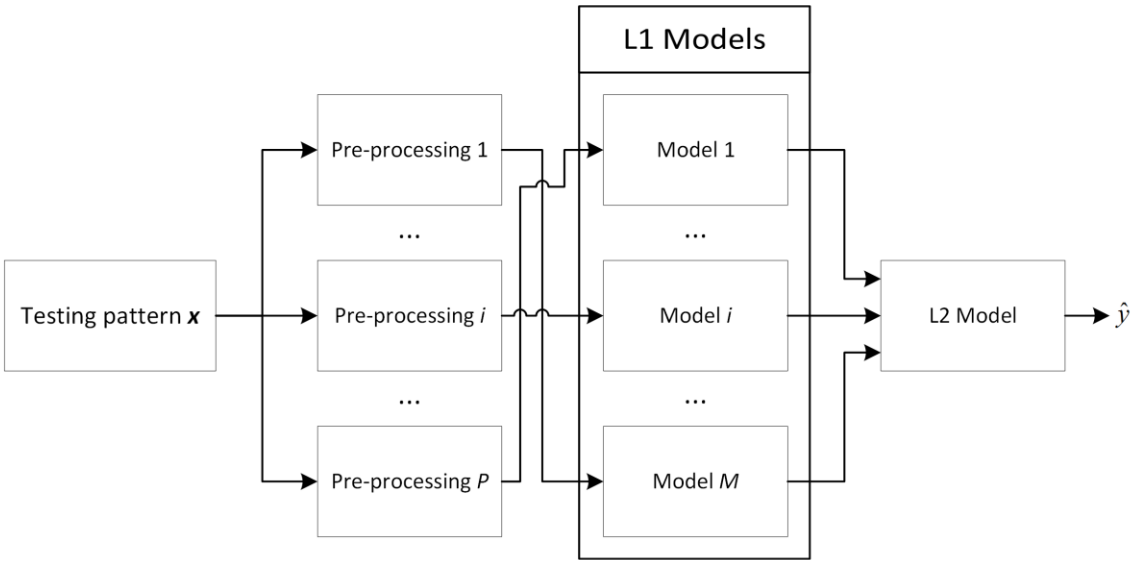

Figure 2-85: The prediction of a testing pattern according to the algorithm – once a configuration of the L2 model is established, the predictions of the constituent models are utilized to produce a new estimation

This novel algorithm was developed and applied in the GEO-CRADLE SSL to predict three common soil properties, namely the Organic Matter, the Clay Fraction, and calcium carbonate ($CaCO_3$). Table 2-17 presents the results of the algorithm. It must be noted that the results were improved over the best models (see Table 2-14 for the base L1 models results). One can observe the ability of the L2 model to offset large prediction errors and establish better predictions.

Table 2-17. Results of the GA-based stacking algorithm

Property	L2 model	No. of L1 models	RMSE	R^2	RPIQ
Clay Fraction	PLS	5	7.20	0.84	1.47
OM	PLS	4	0.47	0.58	2.01
$CaCO_3$	Cubist	4	1.24	0.67	0.56

2.5. Feasibility study – The case of the Drin River basin

2.5.1. **Description of the task**

2.5.1.1. *Scope*

The main scope of this study was to explore the capability of in situ sampling, open soil spectral libraries, and ancillary satellite image data fusion to produce high resolution soil thematic maps, in a catchment basement scale. More specifically, the Drin river basin inside the Albanian territory was considered, and the properties of soil moisture and soil clay content were chosen between others because of their importance in the development of hydrological models. The following subsections will highlight how the developed GEO-CRADLE SSL can be exploited by other users in the future to produce relevant EO products.

2.5.1.2. *The Drin river basin*

The Drin is the longest river in Albania, having a total length of 335 km, 285 of which flow within. The river has two distributaries; one flowing directly into the Adriatic Sea in the west, the other one into the Buna River. Its catchment area concludes an area of 19,686 square kilometres. It includes the Black Drin, which drains from Lake Ohrid and flows northwards to the outskirts of Kukës, where it merges with White Drin and forms the Drin River. The river basin is one of the most biodiverse hotspots in Europe.

The extended Drin Basin:

- Extends to Albania, FYROM, Greece, Kosovo, and Montenegro (Drin Riparians)
- Comprises the sub–basins of five water bodies, each one of them shared by two or three Riparians:
    - Two rivers: the Drin and its two major tributaries, the Black Drin and the White Drin, as well as the Buna/Bojana River
    - Three lakes: Prespa, Ohrid and Skadar/Shkoder
- Covers a geographical area of about 19,000 km^2
- Is characterized by mountainous relief - the highest peaks are over 2,500 m - with flat land in the coastal area.

- Hosts about 1.5 million people who rely on the water resources of the basin for drinking water, agriculture, fisheries, industry, hydropower and for supporting the provision of tourism services.
- Encompasses complex ecosystems of major importance that host unique biotopes with many indigenous species, important both from a European and an international conservation perspective.
- Sustains and affects coastal and marine ecosystems in the Adriatic, through its freshwater flow.

The Drin River is the “connecting body” of the extended Drin Basin, linking the lakes, wetlands, rivers and other aquatic habitats, with a system of groundwater bodies, into a single, yet complex, hydrological ecosystem.

The basin is depicted in Figure 2-86. Setting out from the two Prespa Lakes, linked to each other by a small channel, water flows through underground karst cavities to Lake Ohrid, the largest lake in terms of water volume in South East Europe. The only surface outflow of Lake Ohrid, the Black Drin River flows north through FYROM and enters Albania. The White Drin River rises in Kosovo, flows into Albania, where it meets the Black Drin and forms the Drin River. Flowing westward through Albania, the Drin River meets the Buna/Bojana River, close after the outflow of the latter from Lake Skadar/Shkoder, the largest lake in terms of surface in South East Europe. The Buna/Bojana River directly discharges into the Adriatic Sea.

The area examined in this pilot and in the subsequent sections covered the basin located within the geographical area of Albania.

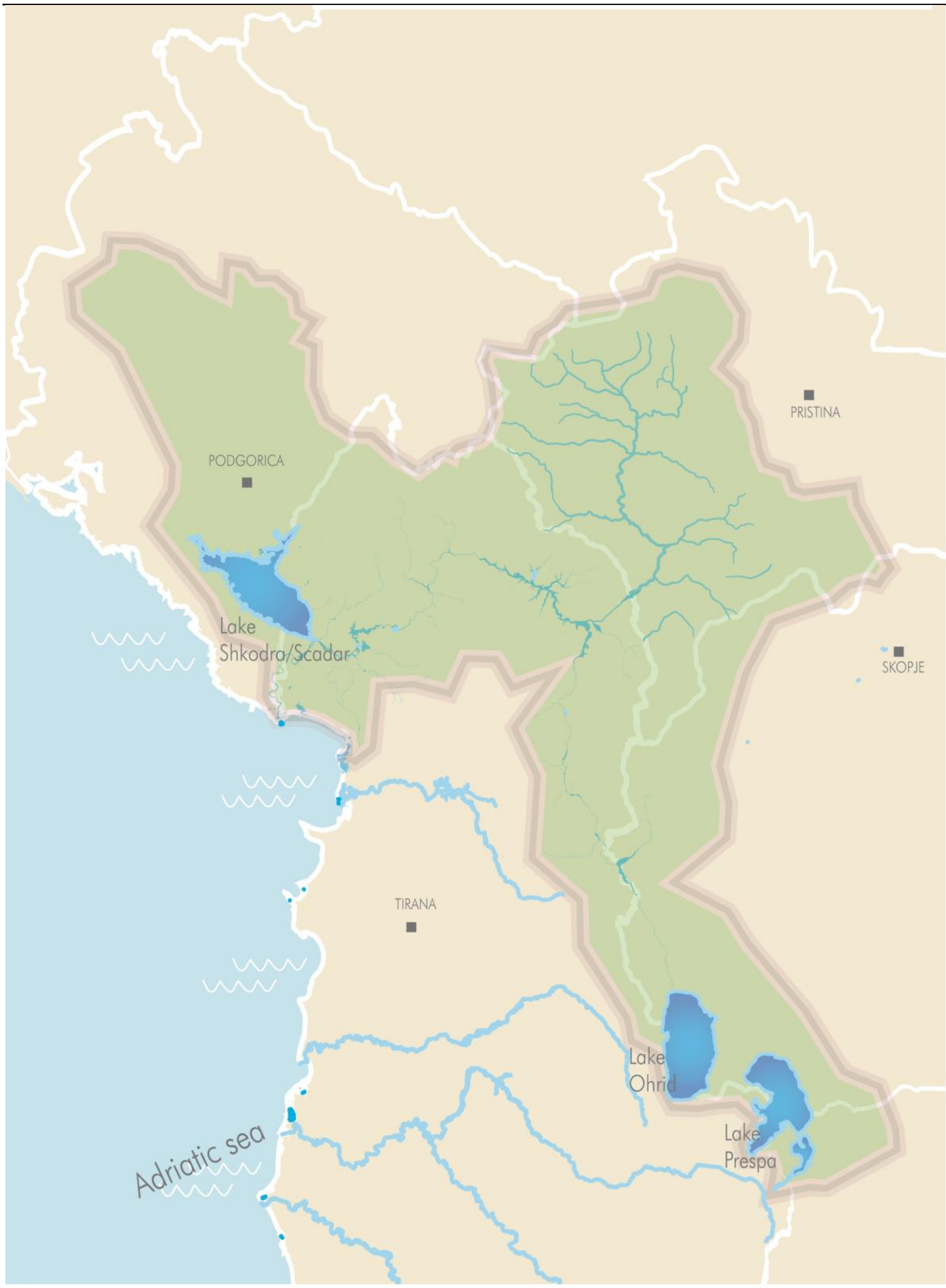

Figure 2-86. The Drin River basin

2.5.1.3. *Overview*

To derive EO-driven soil thematic maps, it is necessary to have the following layers of observation (visualized in Figure 2-87):

- Ground reference data, containing the geo-referenced ground truth for the examined properties, coupled with their spectral signatures
- Earth observation data of the same period of the sampling data
- Auxiliary data which may assist the prediction of the properties

Typically, there are four major groups of the auxiliary information: climate, organism, relief, parent material and time. McBratney et al. [44] further added to this list the geographical location of the soil profiles and the available soil properties that show correlation with the ones to be estimated. These are the major inputs of a statistical framework – also known as SCORPAN – used to predict soil variables at each location of the study of interest. SCORPAN is a conceptual model of soil spatial inference.

A common spatial prediction technique that can be used to apply SCORPAN model is the regression-kriging, which we use to illustrate the general flow of data through the system to estimate the unknown soil parameters. These models assume that there is a stochastic relationship between various predictors and target soil variables, although it can also be used to improve the deterministic models of soil genesis [45].

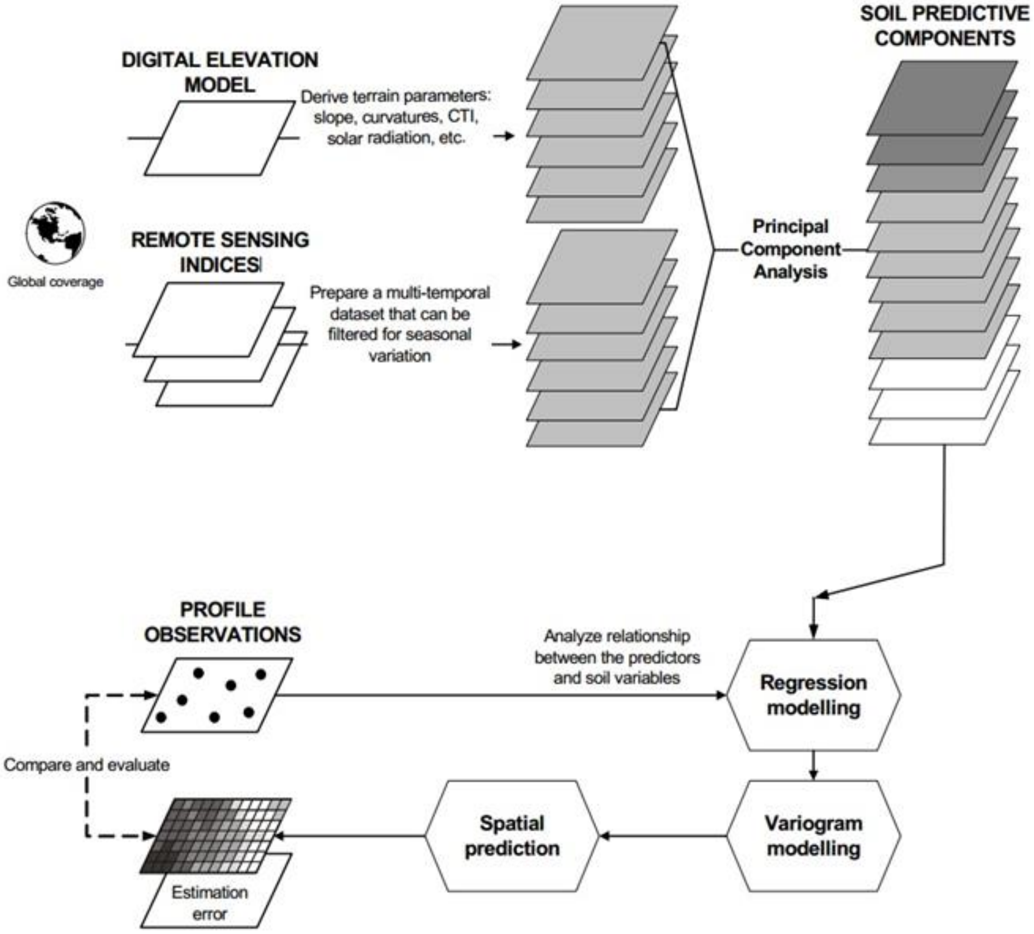

Figure 2-87. The high-level overview of the approach needed to derive EO-based maps

2.5.2. **Sampling planning**

For the implementation of the case study of soil thematic mapping in the Drin basin, a soil sampling field campaign was scheduled to take place during the mid-July – August time frame. This was important as ground truth data are imperative in order to calibrate the models.

A preliminary sampling planning was considered, and for that reason a Sentinel 2 satellite multispectral image of early July date was analysed. Land cover classification indicated the bare soil sites within the basin. The basic principle that dictated the soil sampling procedure was that it would be conducted on lowland flat areas, with low terrain slope, on sites with bare soil land cover, agricultural land use and a certain variation of Soil Taxonomy. The concept behind it is that these are regions that mostly face serious flooding issues, and alongside, display a visible soil colour range, on the satellite image. A stratified random sampling approach was followed.

In stratified sampling, the population is partitioned into non-overlapping groups, called strata and a sample is selected by some design within each stratum.

A stratification of the sampling was achieved, as far as the variability of the USDA Soil Taxonomy Classes involved, with the aid of a regional soil map layer provided by INCA. The abovementioned criteria were spatially implemented, with the aid of a Drin Bruna basin DEM and a CORINE Land Cover Map Layer, in the form of some initial sampling region polygons in Shkodra, Kukes, and Pogradec plains.

These sites were extracted from the satellite image, and their spectral variability in the Vis –NIR range was examined in comparison to Munsell soil color system. An ISODATA unsupervised classification of the multispectral image resulted in the segmentation and subdivision of the initial sampling regions into 3 spectral clusters. As a result, a random point set was created for each sampling region and each spectral cluster. Some further modifications of the sampling pattern, involved the reposition of some sampling points in order to achieve a better accessibility for the sampling field team. Ultimately the sampling planning resulted in 150 sampling locations within the Drin river basin.

2.5.3. **Soil Field Survey**

Soil Sampling was performed by INCA, on the Shkodra, Kukes and Pogradec regions, and the field campaign lasted from 15 July to 04 August. Sampling was performed with an Edelman soil auger, and sampling points were logged in a GNSS - DGPS with submeter accuracy. At each recorded location, 3 augers for a subsample of 20 gr of soil, were obtained from the topsoil layer (0-30 cm), within a 5m radius. Additionally, soil color was noted according to a Munsell soil color chart. Subsamples were mixed to form the sample. Sampling coding followed the nomenclature used in the whole spectral library project (County Code – Soil Taxonomy Code – Soil Horizon Code – Sampling ID, e.g AL-LV-030-00123). At every site, the sample was measured for total wet weight, and in situ conditions of Air Temperature and Relative Humidity were noted. Due to field accessibility reasons the total number of samples was reduced to 107.

Testing the actual sampling dataset for Complete Spatial Randomness of the pattern, an Average Nearest Neighbor z –score testing was performed for Shkodra, Kukes and Pogradec respectively.

The spatial pattern analysis of the results showed that:

- At Shkodra the sampling presented a spatially clustered pattern, at a 10% significance level (Figure 2-88).
- At Kukes the sampling presented a spatially dispersed pattern, at a 5 % significance level (Figure 2-89).
- At Pogradec the sampling presented a complete spatial random motif (Figure 2-90).

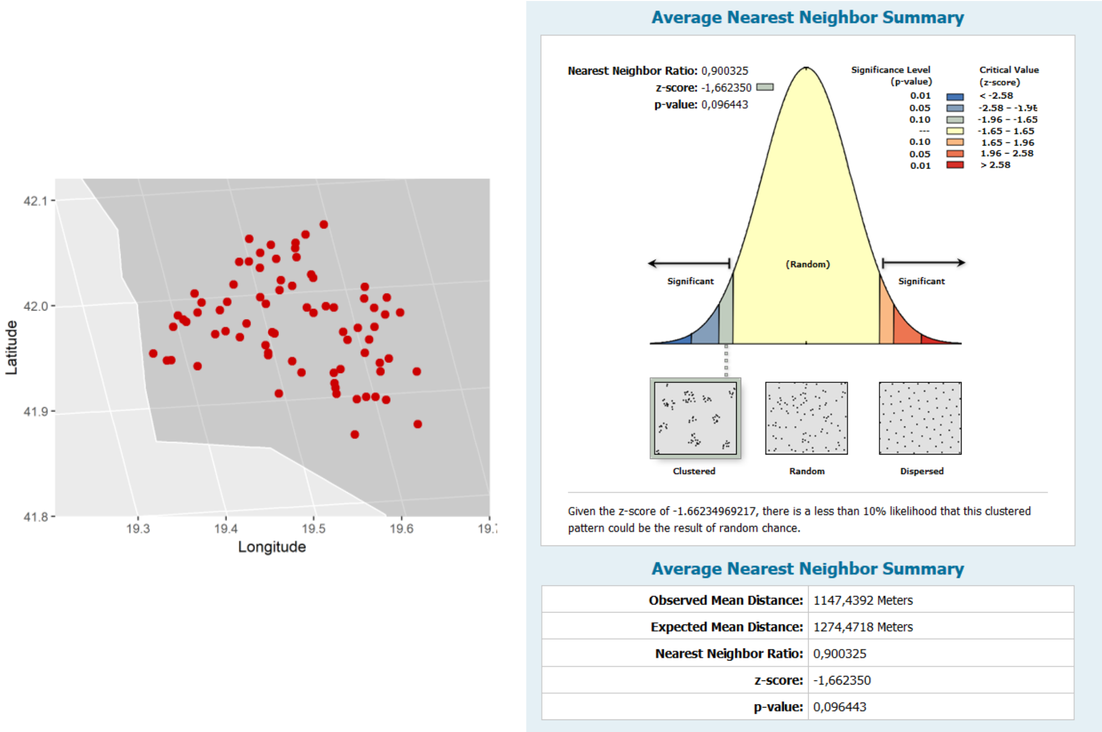

Figure 2-88. The sampling points of the Shokdra region (left) and the spatial pattern analysis (right)

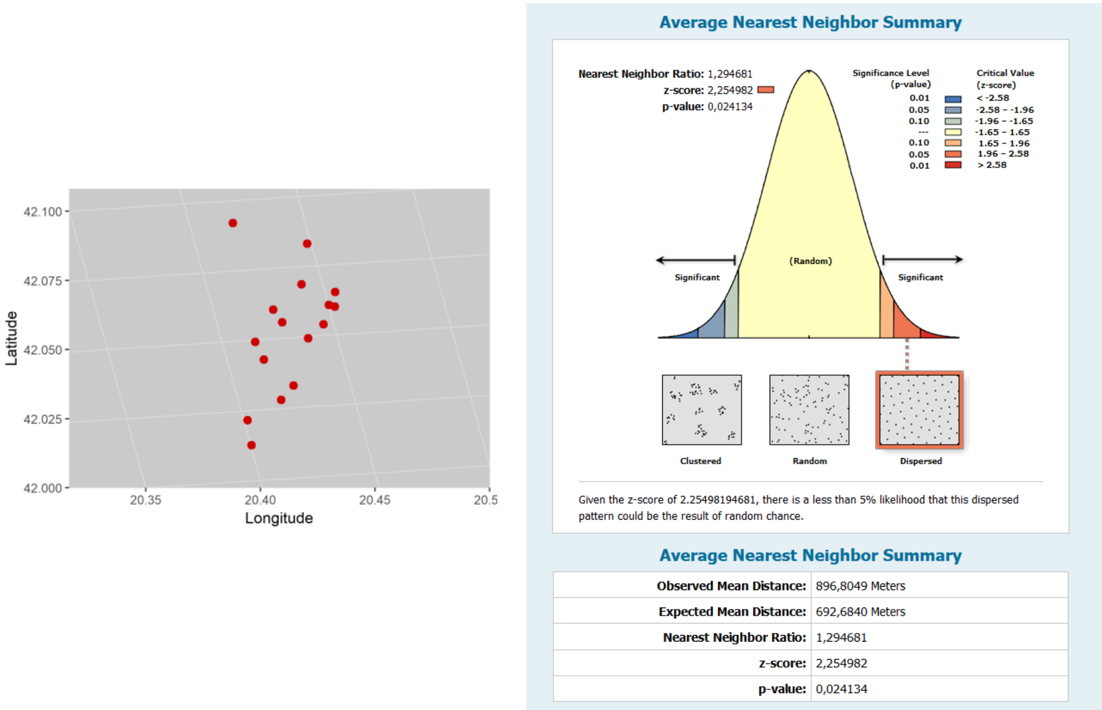

Figure 2-89. The sampling points of the Kukes region (left) and the spatial pattern analysis (right)

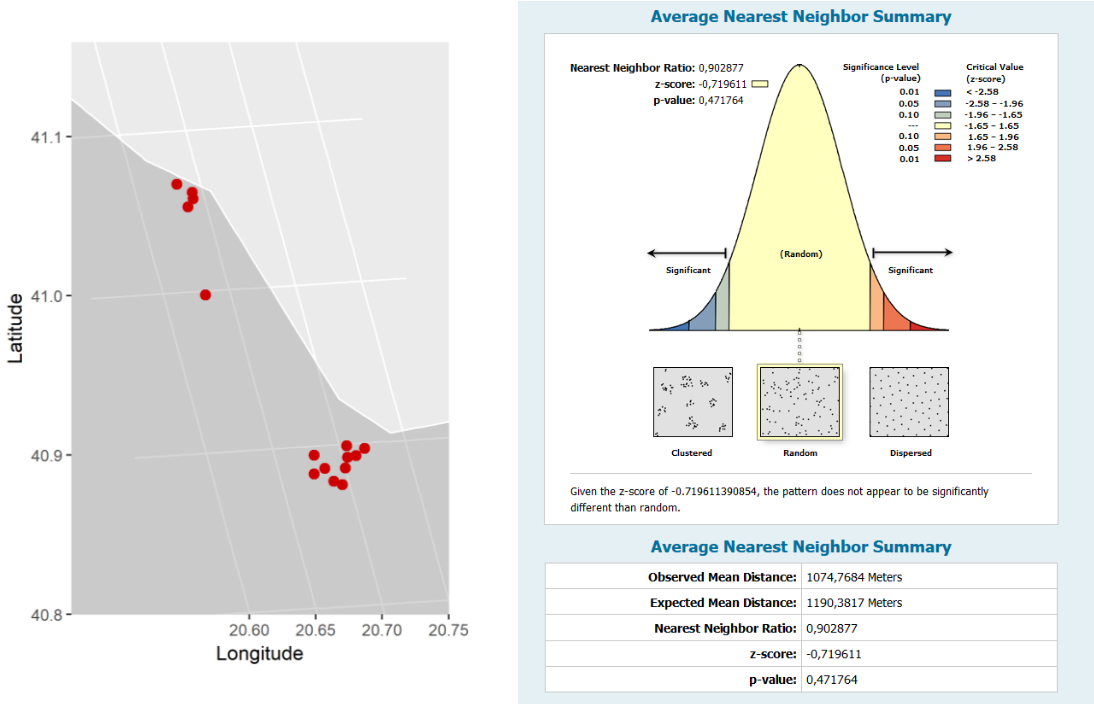

Figure 2-90. The sampling points of the Pogradec region (left) and the spatial pattern analysis (right)

2.5.4. **Earth Observation data considered**

The Earth Observation data considered for this study are the multi-spectral maps of Sentinel-2 (Copernicus data) and of Landsat-8 (NASA data). More concretely, the use of these specific products was decided, due to their individual advantages they offer for soil mapping applications. The soil Clay composition mas mapped with ancillary Sentinel 2 images, to make use of a more detailed spectral resolution on the Red/ Red Edge and NIR parts of the Soil Spectral Curve (Bands 4/5/6/7/8A), as well as the 2 SWIR Bands. On the contrary, soil Moisture mapping was based on the Mid Infrared absorption bands in combination with Land Surface Temperature information found on the Thermal Infrared part of the spectrum. For that reasons, the use of Landsat 8 data was preferred.

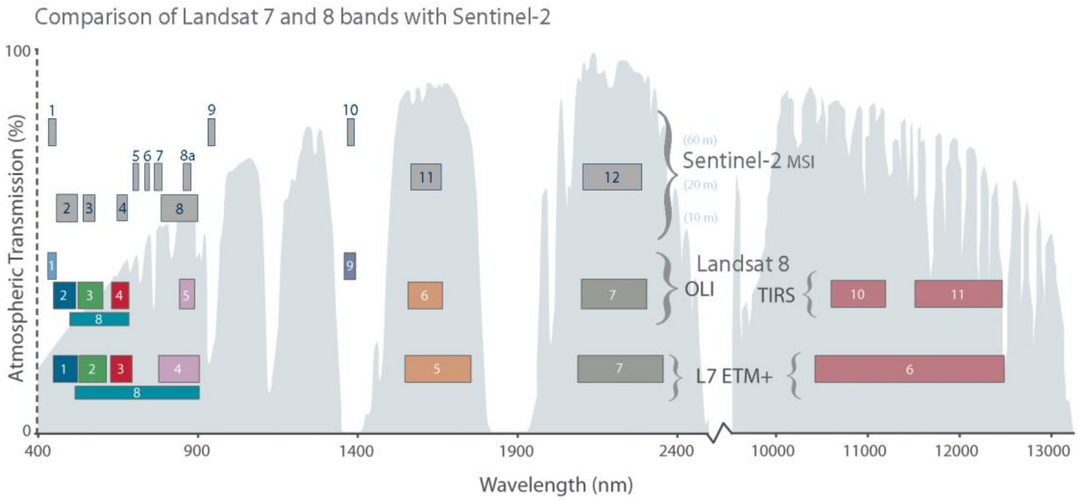

Figure 2-91. Comparison between the bands of Landsat 7 and 8 with the bands of Sentinel-2

2.5.4.1. *Sentinel-2*

Sentinel-2 is an Earth observation mission developed by ESA as part of the Copernicus Programme to perform terrestrial observations in support of services such as forest monitoring, land cover changes detection, and natural disaster management. It consists of two identical satellites built by Airbus DS, Sentinel-2A and Sentinel-2B, with two additional satellites being constructed by Thales Alenia Space.

The Sentinel-2 mission has the following capabilities:

- Multi-spectral data with 13 bands in the visible, near infrared, and short-wave infrared part of the spectrum
- Systematic global coverage of land surfaces from 56° S to 84° N, coastal waters, and all of the Mediterranean Sea
- Revisiting every 5 days under the same viewing angles. At high latitudes, Sentinel-2 swath overlap and some regions will be observed twice or more every 5 days, but with different viewing angles.
- Spatial resolution of 10 m, 20 m and 60 m
- 290 km field of view
- Free and open data policy

To achieve frequent revisits and high mission availability, two identical Sentinel-2 satellites (Sentinel-2A and Sentinel-2B) are operating simultaneously. The orbit is Sun synchronous at 786 km (488 mi) altitude, 14.3 revolutions per day, with a 10:30 a.m. descending node. This local time was selected as a compromise between minimizing cloud cover and ensuring suitable Sun illumination. It is close to the Landsat local time and matches SPOT's, allowing the combination of Sentinel-2 data with historical images to build long-term time series.

The Sentinel-2 satellites each carry a single multi-spectral instrument (MSI) with 13 spectral channels in the visible/near infrared (VNIR) and short-wave infrared spectral range (SWIR), with the complete spectral configuration presented in Table 2-18.

Table 2-18. The spectral configuration of the on-board MSI of Sentinel-2

Sentinel-2 Bands		Central Wavelength (nm)	Resolution (m)	Bandwidth (nm)
Band 1	Coastal aerosol	0.433	60	20
Band 2	Blue	0.490	10	65
Band 3	Green	0.560	10	35
Band 4	Red	0.665	10	30
Band 5	Vegetation Red Edge	0.705	20	15
Band 6	Vegetation Red Edge	0.740	20	15
Band 7	Vegetation Red Edge	0.783	20	20
Band 8	NIR	0.842	10	115
Band 8A	Narrow NIR	0.865	20	20
Band 9	Water vapor	fF0.945	60	20
Band 10	SWIR-Cirrus	1.375	60	20
Band 11	SWIR	1.610	20	90
Band 12	SWIR	2.190	20	180

The Sentinel Level-2A product offers bottom of atmosphere reflectances in cartographic geometry, performing appropriate atmospheric correction through ATCOR code. This product is currently processed on the user side by using a processor running on ESA's Sentinel-2 Toolbox. The possibility of making a standard core product systematically available from the Sentinels core ground segment is currently being assessed as part of the CSC evolution activities.

2.5.4.2. *Landsat-8*

Landsat 8 is an American Earth observation satellite launched on February 11, 2013. It is the eighth satellite in the Landsat program; the seventh to reach orbit successfully. Originally called the Landsat Data Continuity Mission (LDCM), it is a collaboration between NASA and the United States Geological Survey (USGS). NASA Goddard Space Flight Center in Greenbelt, Maryland, provided development, mission systems engineering, and acquisition of the launch vehicle while the USGS provided for development of the ground systems and will conduct on-going mission operations

Landsat 8's Operational Land Imager (OLI) improves on past Landsat sensors and was built, under contract to NASA, by Ball Aerospace. OLI uses a technological approach demonstrated by the Advanced Land Imager sensor flown on NASA's experimental EO-1 satellite. The OLI instrument uses a push-broom sensor instead of whiskbroom sensors that were utilized on earlier Landsat satellites. The push-broom sensor aligns the imaging detector arrays along Landsat 8's focal plane allowing it to view across the entire swath, 115 miles (185 kilometres) cross-track field of view, as opposed to sweeping across the field of view. With over 7,000 detectors per spectral band, the push-broom design results in increased sensitivity, fewer moving parts, and improved land surface information.

OLI collects data from nine spectral bands (Table 2-19). Seven of the nine bands are consistent with the Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors found on earlier Landsat satellites, providing for compatibility with the historical Landsat data, while also improving measurement capabilities. Two new spectral bands, a deep blue coastal / aerosol band and a shortwave-infrared cirrus band, will be collected, allowing scientists to measure water quality and improve detection of high, thin clouds.

Table 2-19. The spectral configuration of the on-board OLI of Landsat 8

OLI Bands		Wavelength (nm)	Resolution (m)
Band 1	Coastal aerosol	433-453	30
Band 2	Blue	450-515	30
Band 3	Green	525-600	30
Band 4	Red	630-680	30
Band 5	Near Infrared	845-885	30
Band 6	Short Wavelength Infrared	1560-1660	30
Band 7	Short Wavelength Infrared	2100-2300	30
Band 8	Panchromatic	500-680	15
Band 9	Cirrus	1360-1390	30

The Thermal InfraRed Sensor (TIRS), built by the NASA Goddard Space Flight Center, conducts thermal imaging and supports emerging applications such as evapotranspiration rate measurements for water management. The TIRS focal plane uses GaAs Quantum Well Infrared Photodetector arrays (known as QWIPs) for detecting the infrared radiation—a first for the Landsat program. The TIRS data will be registered to OLI data to create radiometrically, geometrically, and terrain-corrected 12-bit Landsat 8 data products. Like OLI, TIRS employs a push-broom sensor design with a 185-kilometre swath width. Data for two long wavelength infrared bands will be collected with TIRS. This provides data continuity with Landsat 8's single thermal IR band and adds a second.

Table 2-20. The spectral configuration of the on-board TIRS of Landsat 8

TIRS Bands		Wavelength (μm)	Resolution (m)
Band 10	Thermal Infrared	10.30-11.30	100
Band 11	Thermal Infrared	11.50-12.50	100

The U.S. Geological Survey (USGS) offers on-demand production of Landsat 8 Operational Land Imager/Thermal Infrared Sensor (OLI/TIRS) Surface Reflectance data through EarthExplorer. Surface Reflectance products provide an estimate of the surface spectral reflectance as it would be measured at ground level in the absence of atmospheric scattering or absorption. The Surface Reflectance products are generated at the Earth Resources Observation and Science (EROS) Centre at a 30-meter spatial resolution. The EROS Science Processing Architecture (ESPA) on-demand interface corrects satellite images for atmospheric effects to create Level-2 data products. Landsat 8 Surface Reflectance data are generated from the Landsat Surface Reflectance Code (LaSRC). LaSRC makes use of the coastal aerosol band to perform aerosol inversion tests, uses auxiliary climate data from MODIS and uses a unique radiative transfer model. Additionally, LaSRC hardcodes the view zenith angle to “0”, and the solar zenith and view zenith angles are used for calculations as part of the atmospheric correction.

2.5.4.3. *Land Cover Classifications – Cloud Masks*

Both Sentinel 2 and Landsat 8 Level 2 Data Products include scene Land Cover classification layers, that were used in this study for the identification of the basic land cover types (Water, Vegetation, Bare Soil, Urban) and the creation of cloud cover mask layers

2.5.4.4. *Data Averaging*

In order to get a mean value of the Soil Clay/ Moisture over the sampling survey timespan, two image mosaics for each satellite platform were created, and were ultimately averaged to create Landsat 8 / Sentinel 2 mean reflectance images (named $L8_{mr}$ and $S2_{mr}$).

2.5.4.5. *Earth Observation data derivative products*

Prior to processing of the raw EO data, it is important to first apply data transformation techniques which elucidate the underlying information. We derived a number of data products, which are listed below [46], [47].

Land Surface Temperature from Landsat 8 images

The brightness temperature is a measurement of the radiance of the microwave radiation traveling upward from the top of the atmosphere to the satellite, expressed in units of the temperature of an equivalent black body. The brightness temperature (or TB) is the fundamental parameter measured by passive microwave radiometers. The brightness temperatures, measured at different microwave frequencies, are used at Remote Sensing Systems to derive wind, vapor, cloud, rain, LST and SST products.

Conversion to Top of Atmosphere Radiance

OLI and TIRS band data can be converted to TOA spectral radiance using the radiance rescaling factors provided in the metadata file:

$$L_\lambda = M_L Q_{cal} + A_L$$

where:

L_λ	=	TOA spectral radiance (Watts / (m^2 × srad × µm))
M_L	=	Band-specific multiplicative rescaling factor from the metadata (RADIANCE_MULT_BAND_x, where x is the band number)
Q_{cal}	=	Quantized and calibrated standard product pixel values (DN)
A_L	=	Band-specific additive rescaling factor from the metadata (RADIANCE_ADD_BAND_x, where x is the band number)

Conversion to Top of Atmosphere Brightness Temperature

TIRS band data can be converted from spectral radiance to top of atmosphere brightness temperature using the thermal constants provided in the metadata file:

$$T = \frac{K_2}{\ln(\frac{K_1}{L_\lambda} + 1)}$$

where:

T	=	Top of atmosphere brightness temperature (K)
L_λ	=	TOA spectral radiance (Watts / (m^2 × srad × μm))
K_1	=	Band-specific thermal conversion constant from the metadata (K_1_CONSTANT_BAND_x, where x is the thermal band number)
K_2	=	Band-specific thermal conversion constant from the metadata (K_2_CONSTANT_BAND_x, where x is the thermal band number)

Reclassification of Land Cover Classification to Emissivity Values

Several studies used NDVI for the estimation of land surface emissivity [48]; other studies used a land cover classification for the definition of the land surface emissivity of each class [49] .For instance, the emissivity (e) values of various land cover types are provided in Table 2-21 (from [50]).

Table 2-21. Emissivity values for different land surface types

Land surface type	Emissivity (ε)
Water	0.980
Built-up	0.937
Vegetation	0.982
Bare soil	0.928

Estimation of Land Surface Temperature

Several studies have described the estimation of Land Surface Temperature. Land Surface Temperature can be calculated from At-Satellite Brightness Temperature TB as in [49] :

$$T = \frac{TB}{1 + \left(\lambda \times \frac{TB}{c_2}\right) \times \ln(\varepsilon)}$$

where

λ Wavelength of emitted radiance

c_2	$h \times c/s$	$1.4388\ 10^{-2}$ m K
h	Planck's constant	$6.626\ 10^{-34}$ J s
s	Boltzmann constant	$1.38\ 10^{-23}$ J/K
c	Velocity of light	$2.998\ 10^{8}$ m/s

The values of λ for Landsat bands are listed in Table 2-22.

Table 2-22. Central wavelength of Landsat bands

Satellite	Band	λ (μm)
Landsat 4, 5, and 7	6	11.45
Landsat 8	10	10.8
Landsat 8	11	12

Spectral Transformation – Principal Component Analysis

PCA is used to transform the data in the input bands from the input multivariate attribute space to a new multivariate attribute space whose axes are rotated with respect to the original space. The axes (attributes) in the new space are uncorrelated. The main reason to transform the data in a principal component analysis is to compress data by eliminating redundancy.

Conceptually, using a two-band raster, the shifting and rotating of the axes and transformation of the data is accomplished as follows:

- The data is plotted in a scatterplot.
- An ellipse is calculated to bound the points in the scatterplot Boundary of ellipse plotted
- The major axis of the ellipse is determined. The major axis becomes the new x-axis, the first principal component (PC1). PC1 depicts the greatest variation because it is the largest transect that can be drawn through the ellipse. The direction of PC1 is the eigenvector, and its magnitude is the eigenvalue. The angle of the x-axis to PC1 is the angle of rotation that is used in the transformation.
- An orthogonal line perpendicular to PC1 is calculated. This line is the second principal component (PC2) and the new axis for the original y-axis (see the figure below). The new axis describes the greatest variance not described by PC1.

Using the eigenvectors, the eigenvalues, and the calculated covariance matrix of the input of the multiband raster, a linear formula defining the shift and rotation is created. This formula is applied to transform each cell value relative to the new axis. For the decorrelation of the initial EO data, bands 2,3,4,6,8, 11, and 12 from Sentinel 2 were factorized into 7 PCs. Similarly, bands 2,3,4,5,6,7, and Land Surface Temperature from thermal 10 and 11 were transformed into 8 PCs. These transformations were calculated on the whole raster, with presenting each transformation's eigenvalues and explained variance for Sentinel-2 and Landsat-8, respectively.

Table 2-23. Sentinel-2 Principal Components

Sentinel 2 PCA	EigenValue	Explained Variance %	Accumulative Explained Variance %
S2_PC1	1.92E+01	96.9578	96.9578
S2_PC2	5.38E-01	3.0313	99.9891
S2_PC3	1.63E-03	0.0095	99.9986
S2_PC4	1.88E-04	0.0011	99.9997
S2_PC5	3.98E+00	0.0002	99.9999
S2_PC6	1.47E+00	0.0001	100
S2_PC7	6.44E-01	0	100

Table 2-24. Landsat-8 Principal Components

Landsat 8 PCA	EigenValue	Explained Variance %	Accumulative Explained Variance %
L8_PC1	2.22E+01	90.88	90.88
L8_PC2	4.12E+00	5.32	96.20
L8_PC3	4.68E-01	2.87	99.07
L8_PC4	4.64E-02	0.92	99.99
L8_PC5	4.99E-05	0.01	100.00
L8_PC6	1.76E-05	0.00	100.00
L8_PC7	8.56E-06	0.00	100.00
L8_PC8	1.40E-06	0.00	100.00

Spectral transformation – the Tasseled-Cap Transform

The Tasseled Cap transformation [51] is designed to analyse and map vegetation and urban development changes detected by various satellite sensor systems. It is known as the Tasseled Cap transformation due to the shape of the graphical distribution of data. It was developed in 1976 by R.J. Kauth and G.S. Thomas of the Environmental Research Institute of Michigan (ERIM). In the paper (Kauth and Thomas, 1976), the researchers provided a rationale for the patterns found in Landsat MSS data of agricultural fields as a function of the life cycle of the crop. Essentially, as crops grow from seed to maturity, there is a net increase in near-infrared and decrease in red reflectance based on soil colour.

The utility of this transformation has expanded from monitoring crops to analysing and mapping vegetation to support a variety of applications such as forestry, industrial vegetation management, ecosystem mapping and management, inventory and monitoring for carbon sequestering and credits, urban development, and more.

In remote sensing analysis it is common to ratio and plot different combinations of multispectral bands to examine relationships between the bands. The Tasseled Cap transformation is a special case of principal components analysis which transforms the image data to a new coordinate system with a new set of orthogonal axes.

The primary axis, called brightness, is statistically derived and is calculated as the weighted sum of reflectances of all spectral bands and accounts for the most variability in the image. Brightness is associated with bare or partially covered soil, man-made, and natural features such as concrete, asphalt, gravel, rock outcrops, and other bare areas.

Table 2-25. Weights for the Tasseled-Cap transformation

	Blue	**Green**	**Red**	**NIR**	**SWIR 1**	**SWIR 2**
Brightness	0.3037	0.2793	0.4343	0.5585	0.5082	0.1863
Greenness	-0.2848	-0.2435	-0.5436	0.7243	0.0840	-0.1800
Wetness	0.1509	0.1793	0.3299	0.3406	-0.7112	-0.4572

Orthogonal to the first component, the second component greenness is associated with green vegetation, while the third component wetness is orthogonal to the first two components and is associated with soil moisture, water, and other moist features.

The other additional components contain image noise and atmospheric influences, such as clouds, haze, sun angle differences, and so on, that have been removed from the first three more significant components. The first three components of the Tasselled Cap transformed imagery contain about 97 percent of the meaningful information available in the image.

2.5.5. **Other datasets considered**

2.5.5.1. *Soil Moisture*

Soil moisture is an important component in the atmospheric water cycle, both on a small agricultural scale and in large-scale modelling of land/atmosphere interaction. Vegetation and crops always depend more on the moisture available at root level than on precipitation occurrence. Water budgeting for irrigation planning, as well as the actual scheduling of irrigation action, requires local soil moisture information.

Knowledge of the degree of soil wetness helps to understand the initiation of convective events, and to forecast the risk of flash floods, or the occurrence of fog.

2.5.5.2. *Gravimetric Soil Moisture Analysis*

The gravimetric soil moisture content is typically determined directly. Soil samples of about 50 g are removed from the field with the best available tools (shovels, spiral hand augers, bucket augers, perhaps power-driven coring tubes), disturbing the sample soil structure as little as possible [52]. The soil sample should be placed immediately in a leak-proof, seamless, pre-weighed and identified container. As the samples will be placed in an oven, the container should be able to withstand high temperatures without melting or losing significant mass. The most common soil containers are aluminium cans, but non-metallic containers should be used if the samples are to be dried in microwave ovens in the laboratory. If soil samples are to be transported for a considerable distance, tape should be used to seal the container to avoid moisture loss by evaporation.

The samples and container are weighed in the laboratory both before and after drying, the difference being the mass of water originally in the sample. The drying procedure consists in placing the open container in an electrically heated oven at 105°C until the mass stabilizes at a constant value. The drying times required usually vary between 16 and 24 h. Note that drying at 105°±5°C is part of the usually accepted definition of “soil water content”, originating from the aim to measure only the content of “free” water which is not bound to the soil matrix

Soil samples acquired from Drin Basin were measured in the lab for soil moisture content by the gravimetric method. The soil moisture content θ_d may be expressed by weight as the ratio of the mass of water present to the dry weight of the soil sample, or by volume as ratio of volume of water to the total volume of the soil sample. To determine any of these ratios for a particular soil sample, the water mass must be determined by drying the soil to constant weight and measuring the soil sample mass after and before drying. The water mass (or weight) is the difference between the weights of the wet and oven dry samples. The criterion for a dry soil sample is the soil sample that has been dried to constant weight in oven at temperature between 100 – 110 °C (105 °C is typical). It seems that this temperature range has been based on water boiling temperature and does not consider the soil physical and chemical characteristics. Wet soil weight was measured in situ, directly on the field with the use of an electrical balance with precision of 0, 01 g.

Gravimetric Soil Water Content is computed as:

$$\theta_d = \frac{\text{weight of wet soil} - \text{weight of dry soil}}{\text{weight of dry soil}}$$

2.5.5.3. *Digital Elevation Models and derived terrain parameters*

Relief or topography can be characterized with the use of digital elevation models (DEM). DEM is used to derive quantitative measures of soil forming processes, also called terrain parameterization [53]. This is a process of quantitative description of terrain by terrain parameters. These can be derived using various algorithms that quantify morphological, hydrological, ecological and other aspects of a terrain. In simple terms, terrain parameterization is extraction of terrain parameters using input digital elevation models and terrain parameterization software. Extracted terrain parameters can then be used, for example, to improve mapping and modelling of soils, vegetation, land use, geomorphologic and geological features and similar. There are relatively simple and easy to derive terrain parameters (the slope gradient, aspect, curvature) and there are some more complex ones which are derived with the combined use of the primary terrain parameters. The primary features are direct descriptors of the terrain features, like the slope, curvature or aspect, while secondary features describe more complex characteristics of the landform, which are linked to certain terrain-regulated processes, like stream power index or the compound topographic index (CTI).

The terrain defines the way how the water moves through the landscape and transport soil materials in solid or soluted forms. Thus, the variables, which controls the way of water flow have the greatest significance in explaining the spatial distribution of numerous soil properties. The majority of the studies use slope gradient, curvature and CTI variables among others, which are proved to describe these water-movement–controlled material transport through the landscape. Many of the soil landscape elements, variables have been translated to DEM-derivable format. There is a good and commonly accepted toolkit of digital terrain variables, but the need to develop new variables and approaches to improve our capability of soil-landscape modelling and decrease the unexplained portion of the soil-landscape relationship is still evident.

In this particular study, an ASTER Digital Elevation Model (GDEM V2) covering the Drin Bruna river basin, was used as primary Terrain Dataset. The DEM's spatial resolution is 15 m on the horizontal plane, resampled into a 20-m pixel size grid. From the elevation dataset 2 ancillary datasets were calculated: **slope,** and the **topographic wetness index**. The first one was calculated as the 1st derivative of the elevation, using GIS tools. The calculation of the second one is detailed below.

Topographic Wetness Index

The topographic wetness index (TWI), also known as the compound topographic index (CTI), is a steady state wetness index. It is commonly used to quantify topographic control on hydrological processes. The index is a function of both the slope and the upstream contributing area per unit width orthogonal to the flow direction. The index was designed for hillslope catenas. Accumulation numbers in flat areas will be very large, so TWI will not be a relevant variable. The index is highly correlated with several soil attributes such as horizon depth, silt percentage, organic matter content, and phosphorus. Methods of computing this index differ primarily in the way the upslope contributing area is calculated.

The topographic wetness index is defined as:

$$TWI = \ln \frac{a}{\tan b}$$

Where a is the local upslope area draining through a certain point per unit contour length and tan b is the local slope in radians. The TWI has been used to study spatial scale effects on hydrological processes. The topographic wetness index (TWI) was developed by Beven and Kirkby within the runoff model TOPMODEL. The topographic wetness index is unit less. TWI was utilized in the SAGA GIS environment, by the following steps:

- **Slope Derivation (b)**

The description of how the methods work are in reference to [54].

The computation of a slope parameter in a GIS environment is usually considered a trivial one-step algorithm that outputs visually appealing rasters. In SAGA GIS however, there are multiple methods for the calculation of slopes based on different considerations of geometry, polynomial order, and scale.

SAGA offers 8 different options for calculating slope:

- Maximum slope
- Maximum triangle slope
- Least squares fitted plane
- 6 parameter 2nd order polynomial
- 6 parameter 2nd order polynomial
- 6 parameter 2nd order polynomial
- **9 parameter 2nd order polynomial**
- 10 parameter 3rd order polynomial

The method used here is the one proposed by Zevenbergen & Thorne which is the 9 parameter 2nd order polynomial where slope is the first derivative of elevation. It uses a rectangular matrix of evenly spaced elevations that covers the entire area of the input layer where elevation points are used to determine the parameters in the quadratic equation. In other words, equation fitting is used to map out the surface and derive different indices (slope, aspect, curvature, etc.). This method uses an equation with 9 parameters to resolve for indices. It is more general than the other methods using 6 and more flexible (if a surface is of lower order than the equation, the corresponding coefficients will equal zero and won't influence the equation).

An advantage of this method is also the fact that it not only is able to yield an up slope area (drainage area contributing to a pixel) but also an up slope distance (longest travel path to an up slope divide).

- **Local Upslope Draining Area (a)**

- Sink Drainage Route Detection

This method is used to correct for dips and sinks inherent in the DEM. The algorithm works by identifying sinks or pits in the digital terrain and the direction that water flows out of the sink. Such sinks are usually depressions in the DEM that are lower than their surrounding pixels which introduce error while using flow routing algorithms in hydrological analysis

- Sink Removal
- Catchment Area Calculation

In their basic forms, catchment area methods use a flow accumulation algorithm that saturates a pixel depending on how many pixels upstream (area wise) lead into it across a topographic/directional gradient. The logic that determines how flow is transferred from one pixel to the other however is very variable and yield highly distinctive results.

The catchment area algorithm used in this workflow is recursive. This means it takes in the filled DEM and recursively processes all upward connected pixels until all pixels have been processed and a value is computed. The computation of the area/accumulation value is selected to be done through a Multiple Flow Direction method. The available options for flow accumulation method are D8, D Infinity, and Modified Flow Distribution (MFD) out of which the latter one was chosen for implementation.

MFD: this method computes the flow distribution with divergence. This means that instead of a total flow from one cell to the other based on gradient steepness, flow is distributed based on slope-weight basis and thus fractions of the flow are passed to neighboring cells differentially. This yields a more accurate representation of flow in variable terrain.

2.5.6. **Utilizing the regional soil spectral library**

The regional soil spectral library is comprised of laboratory spectra, allowing the estimation of soil variables at the sampling points. If airborne or satellite data are used, quantitative soil information over large areas may be provided. As noted before, the establishment of a stable and accurate SSL sufficiently sampling the covered region is not panacea. In particular, utilizing the database for its use in conjunction with satellite imagery should be done carefully, because there is a mismatch between the laboratory spectra and the satellite spectra. This discrepancy is due to the following factors:

- The spectral resolution is not the same – whereas laboratory spectra detailly cover the vis-NIR region with very fine spectral resolution, contemporary satellite data (e.g. Sentinel-2 data) have wider resolutions. Consequently, not all of the underlying information is sufficiently captured in satellite imagery. It is noted that the band width of the main airborne sensors is smaller than 40 nm, and thus their spectral range and accuracy is more suitable, if the use of airborne sensors is an option.
- Remote-sensing data are usually not sufficiently stable due to the change of the atmospheric and soil conditions between two different acquisition times. An additional predicament is the uncertainty inserted by the pre-processing procedures (radiometric, atmospheric, and geometric correction); these corrections usually cause non-systematic or random differences between the measurements.
- It is only possible to extract information about the topsoil (i.e. the upper layer). The upper layer of the soil is also not always directly visible by the satellite – depending on the time of the year and period, the soil might not be barren but riddled with weeds or plantations, which can obfuscate the spectra.

Notwithstanding the above, the predictive accuracy of EO data for soil properties can be quite substantial, underscoring the potential of EO data in this domain. To understand the impact of the regional SSL, first the process followed *without* a regional SSL must be described. The contemporary approach involves the collection of soil samples within bare fields, laboratory analytical measurement of the target variables, and calibration of a multivariate model linking the quantity of the target variable with the spectra extracted from the remote-sensing instrument at the sampling points. This procedure generally ensures a high estimation accuracy within the investigated area, but it entails time-consuming and expensive data collection (extensive sampling, and chemical or physical laboratory analysis). However, the calibration of such models allows only local efficiency and their application to different conditions or areas is often difficult or impossible, even for neighbouring fields.

In this context, the regional SSL can be exploited to decrease the efforts concerning the soil-variable estimation and obtaining more widely applicable models. A high-level overview of this approach is presented in Figure 2-92.

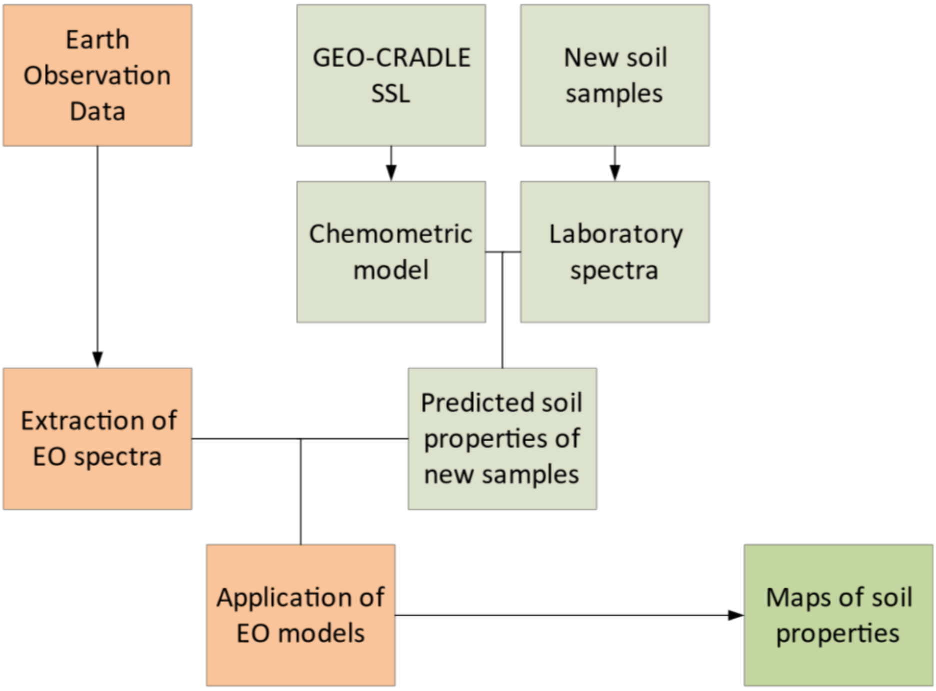

Figure 2-92. A flow-chart depicting how the GEO-CRADLE SSL can be used to derive EO-related soil thematic maps.

In short, the approach still relies on the collection of some new soil samples from the region that will be examined using Earth Observation need. It however circumvents the laborious and costly need to use a chemical laboratory every time a EO product must be produced. Instead, the soil properties of the samples are predicted using the established SSL. These are then correlated with the EO spectra, so that in turn the whole examined region may be mapped with sufficient accuracy for the desired soil properties.

To demonstrate this approach, some samples from the Albanian SSL were left-out of the GEO-CRADLE SSL. In other words, the dataset was split into calibration and validation. The calibration dataset was used in the development of the chemometric model. The model then predicted the outputs of the unknown validation sample. The joined set of calibration and validation was then used indistinguishably for the development of the Earth Observation models.

2.5.7. **Development of soil thematic maps**

2.5.7.1. *Spatial Data Models – Regression Kriging method*

In detail, for the Soil Clay and Moisture variables, values were regressed over the Predictor layers. Evaluating the regression model on the predictor layers gave the global trend surface of the soil variables over the study area. The significance of this trend depends on the R^2 of validation. According to the geostatistical theory of regionalized variable this trend constitutes the structural – deterministic part of the modelled property and lacks information about the local variability. This correlated variation is usually present in the regression residuals that make up the stochastic part of the modelled variable, with the precondition that they are spatially autocorrelated and follow a normal distribution.

Geostatistical interpolation methods like Regression – Kriging combine a regression of the dependent variable on auxiliary predictors with kriging of the regression residuals in order to handle the local variability present in the data [55], [56]. The following diagram shows a decision tree for selecting a suitable spatial prediction model. Consequently regression (OLS or GWR) residuals were tested for normality and spatial autocorrelation. Normality testing showed that the assumption of normality of residuals could not be rejected.

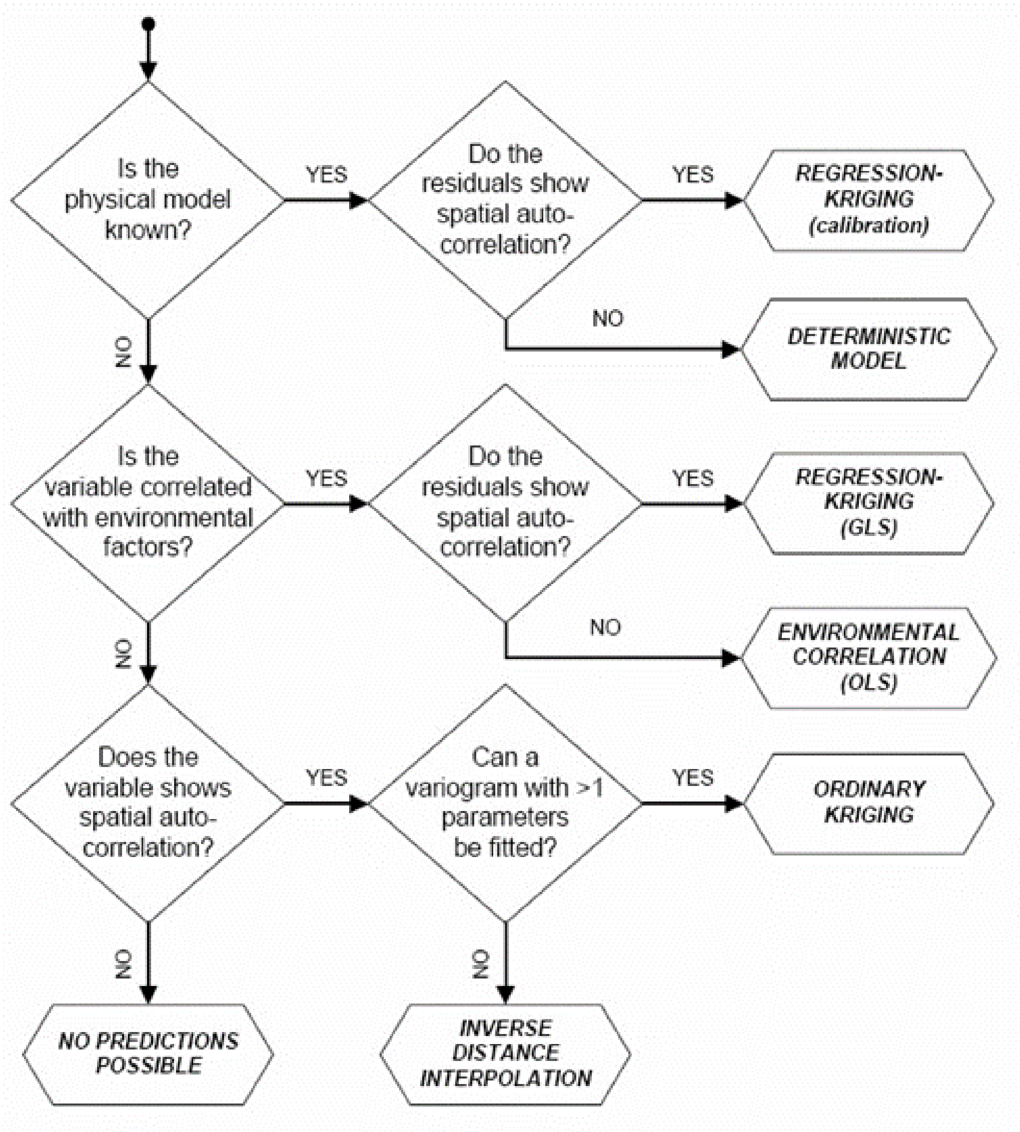

Figure 2-93. Decision tree for selecting a suitable spatial prediction model

In a spatial context, the sampling pattern was analysed, to detect its deviation from the Complete Spatial Random distribution (CSR), find out spatial clustering or dispersion, and derive the Average Nearest Neighbor (AvNN = 1147 for Shkodra, 896 for Kukes and 1076 for Pogradec). AvNN is a good measure of the spatial pattern density, necessary for the determination of spatial lags, and the cartographic scale of the data.

Each soil property was tested for spatial autocorrelation, in a global context with the use of Moran Index, and in a local scale with the use of incremental Moran correlograms. The aim of the analysis was to detect statistically significant spatial autocorrelation, a key property of geostatistics, as well as the lag distance within which it is observed.

Data stationarity, another geostatistics precondition, was tested with the graphical tool of Voronoi polygons, of mean and variance values to detect whether statistical measures remain constant in the study area, or if there exists some spatial trend in the data. Stationarity analysis showed that regression residuals had almost no further trend in them. Furthermore, the Voronoi tool, with the aid of the "cluster" command, showed some possible spatial outliers, in the sense of data points with highly different values for their neighbours. As a result from this analysis, a Simple Kriging of the Residual values was decided. The interpolated layer would be added to the regression layer in order to give the final Soil Property layer.

2.5.7.2. *Multivariate Data – Pairwise Correlation*

An initial pairwise correlation was calculated between all the inputs and the outputs. The inputs are: the initial satellite bands, their principal components, as well as the ancillary data used. The results can be found in Table 2-26. As far as the Clay Content is concerned, the best correlation is produced from the first PC and the TWI, indicating that a combined use of the spectral channels and local topography are the most important predictors. For soil moisture, a number of channels (in particular the SWIR and TIR channels), as well as the combination of the spectral channels via the PC transform produce the best correlations.

Table 2-26. Pairwise correlations between the bands, their PCs, and the ancillary data with the output for Clay Content (left) and Soil Moisture (right)

Soil Clay Content	
S2_BLUE	-0.0091
S2_GREEN	-0.0264
S2_RED	0.0385
S2_RED EDGE	0.0261
S2_NIR1	0.1036
S2_SWIR1	0.1072
S2_SWIR2	0.1469
S2_PC1	0.4127
S2_PC2	0.1350
S2_PC3	0.1415
S2_PC4	0.0234
S2_PC5	-0.1518
S2_PC6	-0.0294
S2_PC7	-0.0072
S2_TC_Br	0.0350
S2_TC_Gr	-0.0409
S2_TC_We	0.0649
Slope	0.2526
TWI	-0.4920
Elevation	0.4174

Soil Moisture	
L8_BLUE	-0.3169
L8_GREEN	-0.3519
L8_RED	-0.4531
L8_NIR	-0.0563
L8_SWIR 1	-0.4932
L8_SWIR 2	-0.5770
LST_L8_TIR 1	-0.4723
LST_L8_TIR 2	-0.5029
L8_PC1	-0.5123
L8_PC2	-0.3971
L8_PC3	-0.4748
L8_PC4	0.3173
L8_PC5	-0.0990
L8_PC6	0.0232
L8_PC7	0.2439
L8_PC8	-0.3338
L8_TC_Br	-0.3943
L8_TC_Gr	0.3655
L8_TC_We	0.5147
Slope	0.1336
TWI	0.2109
Elevation	-0.3757

2.5.7.3. *Regression Modelling*

Soil Moisture - MLR Model

The general model of Multiple Linear Regression, relating a response variable to several predictors by means of regression coefficients, has the following shape:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + f$$

The SVD (Singular Value Decomposition) algorithm is the most widely used algorithm to compute the estimated regression coefficients for MLR.

The Soil Moisture regression model was calibrated globally with the use of all 107 samples, using the first 5 Principal Components of the Landsat 8 mosaic as trend predictors. The MLR model was evaluated with a Leverage Correction validation method, an alternative to the Leave-One-out Cross Validation.

The leverage of an object, a sample or a variable, describes its influential X-"uniqueness" or its actual contribution to the calibration model. A leverage close to zero indicates that the corresponding sample or variable had very little importance for the calibration model.

For MLR, sample leverages are computed according to the following equation:

$$h_i = \frac{1}{Ic} + x_{S,i}^T (X_S^T X_S)^{-1} \quad X_{S,i} \quad , i = 1, \dots, Ic$$

The validation method Leverage Correction uses the leverages to estimate the prediction error without actually performing any predictions. The correction is done by correcting the y-residuals f with the sample leverage h_i:

$$f_{ij}^{\text{corrected}} = \frac{f_{ij}}{1 - h_i}$$

The results on the calibration and validation sets may be found in Table 2-27, where a complete analysis of variance (ANOVA) table is presented.

Table 2-27. Anova Table for the soil moisture regression

Anova Table							
Multiple Correlation			0.757 (cal)		0.727 (val)		
R-Square			0.573 (cal)		0.527 (val)		
RMSE			3.4393 (cal)		3.6100 (val)		
	SS	**df**	**MS**	**F ratio**	**p value**	**B-coefficients**	**STDerr**
Summary							
Model	13590.14	5	2718.0280	216.7360	0.0000		
Error	1040.88	83	12.5410				
Total	14631.02	88	166.2620				
Variable							
Intercept	0.00	0				0.0000	0.0000
L8_PC1	518.53	1	518.5280	41.3480	0.0000	-2.3150	0.3600
L8_PC2	288.39	1	288.3890	22.9960	0.0000	107.7300	22.4650
L8_PC3	1268.05	1	1268.0500	101.1150	0.0000	-183.0750	18.2060
L8_PC4	226.47	1	226.4710	18.0590	0.0000	141.9070	33.3930
L8_PC5	97.69	1	97.6860	7.7890	0.0070	186.6690	66.8830

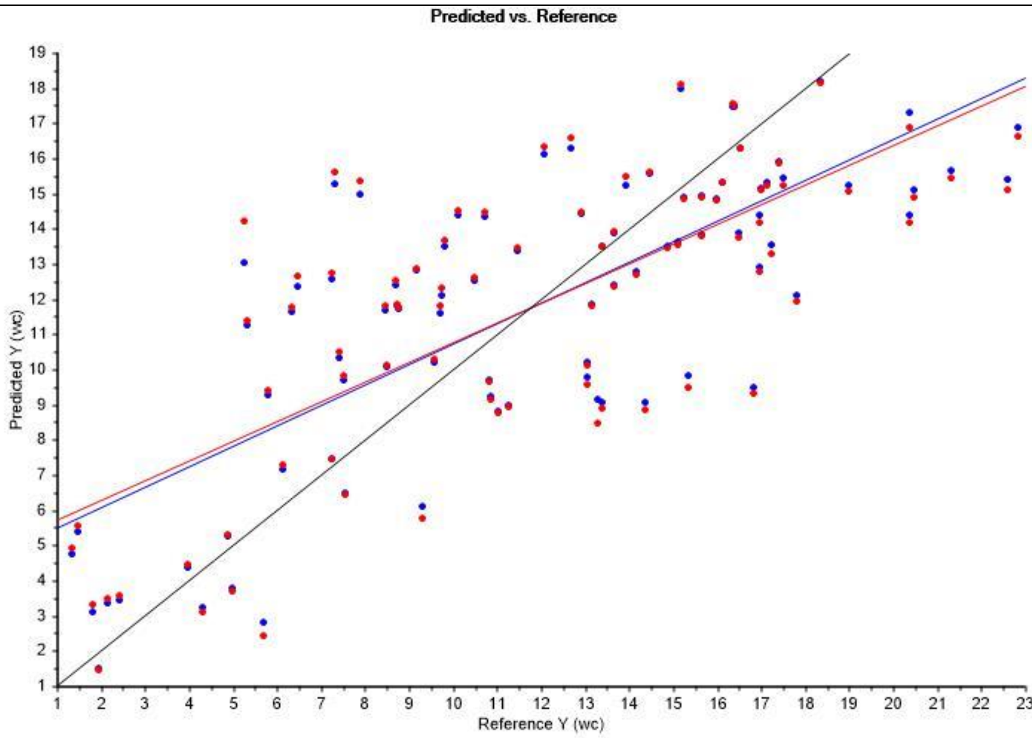

Figure 2-94. Scatter plot of predicted VS observed values - Blue: calibration, Red: validation

Soil Clay Content – GWR Model

Geographically Weighted Regression (GWR) is a regression technique that extends the traditional regression framework by allowing the estimation of local rather than global parameters [57], [58]. In other words, GWR runs a regression for each location, instead of a sole regression for the entire study area. GWR is a useful regression model to work with non-stationary data. The term stationarity refers to relationships in which the influences of the independent variables remain constant over the dependent variable throughout time and space. On the other hand, local or non-stationary models (e.g. GWR) account for different responses in different parts of the study region that the independent variables produce over the dependent variable. In GWR, observations are weighted in accordance with their proximity to point i (determined by the kernel size). This ensures that the weighting of an observation is no longer constant in the calibration, but instead varies with i. As a result, observations closer to i have a stronger influence on the estimation of the parameters for location i. Basically, GWR uses a kernel (also called window or bandwidth) that moves over the study area and seeks to fit the best results for each subarea.

Every single location within the study area has its own set of coefficients; this allows the model to produce an individual r^2 value for each location. It is recommended to map the coefficients and r^2 values to observe how the relationship between the dependent and independent/s variables fluctuates throughout the area under study. This procedure also allows the user to observe how the predictive capabilities of the model vary across space. GWR also provides an overall R^2 output value that can be compared to R^2 values obtained from different regression models, such as Ordinary Least Squares (OLS), and a a t-score output. In this case study, further significant outputs were the regression residuals, and the Beta coefficient raster maps indicating the models' spatial dimensions. The regression coefficients mapping allows for the creation of the global trend surface creation, and the kriging of the residuals allows for mapping the local second order spatial variability.

GWR was performed on Soil Clay data using S2_PC1 and TWI as predictor variables giving a moderate goodness of fit of 53.1 % for the trend surface.

Table 2-28. Results and parameters of the developed GWR model for Soil Clay Content

Soil Clay GWR Model	
Predictors	S2_PC1, TWI
Bandwidth Kernel	57273.47
Residual Squares	17439.98
Effective Number	7.87
Sigma	13.40
AICc	952.09
R^2	0.58
Adjusted R^2	0.53

2.5.7.4. *Development of geospatial maps*

As seen in Section 2.3, the use of geostatistical techniques, enables the creation of continuous raster layers incorporating the statistical properties of the measured data. Because geostatistics is based on statistical assumptions, these techniques produce not only prediction surfaces but also error, or uncertainty surfaces, giving an indication of how good the predictions are.

In geostatistics, Kriging is a method of interpolation for which the interpolated values are modelled by a Gaussian process governed by prior covariances, as opposed to a piecewise-polynomial spline chosen to optimize smoothness of the fitted values.

The basic idea of Kriging is to predict the value of a function at a given point by computing a weighted average of the known values of the function in the neighbourhood of the point. The method is mathematically closely related to regression analysis. Both theories derive a best linear unbiased estimator, based on assumptions on covariances, make use of Gauss-Markov theorem to prove independence of the estimate and error, and make use of very similar formulae. Performing Kriging

Interpolation with polynomial / regression modelled trend surfaces is mathematically identical to generalized least squares polynomial curve fitting.

Kriging is divided into two distinct tasks: quantifying the spatial structure of the data and producing a prediction. Quantifying the structure, known as variography, is a procedure of fitting a spatial-dependence model to the data. To make a prediction for an unknown value for a specific location, Kriging will use the fitted model from variography, the spatial data configuration, and the values of the measured sample points around the prediction location. Arc GIS Geostatistical Analyst has many tools to help you determine which parameters to use and also provides reliable defaults that you can use to make a surface quickly.

The geostatistical interpolation methods used in this study were, Simple Kriging of the Regression residuals. As already mentioned the mean/variance stationarity assumptions were satisfied, after the detrending via regression modelling of soil parameters on EO/ Terrain data layers. Geostatistical Interpolation Models were created for Soil Moisture / Soil Clay at the three sampling territories of Shkodra /Kukes/Pogradec.

Before the production of the final interpolation surface, some idea of how well the model predicts the values at unknown locations must be known. Therefore, Cross-validation helped to make an informed decision as to which model provides the best predictions. The calculated statistics serve as diagnostics that indicate whether the model and/or its associated parameter values are reasonable. The evaluation of the interpolation results of this study was acquired with the Cross-validation method. Consequently, each geostatistical model of Soil variable residuals, created for each sampling territory, was gridded into a raster layer (20 m pixel size for Soil Clay content / 30 m pixel size for Soil Moisture). The addition of the Regression Trend Surface with the Residual Surface gave the Soil Moisture/ Soil Clay map layers.

2.5.7.5. *Maps of soil moisture and clay content*

Soil Moisture

The results of the Variography that took place in order to set up the Kriging parameters (nugget, sill, range) for the Soil Moisture interpolation of the regression residuals are shown in Table 2-29. The determination of the most suitable spatial lag, took under consideration the overall sampling distribution pattern (AvNN) as well as the Moran correlogram. Variogram behaviour was modelled after the exponential model, common in soil chemistry applications. As for the error diagnostics, the geostatistical model was evaluated for performance by the Leave One Out Cross Validation method and gave an RMSE of 2, 97, obviously enhancing, at the actual sampling locations, the global regression model.

Table 2-29. Results of the variography for soil moisture

Soil Moisture Variography	
Lag Size (m)	1150
Model	Exponential
Model Nugget Effect	TRUE
Nugget	0.179
Anisotropy	FALSE
Range (m)	9350
Sill	1.12
Model: 0,179*Nugget+1,12*Exponential (9350)	
Prediction Errors	
Samples	107
Mean	0,00
RMS	2.97
Mean Standardized	0.11
RMS Standardized	1.04
Average Standard Error	2.85

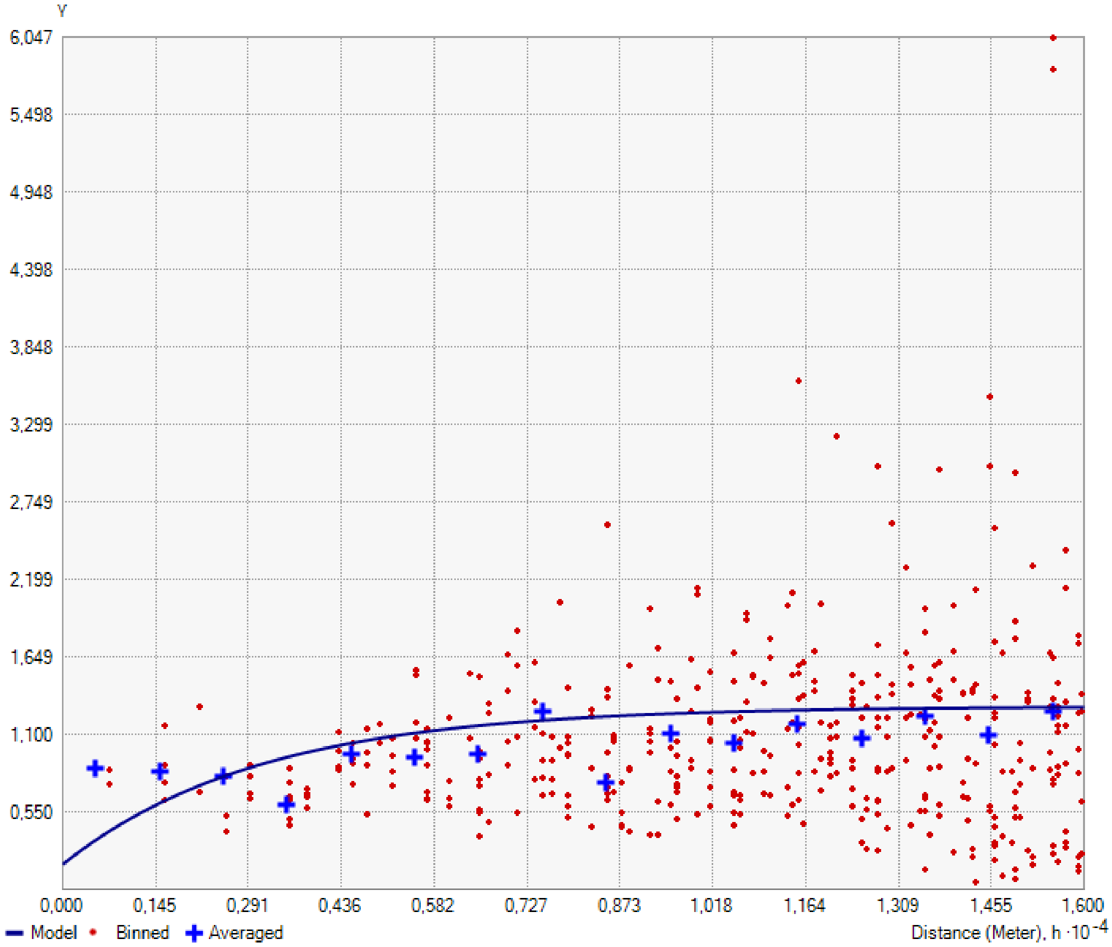

Figure 2-95. Variography graph for soil moisture

On the Variography graph presented in Figure 2-95, one can see the experimental and the modelled Variogram. The interpolation performance can be summarized in the cross-validation error statistics (RMS, RMS Standardized, Av. Standard Error), which can be found in Table 2-29. Finally, the Regression and Interpolation surfaces are added together. The spatial distribution of SM is presented on the Regression Kriging Map clipped on the field river basin borders.

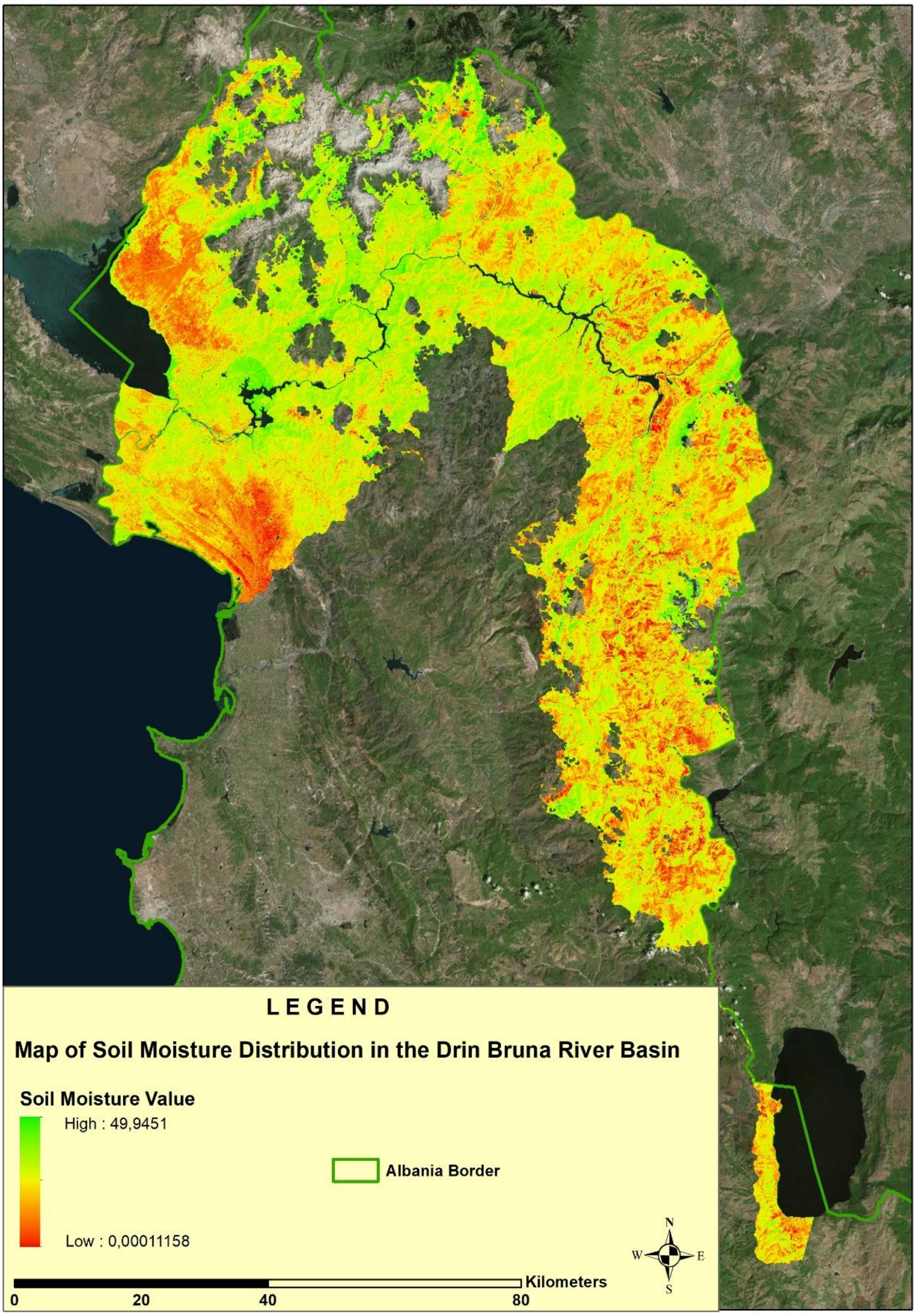

Figure 2-96. Final map of soil moisture in the Drin Bruna River Basin

Soil Clay Content

The results of the Variography that took place in order to set up the Kriging parameters (nugget, sill, range) for the Soil Clay Content interpolation of the regression residuals are shown in the table below. The determination of the most suitable spatial lag, took under consideration the overall sampling distribution pattern (AvNN) as well as the Moran correlogram. Variogram behaviour was modelled after the spherical model. As for the error diagnostics, the geostatistical model was evaluated for performance by the Leave One Out Cross Validation method and gave an RMSE of 8.45.

Table 2-30. Results of the variography for soil clay content

Soil Clay Content Variography	
Lag Size (m)	980985
Model	Spherical
Model Nugget Effect	TRUE
Nugget	0.086
Anisotropy	FALSE
Range (m)	5842
Sill	1
Model: 0,086*Nugget+1,018*Spherical (5842)	
Prediction Errors	
Samples	107
Mean	0.253
RMS	8.450
Mean Standardized	0.110
RMS Standardized	1.140
Average Standard Error	4.730

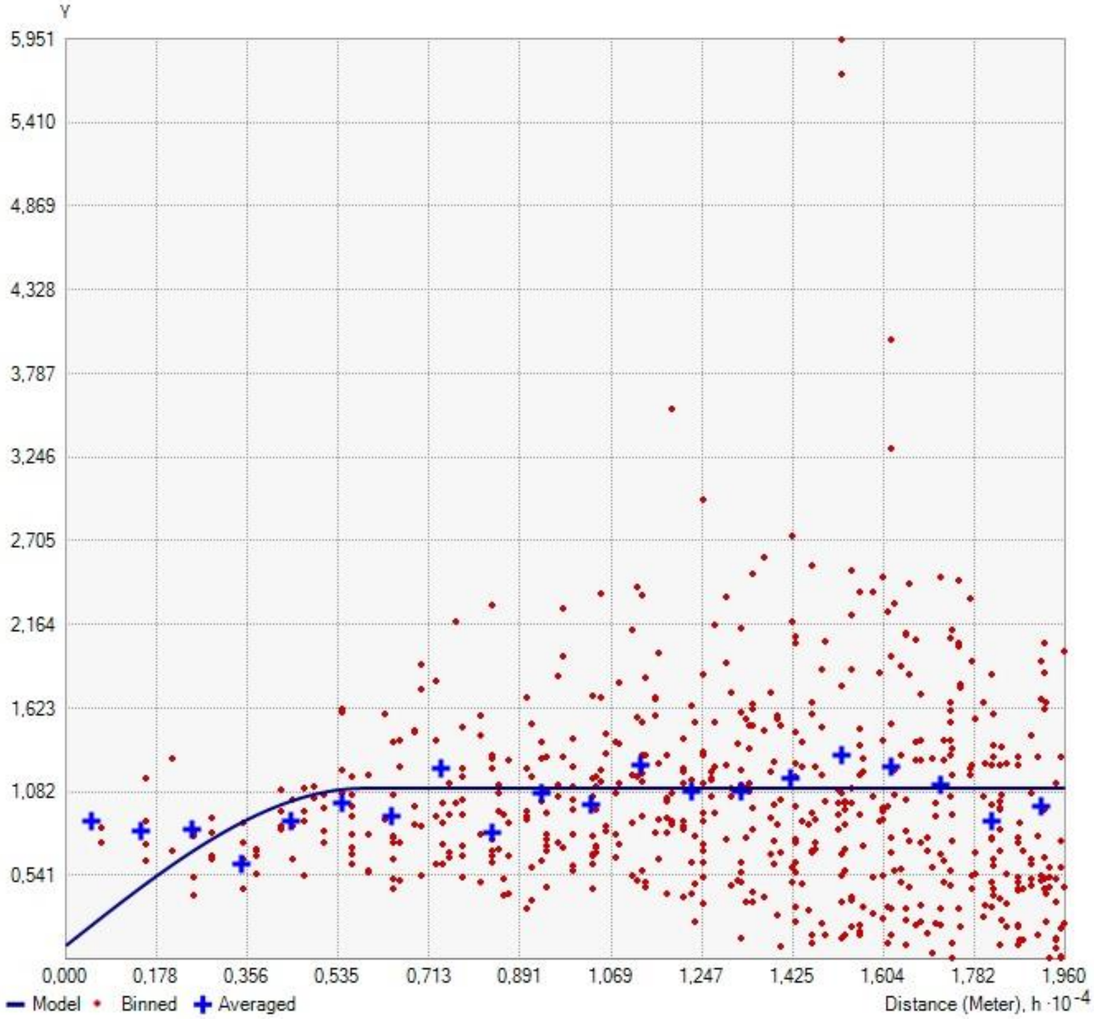

Figure 2-97. Variography graph for soil clay content

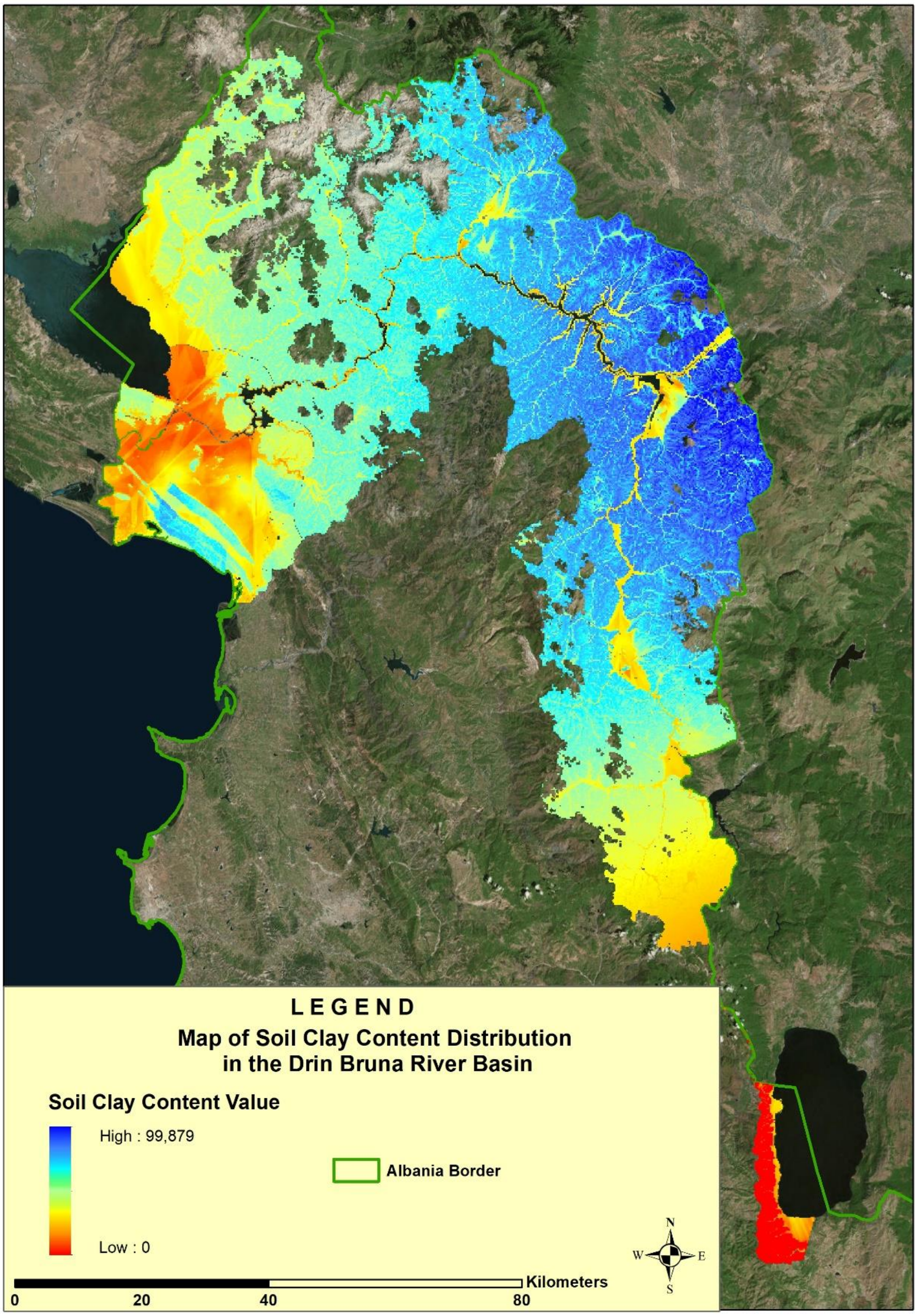

Figure 2-98. Map of soil clay content in the Drin Bruna River Basin

2.5.8. **Integration of soil maps to the myDEWETRA platform**

The Soil Moisture and Clay Content maps obtained by the application of all the steps describes above are available by selecting in the Toolbar of MyDEWETRA the Static section (see Section 2.2.4.3).

When a user selects the Static section, the maps appear available by choosing between three different thematic criteria (for more details see Section 2.2.4.3)

- **Soil moisture**, under which the user can select the map named *Soil Moisture - Drin-Buna*
- **Clay Content**, under which the user can select the map named *Clay content- Drin-Buna*
- **Drin Buna**, under which both two maps are available.

Every time one of this three layers is pulled on, the application uploads it in the ***Control Map***. If the cursor is left on the name of the layer in the ***Layer List*** (top left of the screen) the user enables the tooltip function to open two windows:

- the first one at the top right of the screen (Figure 2-99, panel A2 and B2) which shows the metadata of the layers, such as:
  - the name of the layer
  - the Layer description
  - the reference date
  - the initialization time of the run (if the selected layer is a model's output)
  - the Spatial aggregation (if enabled)
  - the validity interval (if it is a combined variable)
- the second one is placed immediately to the right of the ***Layer List*** (Figure 2-99, panel A1 and B1) and shows the name of the layer, the date of initialization of the run (if the case) and/or the reference date.

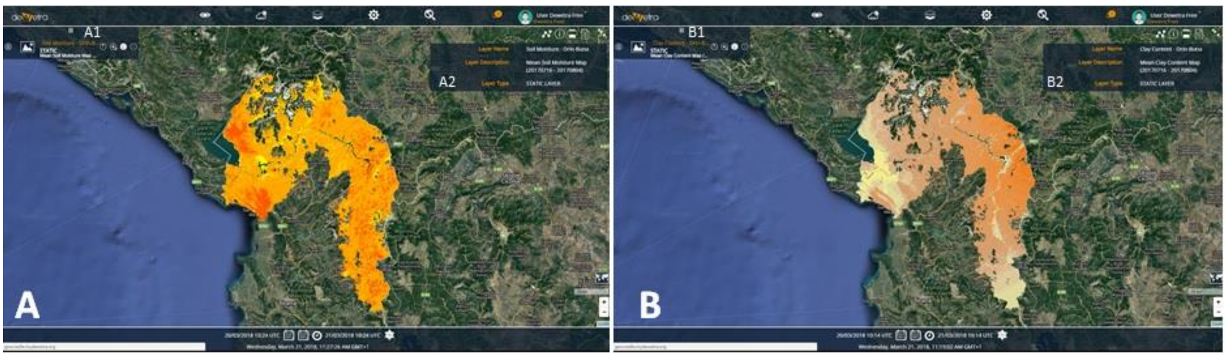

Figure 2-99. Visualization of the Soil Moisture Clay Content maps realized for the Drin-Buna and valid on the period from 15/7/2017 to 4/8/2017.

Clicking on one of the points of the maps, a pop-up screen appears on the left on which is reported the value of that point (Figure 2-100, panel A2- A1).

The values in the Soil Moisture map are expressed in [%] with a range (0-100) as they were measured in situ and consequently in the lab: the color bar used for low values of soil moisture hot colors and for high values of soil cold colors.

The values in the Soil Moisture map are expressed in [%] with a range (0-100) as they were measured in situ and consequently in the lab: the color bar used for low values of soil moisture hot colors and for high values of soil cold colors.

The values in the Soil Clay Content map are expressed in [%] with a range (0-100) as they were measured in the lab.: the color bar used for low values of soil clay light colors and for high values of soil clay orange colors.

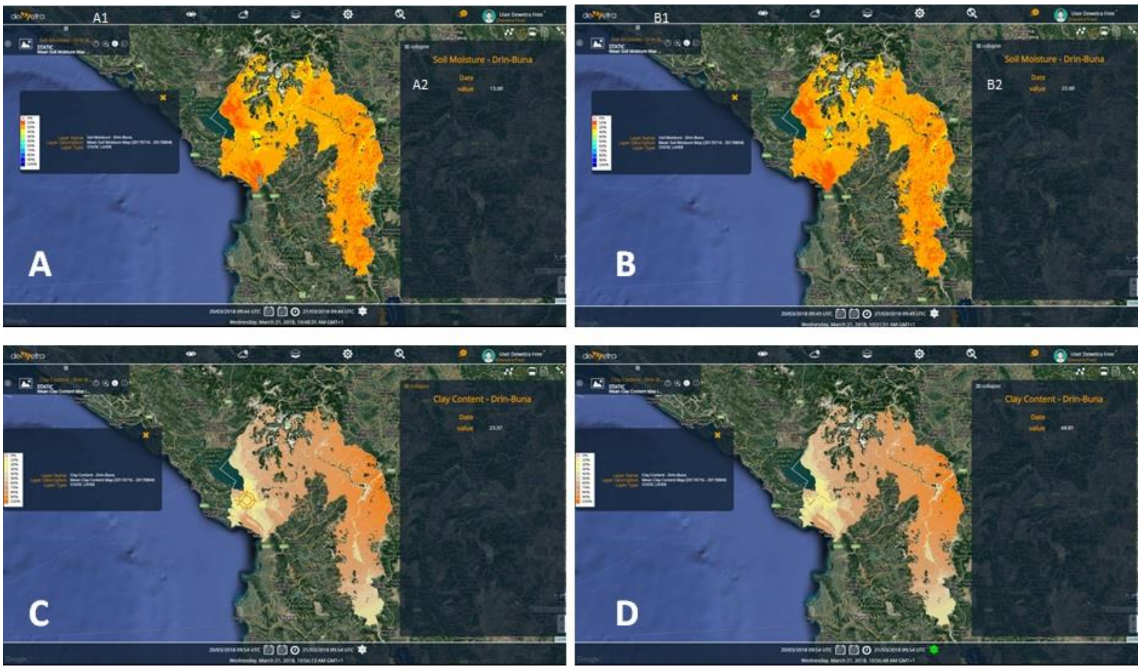

Figure 2-100. Visualization of the Soil Moisture Clay Content maps realized for the Drin-Buna and valid on the period from 15/7/2017 to 4/8/2017.

2.5.9. **Application of EO time-series data to the hydrological model**

Herein, we propose a comparison between modelled soil moisture and satellite soil moisture, and a possible use of the clay content map to evaluate hydrological model’s parameter.

If soil moisture maps are developed in sequential and successive time periods and thus constitute a time-series of EO data, they could be applied to the hydrological model in order to improve its accuracy and robustness.

On the Drin-Buna catchment the Flood-PRObabilistic Operational Forecasting System - Flood-PROOFS - [59] is implemented. It is a system designed to assist decision makers during the operational phases of flood forecasting, nowcasting, mitigation and monitoring. The hydrological core of the system is the Continuum model.

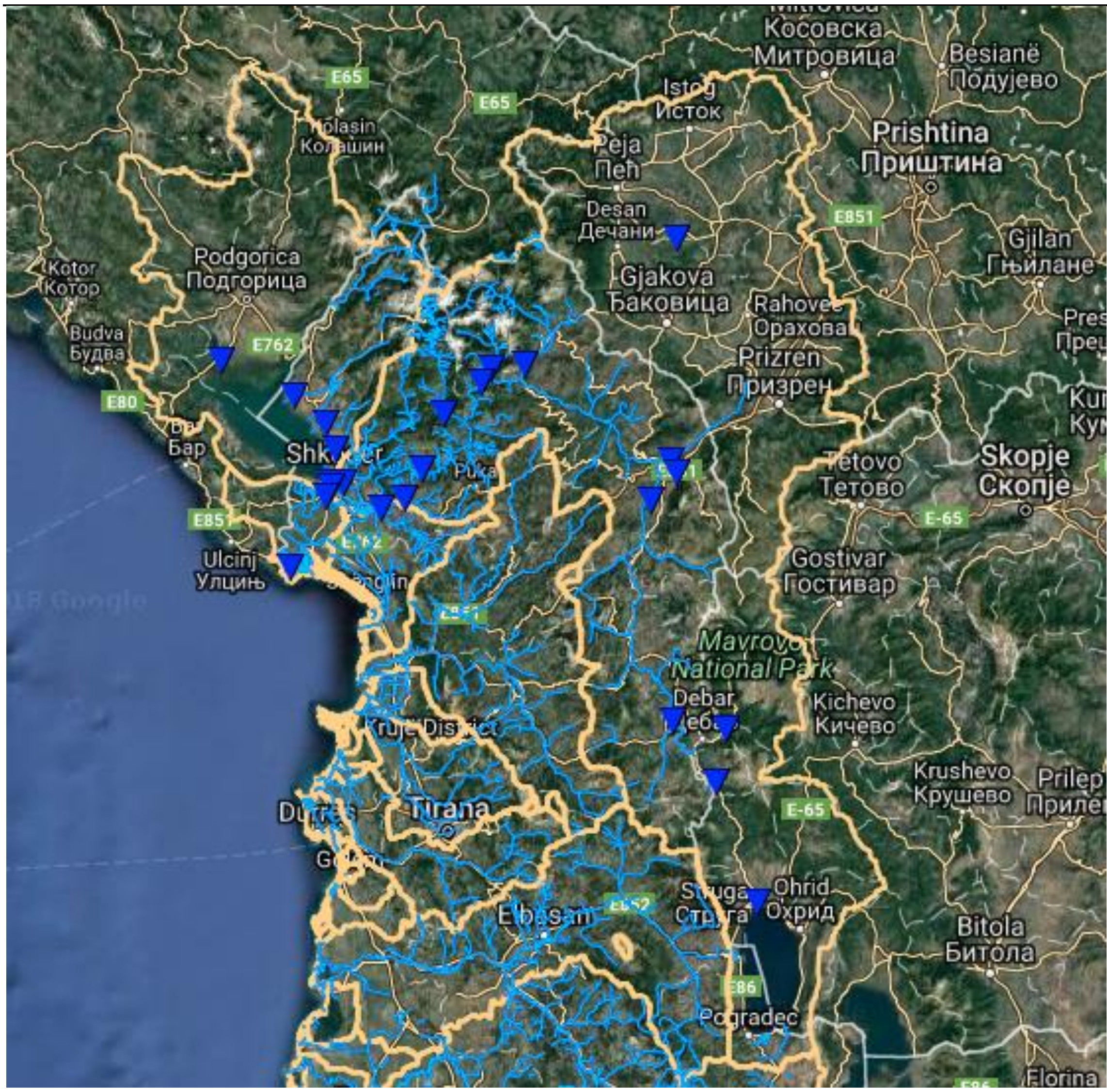

Figure 2-101. River sections on the Drin-Buna river in which discharge outputs of the flood forecasting chain are available.

Continuum is a continuous distributed hydrological model that relies on a morphological approach, based on drainage network components identification [60], [61]. These components are derived from DEMs. The DEM resolution drives the model spatial resolution. Flow in the soil is divided firstly into a sub-surface flow component that is based on a modified Horton schematization (see [62] for details) and that follows the drainage network directions; and secondly, into a deep flow component that moves following the hydraulic head gradient obtained by the water-table modelling. The surface flow schematization distinguishes between channel and hillslope flows. The overland flow (hillslopes) is described by a linear reservoir scheme, while for the channel flow (channel) a schematization derived by the kinematic wave approach is used [63], [64]. The energy balance is solved explicitly at cell scale by means of the force-restore equation, that allows having the LST as a distributed state variable of the model (e.g. [65]–[67]). The snow accumulation-melting module is a simple model that is derived from commonly used equations ([68]) and it is forced by meteorological observations. The mass balance is applied at cell scale for the entire domain of the model, so that a snow cover map can be generated with the same resolution of the DEM. The energy balance and, as a consequence, the evapotranspiration is inhibited for those cells where snow cover is present. The applied approach is very simple and neglects the heat exchanges between the soil and the snow cover, but it is generally sufficient if the goal is the estimation of the snow contribution to the runoff. The precipitation is partitioned into solid or liquid if the air temperature is below or above a fixed threshold [69]. For further details on the model please refer to [23], [69].

Continuum model needs input of: precipitation, air temperature, wind speed, shortwave radiation, air relative humidity. It can consider the presence of reservoirs and hydroelectric infrastructures.

The Continuum model is implemented on the Drin-Buna catchment on a regular with a spatial resolution of 350 meters and 3-hour time step. It gives the temporal evolution of the discharge in different river sections and the spatial and temporal evolution of the distributed soil moisture.

The meteorological inputs of the hydrological model are furnished by the meteorological model COSMO-I7. Unfortunately, no observed meteorological data, from ground stations, are available. Moreover, the Drin-Buna catchment is characterized by the presence of large hydroelectric infrastructures, mainly dams. The dams are modelled within the hydrological model but information about the maneuvers performed on the dams' reservoirs for energy productions are not available. Also observed discharge is not available. These are main limitations of this implementation that prevent an assessment of the performance of hydrological modelling.

For the time period in which the measurements of soil have been realized from 16 July 2017 to 04 August 2017 the hydrological model has been run and discharge and soil moisture computed. The soil moisture derived from satellite analysis can be considered as a mean value on the time period; the mean map from Continuum model has been computed, in Figure 2-102 the comparison between satellite and modelled soil moisture. Modelled soil moisture is drier in comparison to the satellite one as can be see also from the pdf of the soil moisture values (Figure 2-103). The Mean Absolute Error (MAE) between the two maps is 0.174 and the Root Mean Square Difference is 0.048.

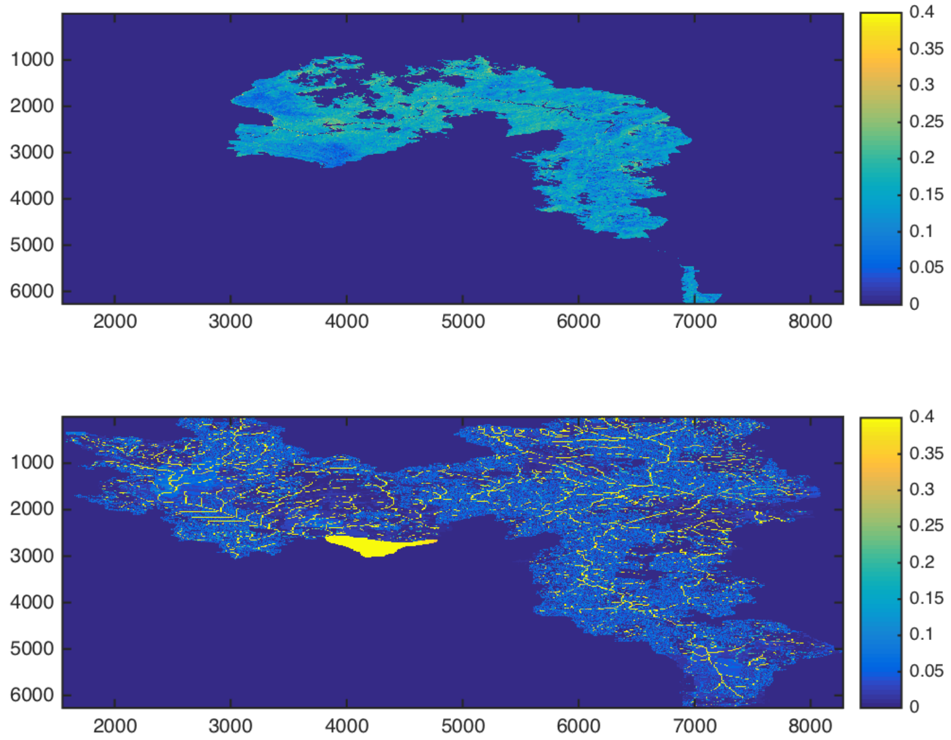

Figure 2-102. Comparison between satellite soil moisture (upper panel) and modelled soil moisture (lower panel) – soil moisture is represented as saturation degree with values in [%].

Unfortunately, during the project implementation it was not possible to assimilate the satellite-derived soil moisture map in the hydrological model because to do so it is necessary to develop a time series of satellite soil moisture maps in order to evaluate the statistics of the observation and implement a bias correction procedure to allow the hydrological model to use the satellite information. More details of the procedure that can be applied to integrate hydrological modelled and satellite soil moisture information can be found in [70]–[72].

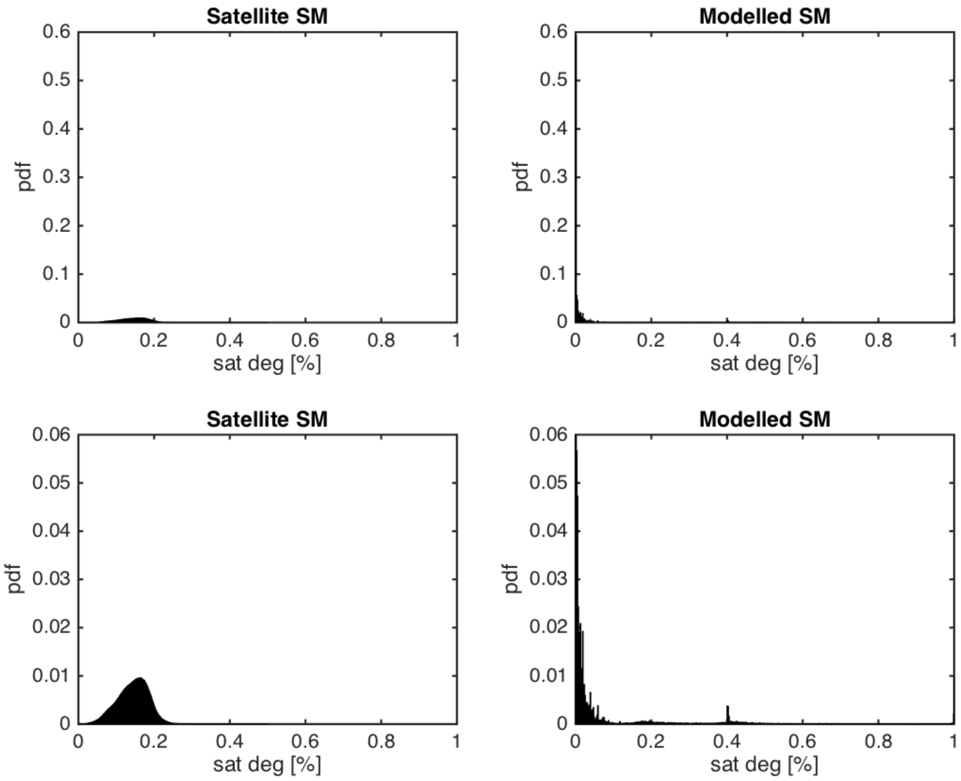

Figure 2-103. Pdf of the satellite and modelled soil moisture (as saturation degree). The lower panels are a zoom of the upper ones.

As described in [62] the Continuum model has one parameter, named ct, that represent the soil field capacity and can be evaluated using soil texture maps. In the actual version of the model the ct is equal to 0.4 (for each point of the basins). Using the clay content map produced within the project it has been evaluated equal to 0.53. considering the mean value of this parameter for each soil type that are in ct_{Sand} = 0.2, ct_{Clay} = 0.7, ct_{Silt} = 0.4.

$$c_t = \frac{\sum_{i=1}^{3} P_i \cdot c_{ti}}{\sum_{i=1}^{3} P_i}$$

in which P_i is the percentage corresponding to the type of soil i (sand, silt, clay).

The hydrological model with the new value of ct has ran in the period of interest, Figure 2-104 shows the comparison between the results with the standard parameter and the value estimate using the clay content map. The differences between the two simulations are about 4%, unfortunately no observed discharge is available to evaluate the improvement of model performance.

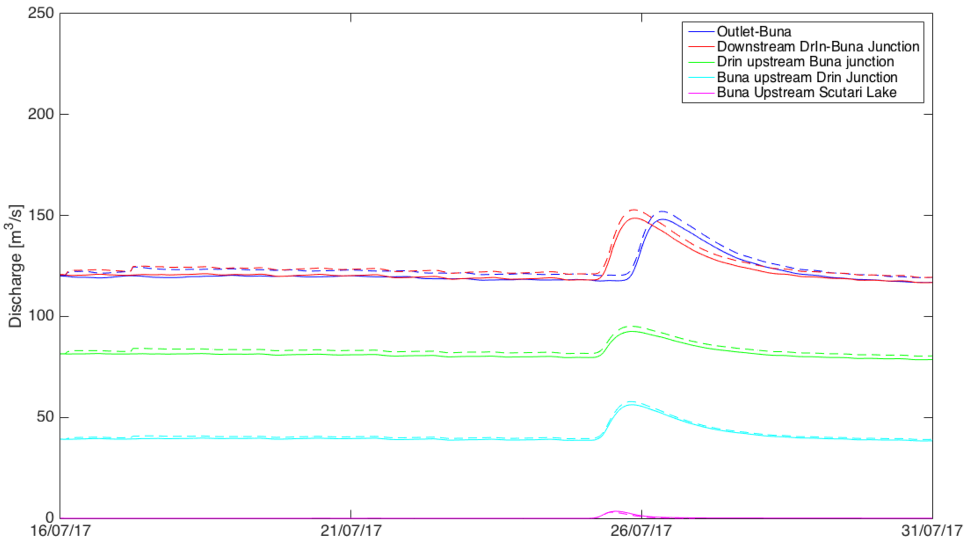

Figure 2-104. Comparison between the modelled discharge in different section of the Drin-Buna river using the standard parametrization of the model (continuous line) and the ri-parametrization using the clay content map (dotted line)

2.6. Training sessions and webinars

One of the most important aspects of the pilot action is the transfer of knowledge and expertise with regards the use of EO data to all the countries and partners in the region. While all partners participated in the collection of the physical soil samples (see Section 2.1.4), the spectral measurements (see Section 2.1.3) were conducted by only two partners, namely i-BEC and TAU, who owned the two spectrometers and were prior to this project sufficiently equipped and trained for the spectral acquisitions and standardization processes. Additionally, a number of partners had no previous expertise or knowledge with the development of spectroscopic multivariate calibration approaches or the application of machine learning algorithms in general (see Section 2.3). To overcome this discrepancy of expertise and to bring the rest of the partners up to date with the current state of the art that was applied during the project implementation in the regional soil spectral library, several knowledge transfer actions were conducted. More concretely, in addition to the meeting and discussions held among all partners involved in this task during the second Project Meeting, and to the material distributed to them, during the course of the task's execution the following actions were done: i-BEC and TAU organized two separate webinars (conducted electronically with all partners being invited), and two training sessions - one in Greece for the partners from the Balkans & Egypt, and one in Israel for the partners from Turkey, with the physical presence of representatives from each partner (Table 2-31). The training sessions involved the training of the partners to the processes of physical sample collection, sample preparation, spectral acquisitions, spectral standardization processes, and development of machine learning algorithms. In the following subsections, the webinars and training sessions conducted are described in detail.

Table 2-31. Overview of the knowledge transfer actions conducted

What	Where	Hosted by	Attended by
Project Meeting 2	Limassol	CUT	All partners
Webinar 1	Web meeting	i-BEC and TAU	CEDARE, CUT, INCA, IPB, SRTI, USCM, UZAY
Webinar 2	Web meeting	i-BEC and TAU	CEDARE, CUT, INCA, IPB, SRTI, USCM, UZAY
Training session 1	Greece	i-BEC	INCA, IPB, SRTI, USCM
Training session 2	Israel	TAU	UZAY

2.6.1. **Webinar #1 – Soil spectroscopy**

The first webinar's main topic was to present a detailed overview of soil spectroscopy. In particular, it covered some introductory information regarding soil science (soil mineralogy, composition, texture, importance in agriculture and food security) and the principles of spectroscopy in general (electromagnetic spectrum, reflectance / absorbance/ emission, absorption bands in elements and molecules), and soil spectroscopy in particular (chromophores, various forms of spectroscopy, point spectroscopy, imaging spectroscopy, sample preparations and spectral acquisitions, importance of standardization and validation, steps to apply standardization). Thus, its main goals were to cover the following areas:

- Key elements of soil science and the importance of soils in agriculture
- Basic principles of spectroscopy
- Application of spectroscopy in soil science

The form of presentation was an oral presentation. The main speaker during this webinar was Prof. Eyal-Ben Dor from TAU who is an expert in soil spectroscopy. Prior to the webinar the presentation as well as some supplementary material was handed out to all participants. The webinar was recorded and uploaded in the shared folder where every partner has access to; thus, every partner can revisit the webinar at any time.

2.6.2. **Webinar #2 – Machine Learning**

This webinar focused on how the processing of the primary source of information that is the spectral signature of each soil is achieved through the means of machine learning (ML) algorithms. The main themes of it were:

- Introduction to ML – Classification and Regression
- Validation of models in ML – avoiding overfitting
- Handling of big data – on the curse of dimensionality
- Introduction to the R programming language
- Developing proper spectroscopic models for the prediction of soil properties using the R language
- Variable importance – identifying significant wavelengths
- Up-scaling of models for EO data

The webinar was conducted as an oral presentation and as a hands-on experience on building the models, since participants were encouraged to follow the speaker and execute the software code presented to them during the webinar on their machines. Prior to the webinar the presentation as well as the necessary software was distributed to all the partners. The webinar was recoded and uploaded in the shared folder, so that every partner might be able to download it and re-watch it whenever they deem fit.

2.6.3. **Training session in Greece**

A training session was held in the premises of i-BEC in Thessaloniki, Greece, on June 1st of 2017. The partners from the Balkans and Egypt (namely CEDARE, INCA, IPB, SRTI and USCM) were invited to Thessaloniki to gain a hands-on experience in soil spectroscopy. The goal was that at the end of this training day the participants would have been able to:

- Operate a spectroradiometer and be able to acquire soil spectra in the laboratory as well as in situ
- Follow a specific measurement protocol to minimize the errors during the measurement
- Be able to standardize the soil spectra
- Understand how to catalogue and archive soil spectral measurements
- Be able to identify potential problems during the measurements and the first steps of spectral processing
- Have all the know-how and tools to apply this to create their country’s soil spectral library

A total of 11 participants travelled to Thessaloniki for the training day and attended the session (Figure 2-105 and Figure 2-106). The training day be ginned with a few short presentations of what was about to unravel and continued with the presentation of the measurement protocol for the acquisition and debugging of laboratory measurements. The participants then had a chance to put the protocol into practice, and after a brief demonstration of the instrument and the process, all participants had a chance to operate the instrument and record soil spectra from the soil samples they contributed to the GEO-CRADLE SSL. A presentation of the standardization procedure then followed, and the participants all applied it on a personal computer. Finally, in-situ measurements were conducted to demonstrate how this process can work on the field in real-life conditions.

Figure 2-105. Professor G. C. Zalidis of i-BEC welcomes the participants to the training day and briefly describes the schedule of the day

The participants were given detailed hard copy and electronic copies of instructions and protocols to follow during the spectral acquisitions. Additionally, the necessary tools needed to automate part of the standardization procedure was given to them. Moreover, contact details of the people in possession of the internal soil samples were given to them (see Subsection 2.1.3 for more details). They were thus equipped with all the tools and means necessary to apply this knowledge to their area.

Figure 2-106. Group photo of the participants to the training day and the involved i-BEC staff

2.6.4. **Training session in Israel**

The Remote Sensing Laboratory hosted a 3 days training session in April of 2017 with the participation of three colleagues from UZAY. The objectives of the training session were similar to the one hosted in Greece:

- Deliver the knowledge of how to perform soil sampling and use it in real-world conditions
- Prepare soils samples for spectral measurements
- Operate a spectroradiometer and be able to acquire soil spectra under various setups
- Follow a specific measurement protocol ensuring the minimizations of errors
- Standardize the soil spectra
- Be able to identify potential problems during the measurements
- Have all the know-how and tools to apply this to create their country's soil spectral library

After the training session ended and participants left back to Turkey, they started performing their own soil survey and collecting samples from various locations in Turkey. Subsequently, they could apply their knowledge and experience on how to perform spectral measurements using a specific protocol.

Few months after the training session, UZAY shared their excellent work of the soil spectral library in order to be included in the GEO-CRADLE database.

Figure 2-107. Members of UZAY in Israel performing soil surveys and spectral measurements

2.7. Brief overview of the outcomes

On the whole, the task succeeded at disseminating the existing state-of-the-art knowledge and essential tools needed for the use of open Earth Observation data for the benefit of food security and water extremes management. Through a series of lectures, meetings, webinars, hands-on demonstrations, training days, dissemination of software, and real-world application of the above the partners tested their new expertise. The culmination of this process was the establishment of the regional SSL which covers a previously underrepresented region (as depicted in Figure 2-52), in addition to a web-GIS platform for the integration of EO data with datasets stemming from other sources. The open, expandable, standardized SSL and the myDEWETRA available platform are available for end-users and researchers alike and can serve as a stepping stone for the successful uptake of EO technologies in the region. The case study considered covered a real-world application of these

3. Analysis of the pilot implementation

3.1. Successes and positive outcomes

3.1.1. **The regional soil spectral library in an under-represented area**

The regional vis-NIR soil spectral library that was developed during the project's implementation is a significant contribution to the LUCAS – European Soil Database and will support future soil mapping activities in the Balkan, Middle East, and North Africa regions.

The development of SSLs is still not adequately taken up or is completely lacking in several Balkan countries according to the LUCAS version that released in 2015. At a first glance the GEO-CRADLE SSL facilitated sampling and a further development and expansion of the LUCAS - European Soil Database in the region of Balkans by around 200%; it further includes samples from Albania, FYROM, and Serbia for the first time (i.e. these countries are not represented in none of other SSLs). In addition, the IFS pilot laid the foundations to create a database with standard soil properties (a similar procedure to the one used for LUCAS), for Northern Africa and Middle East countries for the first time. These efforts aim to kindle a new interest in these regions in the applications of soil spectroscopy and highlight their impact for the region.

Such a regional approach can maximize the innovation potential by exploiting the multiple research and operational assets in the soil spectroscopy domain to effectively develop appropriate national database infrastructures. The results of these activities should not only contribute to climate change mitigation and adaptation actions and but also consider the ongoing work on Sustainable Development Goals implementation (monitoring and reporting obligations). For instance, SSL and advanced machine learning algorithms that are provided by GEO-CRADLE Data Hub could be potentially utilized as key tools to monitor soil organic C stocks for accounting purposes, for monitoring and reporting SDGs (e.g. SDG 15.3) and be important contributors to the adoption of best agronomic and site specific reduced input practices, in a region that faces the first signs of climate change.

Additionally, the GEO-CRADLE SSL is a strong base for the forthcoming hyperspectral remote sensing of soils from space. Whereas modern satellite imagery can be characterised as super-spectral, in the future at least five equipped with hyperspectral imagers are due to be launched: the German Environmental Mapping and Analysis Program (EnMap), the Italian PRecursore IperSpettrale della Missione Applicativa (PRISMA), the U.S. NASA Hyperspectral Infrared Imager (HyspIRI), the Japanese Hyperspectral Imager Suite (HISUI), the Israeli Hyperspectral imager (SHALOM), and the China Commercial Remote-sensing Satellite System (CCRSS). It might then be used for enhanced applications in support of the future Copernicus satellite program in support of Copernicus Land Monitoring program. Last but not least it can be considered as a significant contribution to Copernicus in-situ component in a region with well recognized gaps. Finally, the SSL supports regional contribution to global soil mapping activities and other regional Global Earth Observation System of Systems (GEOSS) hubs.

3.1.2. **The open myDEWETRA platform**

The myDEWETRA platform integrates data and models from heterogenous sources, allowing their easy and effortless visualization and manipulation. The system is a web-GIS platform aimed to multi-risk mapping, forecasting and monitoring. Using the tools of the platform it is possible to aggregate data both in a temporal or spatial way and to build scenarios of risk and damage. Through its web-interface, it allows any computer or device connected to the internet to visualize open geospatial and EO data, as well as link applications to the platform. Its ability to seamlessly integrate models (e.g. for weather, flood, or drought forecasting) enables the continuous monitoring of the RoI, which is so susceptible to water extremes.

The combination and integration of EO data to the platform, and the utilization of a protocol as the one detailed in Section 2.5.9, enables future researchers and end-users alike to utilize this open platform for the modelling of extreme water phenomena.

3.1.3. **Outcome of the feasibility study in the Drin river basin**

The feasibility study highlighted the potential of the pilot activities to offer novel Earth Observation services to a region largely impacted by endangered food security and water extremes. The target area is an important regional basin, on which approximately 1.5 million people rely on for drinking water, agriculture, fisheries, industry, and hydropower. In this context, the utilization of Earth Observation for the conservation and monitoring of soil resources, as well as for the continuous monitoring and risk assessment of water extremes phenomena of utmost importance for the regional economy. The myDEWETRA platform provides a platform for the effortless integration of Earth Observation data, which can assist users with limited knowledge in GIS systems to utilize and explore such datasets. In addition, the visualization of other ancillary data and information is another asset of the tool.

One of the key aspects of the pilot activity was the application of the regional SSL to transform raw Earth Observation data from satellites to end-user services. In particular, maps of two key parameters (i.e. soil moisture and clay content) were developed through a combined synergy of laboratory soil spectral and spectral data recorded from the satellites. This repeatable procedure drives down the time taken and the cost of the contemporary approaches, without sacrificing the predictive accuracy. The developed maps were then integrated into the myDEWETRA platform which is capable of supporting the users in comparing these maps with the hydrograph produced by the hydrological models.

3.2. Identification of weaknesses

3.2.1. **Shortcomings and potential inhibitors**

3.2.1.1. *Insufficient sampling*

Despite the fact that a large number of soil samples were collected to populate the regional SSL, some areas are still under-represented or not sufficiently sampled. For example, Egypt is only represented using 10 soil samples, while Greece, despite being the largest contributor to the SSL, is only represented by one administrative region in the mainland. Accounting for the spatial diversity is also not enough; soil samples should not only cover larger regions, but also sample them adequately to cover the soil type, mineral, and chemical diversity present in the lands. This proper statistical sampling is significant to ensure that the data can be used throughout the whole region.

Moreover, as detailed in previous sections, the reason that some countries are not adequately represented in the GEO-CRADLE lies in the fact that the project partners could not perform the spectral measurements themselves (see also the following point) and were unable to acquire licenses to transfer a large number of soil samples abroad.

3.2.1.2. *Lack of capacities in some countries*

Although an important step was made in building the capacity in some countries, that had no prior knowledge or expertise in soil spectroscopy (but had only in soil science), some partners are markedly lacking in some important tools needed to further populate and extend the SSL. The most important thereof is that out of the around 9 partners, only 3 possess a spectroradiometer in the vis-NIR range, namely i-BEC, TAU, and UZAY. As seen in Section 2.5, to effectively use the SSL with EO data, a pre-requisite is to use a spectroradiometer to measure the newly collected soil samples so that their properties are extracted. Therefore, ownership of such an instrument is a strong asset, as otherwise the scientists will have to resort to sending the soil samples to other laboratories / institutes that possess one, which may delay or hinder the full exploitation of the SSL. As note in the previous point, shipping the soil samples to laboratories in other countries may prove to be a laborious task, since a number of official permits might be needed to be acquired.

3.2.1.3. *Future proofing the SSL*

A potential inhibitor lies in future-proofing the SSL. Although the GEO-CRADLE SSL is open and expandable, it must be ensured that the partners, research institutes, and universities will contribute to it in the future and maintain it. With its open standards, succinct and precise protocols, the GEO-CRADLE SSL should be seen as a starting point for new researchers entering the domain of soil spectroscopy, and not be neglected.

3.2.2. Mitigation strategies

3.2.2.1. *Accounting for the insufficient sampling*

Despite this appearing to be an important inhibitor, it should be underscored that sufficient sampling is a monumental task that is laborious and requires significant funding. While some regions and areas appear to be underrepresented, the knowledge gained during the realization of this task allows

Although some countries are underrepresented in this library, this does not imply that there still exists a potential to be unlocked concerning their more pronounced representation. For example, Egypt and Serbia actually have detailed and large soil libraries comprised of physical soil samples and their respective physical and analytical chemical measurements; thus, it is very possible in the future to expand the SSL with little effort (in comparison to actually having to perform new soil sampling campaigns and the respective analyses).

Moreover, through networking and dissemination activities it is possible to collaborate with more researchers who might be willing to contribute to the regional SSL. Through the establishment of such collaborations it is possible to extend the SSL in other areas.

3.3. Conclusions and future considerations to address the regional challenges

The overarching objective of the IFS-WEM pilot was directly tied with and driven by the GEO's vision stated in the GEO Strategic Plan 2016-2025: Implementing GEOSS, a set of reference EU Strategies (e.g. LULUCF, CAP) and the recently agreed international commitments with regards to sustainable development (e.g. the 2030 Agenda for Sustainable Development). However, to realize its vision and maximize the benefits that EO can bring to users, a **plan for longer term sustainability** (beyond the life of the project) of the tools and services developed through IFS-WEM pilot should be elaborated in close coordination with key players of the relevant ecosystem.

Due to its broad transnational membership (10 countries) and variety of contributing organizations (3 Universities, 8 Research Institutes), IFS-WEM team was able to assemble expertise and perspectives from across different disciplines and communities. In this context, IFS-WEM leveraged this convening power to define **spheres of activity focusing on long term sustainability** of IFS-WEM pilot. The penetration in private markets, further engagement with EO stakeholder communities (e.g. GEO Initiatives, Copernicus), capacity building among current and potential users, the boost of research development and uptake of the provided services and national database of in situ observations serving public government user needs lie at the core of longer term sustainability plan. Accordingly, 5 strategic objectives will guide the "**S5 Objectives for Sustainable IFS-WEM Activities**" beyond the project's life time (Figure 3-1).

Figure 3-1. The S5 objectives for sustainable IFS-WEM activities

The **S5 Objectives**, thoroughly outlined in sections that are mentioned below, builds on achievements from GEOCRADLE's first implementation phase, such as: the development of Regional Data Hub; advances in the areas of national database infrastructure development and data access; gains in regional coordination (e.g. workshops), research and innovation; and the establishment of a diverse and substantial stakeholder-driven network. In doing so, IFS-WEM team will achieve further improvements in the sustainability aspect by proposing a set of core activities which define and focus the scope of actions essential for the attainment of the S5 Objectives.

3.3.1. **Uptake by public domain**

The IFS-WEM pilot delivered to stakeholders, applications, services, data, information that generate social and policy value to the citizens of the RoI. Moreover, the IFS proved that has a broad pool of users from scientists to policy makers, and applications from map validation to modelling. For instance, changes in land and soil management at the farm level could be essential to address challenges emerged by climate change for the future capacity of agriculture.

In this line IBEC, in close collaboration with the JRC European Soil data Centre and other relevant partners (TAU), would take the lead in providing GEOCRADLE soil sampling data and IFS findings to the 'Land Use/Cover Area frame statistical Survey Soil' (LUCAS Soil). The proposed activity surely ensures the sustainability of the IFS findings. Moreover, the IFS-WEM pilot study can be leveraged as a **technical roadmap** with well-defined methodologies and user oriented services to reinforce the interest of decision makers, paving the floor to new EO demands. The national focal points of IFS-WEM pilot could act as the ambassadors of the activities aforementioned in national level.

3.3.2. Private Sector Engagement

As it is already mentioned, IFS-WEM pilot was constructed around a set of reference EU Strategies that allow yielding significant business benefits for the parties involved. In particular, Regional SSL, Mydewetra and networking platform offer unique information and engagement opportunities to the private sector to serve their needs in areas such as agriculture and water extremes management (e.g. insurance sector). They benefited through the access to new types of data, covering new areas, as well as broader community networks.

However, to promote business impacts a socio-economic value analysis (possibly through value tree analysis for IFS- WEM) should be conducted supporting outreach activities and providing impetus to the business development surrounding the uptake of Copernicus, GEOOS and IFS-WEM findings.

3.3.3. Capacity Building

In order to maximize the use and impact of the IFS-WEM solutions, an **EO Capacity Building Program** for the adoption of IFS-WEM components (e.g. Regional SSL, Mydewetra, machine learning models) was developed during the WP4's lifetime, to strengthen participants' knowledge and capabilities (see description of Section 2.6). However, i-BEC's challenge, was not only to train the GEOCRADLE's partners but also **establish a mechanism** that will keep participants' (incl. potentially new additions) knowledge and capabilities sharp and after the project.
The current mechanism addresses the following three pillars:

- **Infrastructure development** by offering a) a concrete methodology (see Appendix A) and a standardization protocol for the development of national database of in situ soil – spectra data able to be implemented in the EUROGEOSS; b) easily deployable models for EO applications.
- **Human capital development** by offering and operating a set of supporting and communicating tools (integrated in the GEOCRADLE Data Hub) that will facilitate the flow of the data, information, knowledge, products and services needed for integrated use across multiple communities.
- **Organizational Development** supported by IBEC/TAU feedback system aiming at a) supporting continuous learning (e.g. modules, inclusion of lessons to 2 Greek University Laboratories via Copernicus Academy) and update the best practices knowledge base (e.g. publications, new models etc.), b) consultancy of relevant stakeholders in adopting the appropriate infrastructure and tools on ways of improving their workflows and services. As a key success story we should highlight that the Laboratory of Remote Sensing and GIS of School of Agriculture (Greece) integrated the IFS results into its research programs and integrate it

into its curriculum and finally was renamed as Laboratory of Remote Sensing, Spectroscopy and GIS.

3.3.4. **Extended Participation**

The wide adoption of IFS-WEM components also includes engagement actions to attract extended participation of relevant communities in different countries in the RoI or key players wishing to join forces (and/or vice versa) towards realizing GEOSS and Copernicus vision.

In this line, IBEC envisages to expand the Regional SSL in other countries, especially those of African territory to support implementation of the EU-Africa R&I Partnership on FNSSA via a diverse open database of agricultural soils information and corresponding EO services. IBEC and NOA have already performed concrete activities in order to expand Regional SSL with sampling points from Tunisia (upcoming ICOSS2018). In the light of the above, synergies with future project funded under H2020 topic SC5-15-2018 and SFS-35 2019B can be explored.

A key tenet of IFS-WEM vision is that EO data should transformed into useable knowledge and information to serve societal needs and sustainable development goals. IBEC as GEO participating organization and contributor to EO4SDGs and GEOGLAM initiatives will convene key stakeholders across the provide-user spectrum (FAO) to co-develop **tailored spatially explicit indicators** for addressing specific issues within the scope of monitoring and reporting SDGs and SBAs. Soil spectroscopy is widely recognized as a cost effective technique for measured SOC content and therefore the Regional SSL may be utilized to report SDG target 15.3.1 as it is stated in the recent released Good Practice Guidance for SDG Indicator 15.3.1.

Last but not least, i-BEC paves the way to **create a community of knowledge exchange** to encourage the interactive sharing of knowledge and best practices between international recognized experts. To that end, building synergies with research international research consortiums (CIRCASA) lies at the core of IBEC's activities in order to explore the mutual benefits of a joint collaboration. Furthermore, IBEC fosters strategic partnerships such as with the Global Partnership for Sustainable Development Data. In particular, the outcomes of IFS pilot are communicating to a wider audience in order to promote best practices and tools to better monitor progress towards sustainable agriculture and natural conservation (see here).

Overall, IFS pilot mobilized actions to mitigate in situ gaps by empowering national focal points to develop their own contributions of EO resources to GEO and Copernicus in situ component.

3.3.5. **Boost Research and Science Development**

From the beginning of the project IBEC, CIMA and NOA recognized that IFS-WEM was not a purely coordination endeavor since it required technical capacities (predicting algorithms, serving platforms) to provide information in support of science-based and data driven decision making. IBEC, TAU and CIMA inspired by the initial ambition of the IFS-WEM pilot and taking into account the global trends in spectroscopy (both space borne and in situ) and recent advances in machine learning technologies envisaged to further promote basic and applied research (e.g. reviewing existing international frameworks) in the areas of the integrated use of remote sensing, and ICT techniques for monitoring the environment (soil and water resources).

The Open Data Commons Open Database License (ODbL) that have been adopted by the regional SSL make it as a valuable source for open, reliable data and information for the academic community. This state can maximise the innovation potential by exploiting the multiple research in soil spectroscopy domain to effectively monitor soil organic C stocks for accounting purposes and be contributor to the adoption of best agronomic and site specific reduced input practices. To that end, based on the initial findings of IFS for the region of our interest there is a need to develop synergies on research at global level in order to further:

- promote guidance on these specialized topics by offering the regional SSL for research purposes as a part of the Global Soil Spectral Library (GSSL). It should be noted that till now there are well reported gaps for the RoI of IFS in the GSSL.
- Utilize IFS tools and findings to improve the understanding of agricultural soil carbon sequestration in these diverse pedoclimatic conditions (synergies with CIRCASA).

Future efforts could be focussed on whether GEOCRADLE findings can be generalized to other soil attributes (bulk density), and micronutrients, and how this approach could be combined with spectra data recorded by proximal soil sensors (relevant work is going to be presented in 21st World Soil Conference). In addition, an analysis of heavy metal content, via national funded projects, in selected sub-samples is proposed to assess the risk of soil contamination in order to serve in the priorities highlighted be well recognized organizations such as FAO (Global Symposium on Soil Pollution).

The core functions to achieve IFS-WEM overall vision drive the implementation of the aforementioned objectives and are summarized in Figure 3-2.

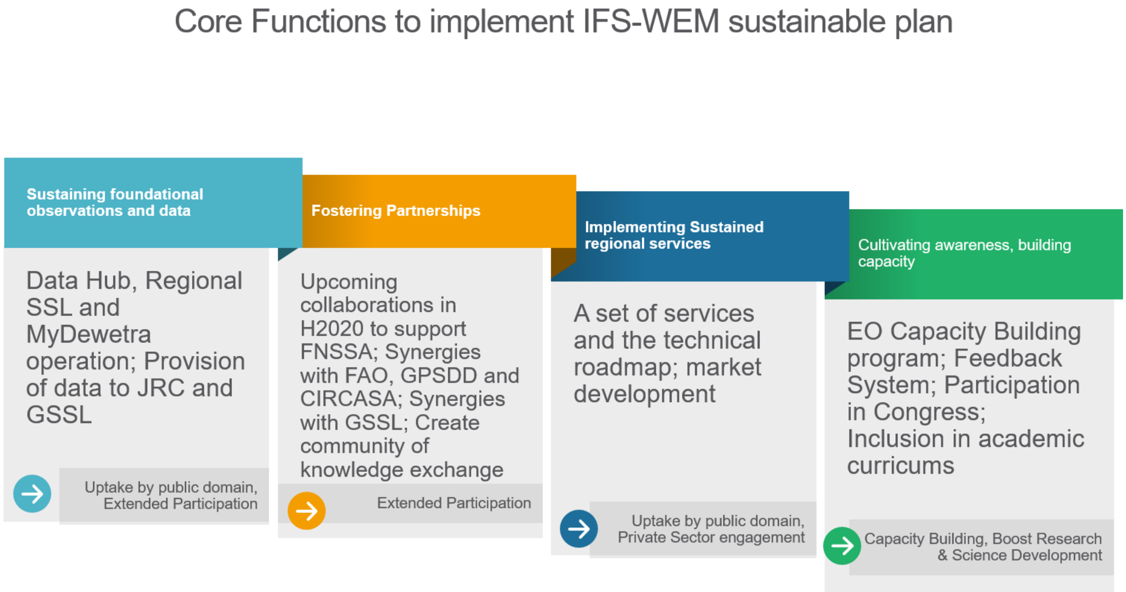

Figure 3-2. Core functions for the implementation of sustainability plan

References

[1] FAO and ITPS, *Status of the World's Soil Resources (SWSR) – Main Report.* 2015.

[2] J. M. Soriano-Disla, L. J. Janik, R. a. Viscarra Rossel, L. M. Macdonald, and M. J. McLaughlin, "The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties," *Appl. Spectrosc. Rev.*, vol. 49, no. 2, pp. 139–186, Feb. 2014.

[3] E. Ben-Dor, S. Chabrillat, J. a. M. Demattê, G. R. Taylor, J. Hill, M. L. Whiting, and S. Sommer, "Using Imaging Spectroscopy to study soil properties," *Remote Sens. Environ.*, vol. 113, no. SUPPL. 1, pp. S38–S55, Sep. 2009.

[4] R. A. V. Rossel and R. Webster, "Predicting soil properties from the Australian soil visible-near infrared spectroscopic database," *Eur. J. Soil Sci.*, vol. 63, no. 6, pp. 848–860, Dec. 2012.

[5] Z. Shi, W. Ji, R. A. Viscarra Rossel, S. Chen, and Y. Zhou, "Prediction of soil organic matter using a spatially constrained local partial least squares regression and the Chinese vis-NIR spectral library," *Eur. J. Soil Sci.*, vol. 66, no. 4, pp. 679–687, Jul. 2015.

[6] G. Tóth, A. Jones, and L. Montanarella, "The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union," *Environ. Monit. Assess.*, vol. 185, no. 9, pp. 7409–7425, Sep. 2013.

[7] R. A. Viscarra Rossel, T. Behrens, E. Ben-Dor, D. J. Brown, J. A. M. Demattê, K. D. Shepherd, Z. Shi, B. Stenberg, A. Stevens, V. Adamchuk, H. Aïchi, B. G. Barthès, H. M. Bartholomeus, A. D. Bayer, M. Bernoux, K. Böttcher, L. Brodský, C. W. Du, A. Chappell, Y. Fouad, V. Genot, C. Gomez, S. Grunwald, A. Gubler, C. Guerrero, C. B. Hedley, M. Knadel, H. J. M. Morrás, M. Nocita, L. Ramirez-Lopez, P. Roudier, E. M. R. Campos, P. Sanborn, V. M. Sellitto, K. A. Sudduth, B. G. Rawlins, C. Walter, L. A. Winowiecki, S. Y. Hong, and W. Ji, "A global spectral library to characterize the world's soil," *Earth-Science Rev.*, vol. 155, no. February, pp. 198–230, Apr. 2016.

[8] B. Kuang and A. M. Mouazen, "Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms," *Eur. J. Soil Sci.*, vol. 62, no. 4, pp. 629–636, 2011.

[9] E. Ben Dor, C. Ong, and I. C. Lau, "Reflectance measurements of soils in the laboratory: Standards and protocols," *Geoderma*, vol. 245–246, no. JANUARY, pp. 112–124, May 2015.

[10] V. Kopačková and E. Ben-Dor, "Normalizing reflectance from different spectrometers and protocols with an internal soil standard," *Int. J. Remote Sens.*,

vol. 37, no. 6, pp. 1276–1290, 2016.

[11] E. Ben-Dor, “Quantitative remote sensing of soil properties,” *Adv. Agron.*, vol. 75, no. July, pp. 173–243, 2002.

[12] T. Paz-Kagan, E. Zaady, C. Salbach, A. Schmidt, A. Lausch, S. Zacharias, G. Notesco, E. Ben-Dor, and A. Karnieli, “Mapping the spectral soil quality index (SSQI) using airborne imaging spectroscopy,” *Remote Sens.*, vol. 7, no. 11, pp. 15748–15781, 2015.

[13] A. Gholizadeh, N. Carmon, A. Klement, E. Ben-Dor, and L. Borůvka, “Agricultural soil spectral response and properties assessment: Effects of measurement protocol and data mining technique,” *Remote Sens.*, vol. 9, no. 10, 2017.

[14] N. L. Tsakiridis, N. Tziolas, A. Dimitrakos, G. Galanis, E. Ntonou, A. Tsirika, E. Terzopoulou, E. Kalopesa, and G. C. Zalidis, “Predicting soil properties for sustainable agriculture using vis-NIR spectroscopy: a case study in northern Greece,” in *Fifth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2017)*, 2017, vol. 10444, p. 28.

[15] N. L. Tsakiridis, J. B. Theocharis, and G. C. Zalidis, “A fuzzy rule-based system utilizing differential evolution with an application in vis-NIR soil spectroscopy,” in *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2017, pp. 1–7.

[16] T. Angelopoulou, A. Dimitrakos, E. Terzopoulou, G. Zalidis, J. Theocharis, T. Stafilov, and A. Zouboulis, “Reflectance Spectroscopy (Vis-NIR) for Assessing Soil Heavy Metals Concentrations Determined by two Different Analytical Protocols, Based on ISO 11466 and ISO 14869-1,” *Water, Air, Soil Pollut.*, vol. 228, no. 11, p. 436, Nov. 2017.

[17] K. U. Totsche, T. Rennert, M. H. Gerzabek, I. Kögel-Knabner, K. Smalla, M. Spiteller, and H. J. Vogel, “Biogeochemical interfaces in soil: The interdisciplinary challenge for soil science,” *Journal of Plant Nutrition and Soil Science*, vol. 173, no. 1. pp. 88–99, 2010.

[18] D. W. Pribyl, “A critical review of the conventional SOC to SOM conversion factor,” *Geoderma*, vol. 156, no. 3–4. pp. 75–83, 2010.

[19] J. Ashworth, “Improvements to two routine methods for calcium carbonate determination in soils,” *Commun. Soil Sci. Plant Anal.*, vol. 28, no. 11–12, pp. 841–848, Jun. 1997.

[20] IUSS Working Group WRB, *World reference base for soil resources 2014. International soil classification system for naming soils and creating legends for soil maps*. 2014.

[21] N. Rebora, L. Ferraris, J. von Hardenberg, and A. Provenzale, “RainFARM: Rainfall

Downscaling by a Filtered Autoregressive Model," *J. Hydrometeorol.*, vol. 7, no. 4, pp. 724–738, 2006.

[22] N. Rebora, L. Ferraris, J. von Hardenberg, and A. Provenzale, "Rainfall downscaling and flood forecasting: a case study in the Mediterranean area," *Nat. Hazards Earth Syst. Sci.*, vol. 6, no. 4, pp. 611–619, Jul. 2006.

[23] F. Silvestro, S. Gabellani, F. Delogu, R. Rudari, and G. Boni, "Exploiting remote sensing land surface temperature in distributed hydrological modelling: The example of the Continuum model," *Hydrol. Earth Syst. Sci.*, vol. 17, no. 1, pp. 39–62, 2013.

[24] R. A. V. Rossel and T. Behrens, "Using data mining to model and interpret soil diffuse reflectance spectra," *Geoderma*, vol. 158, no. 1–2, pp. 46–54, Aug. 2010.

[25] H. W. S Wold, H Martens, "The multivariate calibration problem in chemistry solved by the PLS method," *Matrix pencils*, no. 1981, pp. 286–293, 1983.

[26] H. Wold, "Partial least squares," *Encycl. Stat. Sci.*, no. 2, pp. 581–591, 1985.

[27] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Anal. Chim. Acta*, vol. 185, no. C, pp. 1–17, 1986.

[28] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," in *Chemometrics and Intelligent Laboratory Systems*, 2001, vol. 58, no. 2, pp. 109–130.

[29] W. F. Massy, "Principal Components Regression in Exploratory Statistical Research," *J. Am. Stat. Assoc.*, vol. 60, no. 309, pp. 234–256, 1965.

[30] J. R. Quinlan, "Learning with continuous classes," *Mach. Learn.*, vol. 92, pp. 343–348, 1992.

[31] J. R. Quinlan, "Combining Instance-Based and Model-Based Learning," *Mach. Learn.*, vol. 76, pp. 236–243, 1993.

[32] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY: Springer New York, 2013.

[33] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 1992, pp. 144–152.

[34] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Adv. Neural Inf. Process. Syst.*, vol. 9, no. x, pp. 155–161, 1997.

[35] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005.

[36] R. W. Kennard and L. A. Stone, “Computer Aided Design of Experiments,” *Technometrics*, vol. 11, no. 1, pp. 137–148, 1969.

[37] V. Bellon-Maurel, E. Fernandez-Ahumada, B. Palagos, J.-M. Roger, and A. McBratney, “Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy,” *TrAC Trends Anal. Chem.*, vol. 29, no. 9, pp. 1073–1081, Oct. 2010.

[38] B. Stenberg, R. A. Viscarra Rossel, A. M. Mouazen, and J. Wetterlind, “Visible and Near Infrared Spectroscopy in Soil Science,” in *Advances in Agronomy*, 1st ed., vol. 107, no. 10, Elsevier Inc, 2010, pp. 163–215.

[39] L. Ramirez-Lopez, T. Behrens, K. Schmidt, A. Stevens, J. A. M. Demattê, and T. Scholten, “The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets,” *Geoderma*, vol. 195–196, pp. 268–279, Mar. 2013.

[40] M. Nocita, A. Stevens, G. Toth, P. Panagos, B. van Wesemael, and L. Montanarella, “Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach,” *Soil Biol. Biochem.*, vol. 68, pp. 337–347, 2014.

[41] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[42] L. Breiman, “Stacked regressions,” *Mach. Learn.*, vol. 24, no. 1, pp. 49–64, Jul. 1996.

[43] J. H. Holland, *Adaptation in Natural and Artificial Systems*. 1975.

[44] A. . McBratney, M. . Mendonça Santos, and B. Minasny, “On digital soil mapping,” *Geoderma*, vol. 117, no. 1–2, pp. 3–52, 2003.

[45] T. Hengl, *A Practical guide to Geostatistical Mapping*. 2009.

[46] M. Sadeghi, S. B. Jones, and W. D. Philpot, “A linear physically-based model for remote sensing of soil moisture using short wave infrared bands,” *Remote Sens. Environ.*, vol. 164, pp. 66–76, Jul. 2015.

[47] D. Zhang and G. Zhou, “Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review,” *Sensors*, vol. 16, no. 8, p. 1308, Aug. 2016.

[48] J. A. Sobrino, J. C. Jimenez-Muoz, G. Soria, M. Romaguera, L. Guanter, J. Moreno, A. Plaza, and P. Martinez, “Land Surface Emissivity Retrieval From Different VNIR

and TIR Sensors," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 2, pp. 316–327, Feb. 2008.

[49] Q. Weng, D. Lu, and J. Schubring, "Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies," *Remote Sens. Environ.*, vol. 89, no. 4, pp. 467–483, 2004.

[50] J. Mallick, C. K. Singh, S. Shashtri, A. Rahman, and S. Mukherjee, "Land surface emissivity retrieval based on moisture index from LANDSAT TM satellite data over heterogeneous surfaces of Delhi city," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 19, no. 1, pp. 348–358, 2012.

[51] E. P. Crist and R. J. Kauth, "The Tasseled Cap de-mystified," *Photogramm. Eng. Remote Sens.*, vol. 52, no. 1, pp. 81–86, 1986.

[52] C. Dirksen, *Soil physics measurements*. 1999.

[53] T. Hengl, P. Panagos, A. Jones, and G. Tóth, *Status and prospect of soil information in south - eastern Europe: soil databases, projects and applications*. 2007.

[54] L. W. Zevenbergen and C. R. Thorne, "Quantitative analysis of land surface topography," *Earth Surf. Process. Landforms*, vol. 12, no. 1, pp. 47–56, 1987.

[55] M. Shabou, B. Mougenot, Z. Chabaane, C. Walter, G. Boulet, N. Aissa, and M. Zribi, "Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi-Arid Lands," *Remote Sens.*, vol. 7, no. 5, pp. 6059–6078, May 2015.

[56] C. Gomez, P. Lagacherie, and G. Coulouma, "Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis–NIR data," *Geoderma*, vol. 189–190, pp. 176–185, Nov. 2012.

[57] S. Kumar, R. Lal, and D. Liu, "A geographically weighted regression kriging approach for mapping soil organic carbon stock," *Geoderma*, vol. 189–190, pp. 627–634, Nov. 2012.

[58] X. Tan, P.-T. Guo, W. Wu, M.-F. Li, and H.-B. Liu, "Prediction of soil properties by using geographically weighted regression at a regional scale," *Soil Res.*, vol. 55, no. 4, p. 318, 2017.

[59] P. Laiolo, S. Gabellani, N. Rebora, R. Rudari, L. Ferraris, S. Ratto, H. Stevenin, and M. Cauduro, "Validation of the Flood-PROOFS probabilistic forecasting system," *Hydrol. Process.*, vol. 28, no. 9, pp. 3466–3481, Apr. 2014.

[60] F. Giannoni, G. Roth, and R. Rudari, "Can the behaviour of different basins be described by the same model's parameter set? A geomorphologic framework," *Phys. Chem. Earth*, vol. 28, no. 6–7, pp. 289–295, 2003.

[61] F. Giannoni, G. Roth, and R. Rudari, "A procedure for drainage network identification from geomorphology and its application to the prediction of the hydrologic response," *Adv. Water Resour.*, vol. 28, no. 6, pp. 567–581, 2005.

[62] S. Gabellani, F. Silvestro, R. Rudari, and G. Boni, "General calibration methodology for a combined Horton-SCS infiltration scheme in flash flood modeling," *Nat. Hazards Earth Syst. Sci.*, vol. 8, no. 6, pp. 1317–1327, Dec. 2008.

[63] R. A. Wooding, "A hydraulic model for the catchment-stream problem," *J. Hydrol.*, vol. 3, no. 3–4, pp. 254–267, Nov. 1965.

[64] V. P. Singh and D. A. Woolhiser, "Mathematical Modeling of Watershed Hydrology," *J. Hydrol. Eng.*, vol. 7, no. 4, pp. 270–292, Jul. 2002.

[65] J. D. Lin, "On the force-restore method for prediction of ground surface temperature," *J. Geophys. Res.*, vol. 85, pp. 3251–3254, 1980.

[66] R. E. Dickinson, "The Force–Restore Model for Surface Temperatures and Its Generalizations," *J. Clim.*, vol. 1, no. 11, pp. 1086–1097, Nov. 1988.

[67] F. Sini, G. Boni, F. Caparrini, and D. Entekhabi, "Estimation of large-scale evaporation fields based on assimilation of remotely sensed land temperature," *Water Resour. Res.*, vol. 44, no. 6, Jun. 2008.

[68] D. Maidment, *Handbook of Hydrology*, 1st ed. McGraw-Hill Education, 1993.

[69] F. Silvestro, S. Gabellani, R. Rudari, F. Delogu, P. Laiolo, and G. Boni, "Uncertainty reduction and parameter estimation of a distributed hydrological model with ground and remote-sensing data," *Hydrol. Earth Syst. Sci.*, vol. 19, no. 4, pp. 1727–1751, Apr. 2015.

[70] P. Laiolo, S. Gabellani, L. Campo, F. Silvestro, F. Delogu, R. Rudari, L. Pulvirenti, G. Boni, F. Fascetti, N. Pierdicca, R. Crapolicchio, S. Hasenauer, and S. Puca, "Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 48, pp. 131–145, Jun. 2016.

[71] L. Cenci, P. Laiolo, S. Gabellani, L. Campo, F. Silvestro, F. Delogu, G. Boni, and R. Rudari, "Assimilation of H-SAF Soil Moisture Products for Flash Flood Early Warning Systems. Case Study: Mediterranean Catchments," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 12, pp. 5634–5646, Dec. 2016.

[72] L. Cenci, L. Pulvirenti, G. Boni, M. Chini, P. Matgen, S. Gabellani, G. Squicciarino, and N. Pierdicca, "An evaluation of the potential of Sentinel 1 for improving flash flood predictions via soil moisture–data assimilation," *Adv. Geosci.*, vol. 44, pp. 89–100, Nov. 2017.

Appendix A

A protocol and instructions on how to establish a SSL by Ogen Y. and Ben-Dor E. (Tel Aviv University, The Remote Sensing Laboratory)

Preface

Soil spectral library (SSL) becomes a very important issue in all domains (local, regional and global) (Viscarra et al. 2016). The SSL is used to generate predictors for many soil attributes using a chemometric approach and has a vast potential in precision agriculture activities, ranging from ground to space domains. As growing activity in this direction has been taking place worldwide, it is important to establish a standard protocol on how to build such a SSL. This is in order to allow in the future better collaboration with other colleagues as well as to enable smooth and correct merging of SSLs from various sources.

Introduction

The following document provides a protocol for building a Soil Spectral Library (SSL). The protocol covers how to plan and collect the samples, the sampling procedures followed by the spectral measurements and the analyses of chemical-physical attributes. A special attention is given to the physical construction of the library and how to store soil samples for future utilization.

1. Project planning and samples collection

Building SSL requires a prior planning in order to sample different soil types from diverse locations and also collect the samples from various land uses and horizons. There are two parallel ways to acquire the samples: self-field sampling and samples collection from existing (pedologic) research laboratory (take-away method). If a self-field sampling is required, start at 1a. **If you are familiar with laboratories or organizations which hold soil samples, start at 1b.**

1a. Self-field sampling and procedures

1. Use a local or national soils map (1:100,000 or 1:250,000 scales) and identify the distribution of the soil types (maintain a representative soil types sampling within the study area). It is highly recommended to work with a professional soil surveyor and include a profile sampling.
2. Make a list of all the soils starting from the most common ones to the less common.
3. Mark the geographic locations for sampling on a map. if a certain soil type has a high spatial distribution, its samples should be collected from diverse locations as possible.

4. Organize sampling days to collect the samples together with the following items:
    - Soil map with the marked location
    - Global Positioning System (GPS)
    - Writing paper (for suggested format, see Table 1) and a pen
    - Shovel
    - Sampling spoon
    - Paper bags and a marker
    - Camera
5. Collect at least a 150 g of soil sample for each horizon (surface or profile) and keep it inside a paper bag.
6. For surface sampling: before sampling take a photo, sample the soil without debris. Describe the sample status and surrounding (moisture content, stones).
7. Write the sample number on the paper bag.
8. Fill in the table with the sample information.
9. Continue to collect the samples and edit the table.
10. Continue to section 2.

Additional information can be found at:

1. Smith, D.B., Cannon, W.F., Woodruff, L.G., Solano, Federico, Kilburn, J.E., and Fey, D.L., 2013, **Geochemical and mineralogical data for soils of the conterminous United States**: U.S. Geological Survey Data Series 801, 19 p., https://pubs.usgs.gov/ds/801/downloads/Appendix%201_NASGLP-Soil-sampling-manual.pdf
2. EuroGeoSurveys Geochemistry Working Group, 2008. ***EuroGeoSurveys Geochemical mapping of agricultural and grazing land soil of Europe (GEMAS) - Field manual***. NGU Report 2008.038. Geological Survey of Norway, 46 pp. http://www.ngu.no/upload/Publikasjoner/Rapporter/2008/2008_038.pdf

Table 1

#	Latitude	Longitude	Elevation (m)	Depth (cm)	Land use	# of photo	Remarks
1							
2							
3							
4							

1b. The take-away method

1. Do a desk work and make a list of soil surveyors and any laboratories who perform chemical-physical measurements on soils.
2. Contact the laboratory, explain about the project and ask if they have any unneeded soil samples which have chemical-physical attributes together with coordinates data and additional meta data information. If so, ask permission to collect the samples.
3. Collect the soil samples together with the necessary data (for details, see section 4).
4. Continue to section 2.

2. Initial processing

Before conducting any measurement, it is essential to perform the following steps:

1. Dry the soil samples to room condition (provide the temp. and RH (%) in the room).
2. Ground each sample to 2mm.
3. Sieve (>2 mm) the soil sample and make sure you are left with at least 100 g of sample.
4. Continue to section 3.

3. Coding and maintenance

Coding the samples is essential in order to keep track of the samples and find them easily in the database. The coding should be as follows:

AA-BB-CCC-DDDDD

AA – Country code

BB – Serial number of the soil

CC – Soil type (WRB)

DDDDD – Sampling depth (3 digits - in cm)

For example, the code **IL-VR-030-00001** represent a Vertisol sample taken from a 30cm depth in Israel.

Consider establish a BARCODE system for the soil samples refereeing soil sample to the data basis.

4. Spectral measurements

Spectral measurements should be performed using any available reflectance spectrometer for the 400-2500nm. Provide the spectral resolution and sampling information. Use any protocol that assured by your laboratory (recommended in CSIRO protocol, Ben Dor et al. 2015). Make sure that your protocol will be composed of representative replications with SD less than 4% between each spectrum in a given sample. Convert sample reading to reflectance (either by white reference or any other method). All spectral measurements MUST be accompanied with using the internal soil standard as discussed in (Ben Dor et al. 2015).

5. Chemical-physical attributes

Chemical Physical attributes are essential part of the SSL, for that purpose each sample has to be accompanied with chemical-physical data after using "wet chemistry" using traditional and accepted methods. Please provide a reference to any method used. There are three obligatory attributes **to any SSL**. Additional attributes if available are welcome. The obligatory attributes are: Soil texture (sand, silt and clay distributions), Organic Matter (OM) and calcium carbonate ($CaCO_3$). Highly recommended attributes are: Specific Surface Area (SSA), soil moisture, Cation Exchange Capacity (CEC), Free Iron Oxides (Fed), pH and Electrical Conductivity (EC).

It is important to include in the physical-chemical data base the declared laboratory accuracy for every attribute.

6. Additional data

In addition to the soil attributes, it is important to include in the database other information which is needed in order to build a broad picture of the samples. This information includes:

- Sampling date (DD-MM-YYYY format)
- Data Source
- Latitude and Longitude (decimal degree format)
- Elevation above sea level (in meters)
- Sampling depth (in cm)
- Land use
- Climate (according Koeppen classification)
- Soil type (according the WRB and USDA)

- Photo of the sample
- Spectrometer used and spectral its resolution

An example of the database is given in Table 2.

Table 2

Serial Number	**1**	**2**
Sample Code	**IL-VR-030-XXXXX**	**IL-AT-005-XXXXX**
Date of sampling	25.11.2015	29.07.2016
Photo number	1	2
Latitude	32.595561	33.879545
Longitude	34.542354	36.544444
Elevation (m)	40	350
Depth (cm)	30	5
Soil type (WRB)	Vertisol	Anthrosol
Soil type (USDA)	Vertisol	Anthrepts
Climate (Koeppen)	BWh	BWh
Spectrometer	ASD field spec Pro	ASD field spec Pro
Spectral resolution (nm)	1	1
OM (%)	2.8	3.2
CaCO3 (%)	3.4	2.9
Clay Fraction (%)	30	20
Silt Fraction (%)	25	30
Sand Fraction (%)	45	50

pH	7.5	6.8
EC (μ/S)	2.8	8.6
Soil moisture (%)	1.6	3.5
350	0.049565	0.024690
351	0.049565	0.024690
352	0.048667	0.025166
353	0.048166	0.025108

7. Storage

Samples should be stored inside a dry cool room in glass cans/jars along with their sample code. Samples should be organized in a way that they can be easily found when necessary.

References

Ben Dor, E., Ong, C., and Lau, I.C. (2015). Reflectance measurements of soils in the laboratory: Standards and protocols. Geoderma 245–246, 112–124.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., et al. (2016). A global spectral library to characterize the world's soil. Earth-Sci. Rev. 155, 198–230.

END OF DOCUMENT