

Eberswalde University for Sustainable Development – HNEE  
University of Applied Sciences  
Faculty of Forest and Environment

Warsaw University of Life Sciences WULS – SGGW in Warsaw  
Faculty of Forestry

Akhil Chandran  
Album number HNEE: 19212834  
Album number SGGW: 200854

# Soil organic carbon estimation using proximal and remote sensing at field-scale level

Schätzung des organischen Kohlenstoffs im Boden mit Hilfe der Nah-  
und Fernerkundung auf Feldniveau

Master Thesis on the course of Forest Information Technology

Thesis written under the supervision of

Prof. Dr. Luis Miranda  
Faculty of Forest and Environment  
Eberswalde, Germany

Dr. Kathrin Grahmann  
Leibniz Center for Agricultural Landscape Research (ZALF) e. V.  
Müncheberg, Germany

Eberswalde, 2023



### **Oświadczenie promotora pracy**

Oświadczam, że niniejszy rozdział pracy dyplomowej przygotowanej przez autora został przygotowany pod moim kierunkiem i stwierdzam, że spełnia warunki do przedstawienia tej pracy w postępowaniu o nadanie tytułu zawodowego.

### **Declaration of the promoter**

I declare that this thesis was prepared under my supervision and I state that it meets the conditions for presenting such a body of work in the process of obtaining a professional title.

### **Erklärung des/der Betreuers/-in**

Hiermit erkläre ich, dass die vorliegende Arbeit unter meiner Leitung erstellt wurde und ich bestätige, dass sie die Bedingungen zur Verleihung des Abschlussdiploms erfüllt.

Data .....

Date

Datum

Podpis promotora pracy .....

Signature of the promoter

Unterschrift des/der Betreuers/-in

### **Oświadczenie autora pracy**

Świadom odpowiedzialności prawnej, w tym odpowiedzialności karnej za złożenie fałszywego oświadczenia, oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami prawa, w szczególności ustawą z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz. U. Nr 90 poz. 631 z późn. zm.)

Oświadczam, że przedstawiona praca nie była wcześniej podstawą żadnej procedury związanej z nadaniem dyplomu lub uzyskaniem tytułu zawodowego.

Oświadczam, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Przyjmuję do wiadomości, że praca dyplomowa poddana zostanie procedurze antyplagiatowej.

### **Declaration of the author**

Aware of the legal liability, including criminal liability for submitting a false statement, I declare that this thesis was written by myself alone and does not contain content obtained in a manner breaking applicable laws, in particular the Act of February 4, 1994 on copyright and related rights (Journal of Laws, no. 90, item 631, as amended)

I certify that the work has not previously been the basis for any procedure in connection with obtaining a diploma or professional title.

I declare that this version of the work is identical with the attached electronic version.

I acknowledge that the thesis is subject to anti-plagiarism procedures.

### **Erklärung des Autors**

Gesetzlicher Haftpflicht, besonders strafrechtlicher Verantwortlichkeit für Abgabe einer falschen Erklärung bewusst, erkläre ich hiermit, dass vorliegende Diplomarbeit selbständig angefertigt wurde und keinen Inhalt enthält, der widerrechtlich erworben wurde, insbesondere nicht mit dem Gesetz über Urheberrecht vom 4. Februar 1994 (GB. Nr. 90, Pos. 631 mit späteren Änderungen) übereinstimmend.

Ich erkläre auch, dass die Arbeit bisher keiner anderen Prüfungsbehörde vorgelegt wurde.

Der Durchführung einer elektronischen Plagiatsprüfung stimme ich hiermit zu. Die eingereichte elektronische Fassung der Arbeit entspricht der eingereichten schriftlichen Fassung exakt.

Data: .....

Date

*Datum*

Podpis autora pracy .....

Signature of the author

*Unterschrift des Autors*



## Podsumowanie

### **Tytuł: Szacowanie węgla organicznego w glebie z wykorzystaniem proksymiki i teledetekcji w skali pola.**

Węgiel organiczny w glebie (SOC) jest ważnym czynnikiem środowiskowym, który wpływa na jakość i funkcje gleby, globalne bezpieczeństwo żywnościowe oraz wysiłki zmierzające do złagodzenia zmian klimatu. Niezbędne jest dokładne oszacowanie i przewidzenie poziomów SOC na dużą skalę. Podczas gdy spektroskopia poprzez proksymalną detekcję jest skuteczna w dokładnym przewidywaniu poziomów SOC, jej ograniczenia w szacowaniu SOC na dużą skalę przestrzenną są niepokojące. Dlatego też istnieje potrzeba szybszych i bardziej opłacalnych technik ilościowego określania zawartości SOC. W ostatnich badaniach zaproponowano wykorzystanie metod teledetekcji jako potencjalnego rozwiązania. Głównym celem tego badania była ocena i porównanie proksymalnego sensingu gleby (pomiar spektroskopii polowej) w monitorowaniu i szacowaniu poziomu SOC z danymi uzyskanymi z satelity (Sentinel 2A) i bezzałogowego statku powietrznego (UAV) na polu rolnym. Aby zwiększyć dokładność metod teledetekcyjnych (UAV i Sentinel-2A) w przewidywaniu poziomów SOC, stworzono dziewięć indeksów spektralnych. W procesie modelowania wykorzystano różne pasma i długości fal, a konkretnie cztery pasma dla UAV i osiem dla Sentinel-2A. Dodatkowo, obliczone indeksy spektralne zostały wykorzystane jako zmienne niezależne do stworzenia modeli predykcyjnych dla zawartości gleby, model random forest (RF) jest trenowany z 90 próbek gleby zebranych z pola, oraz dziesięciokrotnej walidacji krzyżowej. Przed przeprowadzeniem predykcji SOC, w pracy zbadano impakt kowariancji. Modele utworzone z danych proksymalnych miały lepszą dokładność w tworzeniu prognoz za pomocą lasu losowego (RF) w porównaniu do dwóch pozostałych typów danych. Badanie wykazało, że technologie proksymalne i teledetekcyjne mogą być skutecznie wykorzystane do prognozowania SOC.

Słowa kluczowe: Soil Organic Carbon, Remote Sensing, Proximal Sensing, Soil sampling, Random Forest

## Summary

**Title: Soil organic carbon estimation using proximal and remote sensing at field-scale level.**

Soil organic carbon (SOC) is an important environmental factor that impacts soil quality and function, global food security, and efforts to mitigate climate change. It is essential to accurately estimate and predict SOC levels on a large scale. While spectroscopy through proximal sensing is effective in accurately predicting SOC levels, its limitation in estimating SOC on a large spatial scale is a concern. Hence, there is a need for faster and more cost-effective techniques for quantifying SOC content. Recent research has proposed the use of remote sensing (RS) methods as a potential solution. The main objective of this research was to assess and compare the proximal soil sensing (field spectroscopy measurements) in monitoring and estimating SOC levels with data obtained from spaceborne Unmanned Aerial Vehicle (UAV) and Sentinel-2A on an agricultural field. To improve the accuracy of the RS methods (UAV and Sentinel-2A) in predicting SOC levels, nine spectral indices were created. The modelling process involved the use of different bands and wavelengths, specifically four bands for UAV and eight for Sentinel-2A. In addition, the computed spectral indices were used as independent variables to create prediction models for soil content, Random Forest (RF) model is trained with 90 soil samplings collected from the field, and validated by ten-fold cross-validation. Prior to conducting the SOC predictions, the study investigated the covariate importance. The models created from proximal sensing data had better accuracy in making predictions with the help of RF compared to the other two methods. The study demonstrated that proximal and remote sensing technologies can be exploited efficiently for SOC prediction.

Keywords: Soil Organic Carbon, Remote Sensing, Proximal Sensing, Soil Sampling, Random Forest.

## **Zusammenfassung**

### **Titel: Schätzung des organischen Bodenkohlenstoffs mit Hilfe von Nah- und Fernerkundung auf Feldniveau.**

Der organische Kohlenstoff im Boden (SOC) ist ein wichtiger Umweltfaktor, der sich auf die Qualität und Funktion des Bodens, die weltweite Ernährungssicherheit und die Bemühungen zur Eindämmung des Klimawandels auswirkt. Es ist von entscheidender Bedeutung, den SOC-Gehalt in großem Maßstab genau zu schätzen und vorherzusagen. Während die Spektroskopie mit Hilfe der proximalen Abtastung bei der genauen Vorhersage des SOC-Gehalts wirksam ist, ist ihre Begrenztheit bei der Schätzung des SOC auf einer großen räumlichen Skala ein Problem. Daher besteht ein Bedarf an schnelleren und kostengünstigeren Verfahren zur Quantifizierung des SOC-Gehalts. Jüngste Forschungsarbeiten haben den Einsatz von Fernerkundungsmethoden als mögliche Lösung vorgeschlagen. Das Hauptziel dieser Studie war die Bewertung und der Vergleich der proximalen Bodenerfassung (Feldspektroskopie-Messungen) bei der Überwachung und Schätzung des SOC-Gehalts mit Daten, die von weltraumgestützten (Sentinel 2A) und unbemannten Luftfahrzeugen (UAV) auf einem landwirtschaftlichen Feld gewonnen wurden. Um die Genauigkeit der Fernerkundungsmethoden (UAV und Sentinel-2A) bei der Vorhersage des SOC-Gehalts zu verbessern, wurden neun Spektralindizes erstellt. Bei der Modellierung wurden verschiedene Bänder und Wellenlängen verwendet, insbesondere vier Bänder für UAV und acht für Sentinel-2A. Darüber hinaus wurden die berechneten Spektralindizes als unabhängige Variablen verwendet, um Vorhersagemodelle für den Bodengehalt zu erstellen. Das Random-Forest-Modell (RF) wurde mit 90 Bodenproben aus dem Feld und einer zehnfachen Kreuzvalidierung trainiert. Vor der Durchführung der SOC-Vorhersagen wurde in der Studie die Bedeutung der Kovariablen untersucht. Die Modelle, die auf der Grundlage von Naherkundungsdaten erstellt wurden, wiesen im Vergleich zu den anderen beiden Datentypen eine höhere Vorhersagegenauigkeit mit Hilfe von RF auf. Die Studie zeigte, dass Nah- und Fernerkundungstechnologien effizient für die SOC-Vorhersage genutzt werden können.

Schlüsselwörter: Organischer Bodenkohlenstoff, Fernerkundung, Naherkundung, Bodenprobenahme, Random Forest.

## Table of Contents

I. INTRODUCTION .....	9
1. Review of the literature.....	11
1.1 Data collection.....	11
1.2 Modelling.....	18
2. Research objective .....	20
II. MATERIALS AND METHODS.....	21
1. Workflow.....	21
2. Description of input data.....	22
2.1 Manual soil sampling data .....	22
2.2 Proximal sensing data, Veris MSP3 .....	23
2.3 Remote sensing data .....	24
3. Preparation of data .....	25
3.1 Preprocessing of soil samplings.....	25
3.2 Preprocessing of Veris MSP3 data .....	26
3.3 Preprocessing of UAV imagery.....	26
3.4 Preprocessing of Sentinel-2A .....	26
4. Ground truth data selection.....	27
5. Modelling and prediction assessment .....	27
III. RESULTS AND DISCUSSION .....	28
1. Data inputs.....	28
2. Data outputs.....	30
3. Model structure .....	31
4. Creation of SOC maps from proximal and remote sensed data at the field scale.....	32
5. Evaluation of heterogeneity pattern of the SOC distribution at the field scale.....	33
6. Comparison of proximal and remote sensed SOC estimations with ground truth measurements.....	35
IV. CONCLUSION.....	37
Future Work.....	37
V. ACKNOWLEDGEMENT .....	38
VI. REFERENCES.....	39
VII. APPENDICES .....	45
List of Figures.....	45
List of Tables .....	47
List of Abbreviations .....	47



## I. INTRODUCTION

In the second half of the twenty-first century, the world's need for food, fiber, and energy is predicted to grow (Godfray et al., 2010), and this poses a challenge for farmers to increase their crop yields while preserving natural resources and promoting environmental sustainability (Preza Fontes et al., 2019). Crop production requires fertile soil, which can be hampered by intensive farming methods.

The scientific community has recognized the consequences of soil degradation. By 1987, the United Nations Environment Programme (UNEP), in collaboration with the International Soil Reference and Information Centre (ISRIC) called for the execution of the Global Assessment of Soil Degradation (GLASOD) project that formed a world map of human-induced soil degradation (Angelopoulou et al., 2020).

The understanding of the variability of soil characteristics allows for an improved demand-oriented and environmentally friendly agricultural management as well as a more efficient use of resources (Gholizadeh et al., 2018). Soil Organic Carbon (SOC) is an important factor in crop growth (Fleming et al., 2000), as it influences soil moisture infiltration, retention, soil texture, rooting depth, soil herbicide activity, nitrogen release, and other nutrient cycling aspects (Kweon et al., 2013). SOC is the organic fraction of the soil and consists of decomposed plant and animal materials as well as microbial organisms, except fresh and un-decomposed plant materials (Ladoni et al., 2010). It is a useful indicator of soil fertility and an important factor in controlling the dynamics of various agrochemicals in the soil. In addition, it affects the chemical, physical and biological properties of a soil ecosystem (Gholizadeh et al., 2018).

Hence, the quantitative and qualitative estimation of soil properties is a laborious process (Angelopoulou et al., 2019). To enhance monitoring and mapping capabilities, stable data sets that can provide credible information for estimating SOC content are needed (Sanchez et al., 2009). SOC and soil texture are vital parameters of agricultural soils and need to be monitored regularly (Chaudhari et al., 2013).

patchCROP is an experimental approach implemented by Leibniz-Centre for Agricultural Landscape Research (ZALF) e.V, to design a multifunctional and sustainable cropping system which was started in March 2020 under on-farm conditions (Figure 1). For this, a

70-ha field was grouped into high and low-yield potential zones (Donat et al., 2022). Further, these zones were divided into 30 patches, 72m x 72m in size. The patches are located in a small area of the agricultural field, and they are large enough to accommodate the use of standard field machinery. Also, a range of soil, crop, ecological, and abiotic data were collected.

The study area extends from latitude 52°26'51.2376" N and 14°8'27.9492" E (Figure 1). This is an agricultural production area, and field experiments are conducted as part of the patchCROP project. The agricultural field is managed by the Komturei Lietzen GmbH in Brandenburg.



Figure 1: Location of the study area

The structure of the thesis is organized as follows: Firstly, the literature review discusses various methods for collecting soil data. This section also provides detailed information on the workings of the Random Forest (RF) algorithm. In section II, the materials and methodology are presented, including the workflow, data preparation and description, ground truth data selection, modelling and prediction assessment. Section III presents the results and discussion, which includes SOC maps from different methods and a comparison of technologies. Section IV presents the conclusion of the study. Finally, the thesis includes a reference section and an appendix.

## **1. Review of the literature**

### **1.1 Data collection**

The different methods for estimating SOC content can be divided into analytical methods (i.e., dry and wet combustion), remote sensing-based methods (i.e., space- and airborne-based methods), spectroscopic methods (i.e., visible near-infrared and shortwave infrared spectroscopy (VNIR-SWIR), and mid-infrared spectroscopy (MIR)), and laser-induced decay spectroscopy and inelastic neutron scattering methods (Angelopoulou et al., 2020). The methods are discussed in the sections below.

#### **a) Laboratory methods**

SOC is typically estimated in the laboratory on soil samples collected from the field (Rochette & Bertrand, 2007). There are two types of tests for SOC estimation, one based on acid digestion and the other on the principle of combustion. The second estimation is based on the total carbon present in the soil sample, while the first one estimates only a part of the organic carbon (Nelson & Sommers, 2018). SOC results are commonly reported as % C by weight (i.e., g C per 100 g of soil). Very often it is more convenient to express SOC on a per ha basis, namely as tons of C per ha (Chan, 2008).

#### **b) Proximal sensing**

The use of field-based sensors that are close to the ground is known as proximal soil sensing, i.e., within a maximum distance of two meters (Rossel & Adamchuk, 2013). According to Kuang et al., (2012) proximal soil sensing in the field became interesting because of its potential benefits. These technologies either connect with on-the-go sensors that are mounted on agricultural vehicles or handheld tools that can be utilized for site-specific management (Christy, 2008). Waiser et al., (2007) said that, for a sufficient field-scale evaluation of soil heterogeneity, the application of proximal soil sensing techniques requires a greater number of measured soil samples. In the next section, the information related to the proximal soil sensor (Veris MSP3) used in this study are provided.

#### **(1) Veris Multi Sensor Platform 3 (MSP3)**

Veris MSP3 technologies launched a proximal optical sensor that estimates soil reflectance at two wavelengths. In 2002, Veris Technologies started the development of soil optical devices and a VNIR spectrometer system for mapping soil (Veris Technologies, Salinas, KS, USA) (Bönecke et al., 2020). It is an on-the-go optical soil

sensor consisting of a single photodiode and two light sources (LEDs) that provide reflectance measurements at 660 nm (red) and 940 nm (near infrared) with a bandwidth of 20 nm each (Bönecke et al., 2020).



Figure 2: Veris MSP3 Unit

According to Kweon & Maxton, (2013), organic carbon is particularly sensitive to these two wavelengths. At the front, the Optic Mapper has an aperture coultter that slices crop residue. The sensing depth is controlled by an optical module which is mounted in the bottom between two sides of the wheel. The wear plate is pushed towards the bottom of the furrow, approximately 0.04 m below the soil surface, at a uniform pressure to ensure a self-cleaning function. It emits alternating light from two LEDs and enters the soil through a sapphire window. The reflected light captures a photodiode and contains the light intensity at dimensionless values. Digital reflectance data and Global Navigation Satellite System (GNSS) coordinates are recorded at a rate of 1 Hz. It can record an average of 260 reflectance data points per hectare at a speed of 10-12 km per hour and a track distance of 12 m (Bönecke et al., 2020). SOC composition was predicted by a combination of red (660 nm), infrared (NIR, 970 nm), and optical reflectance measurements (Bönecke et al., 2020). A sapphire window in the bottom of a furrow 'shoe' is used for the soil measurements. Through the sapphire window modulated light is directed onto the soil. The photodiode received the reflected light and converted it to a modulated voltage, which is further processed and logged.

The device consists of a depth-control row unit, an optical module, electronics for signal conditioning, a data logger and a Global Positioning System (GPS). It contains six coulter electrodes for Electric Conductivity (EC) measurements, a specially-configured row unit for optical measurements and an Ion Selective Electrode (ISEs) and soil sampling shoe for pH measurements. Rolling coulters are inserted into the soil and the module measures EC. Based on the established practice of measuring the soil EC in situ, the system maps soil texture. Smaller soil particles such as clay conduct more current than larger silt and sand particles. Electrical current is injected into the soil by one pair of coulters in the module. Other five coulter measured the voltage changes. The measurement from one pair is for a "shallow" EC (0–30 cm) and the other is for "deep" EC (0–90 cm)(G. Kweon, E. Lund & Veris, 2012).

While travelling across the field, the soil pH mapping unit automatically collected a soil sample and recorded its geographic position. The sampler shank is lowered into the soil surface during soil sampling. The collected soil presses against the two electrodes for two separate measurements and then the arithmetic mean of the measurements are recorded. Then the sampler shank is lowered again into the soil and as the new sample enters from its front end, the old soil core at the back end is replaced. Before the next measurement, the electrodes are cleaned with water by two spray nozzles. By a preceding calibration with pH4 and pH7 standard solutions, voltage is converted to pH value. Every 10 to 12 seconds pH values are recorded (Vogel et al., 2021).

### **c) Remote sensing (RS) methods**

Mulder et al., (2011) suggested that RS can be used to gather qualitative and quantitative data on soil characteristics and classification. RS technologies allow for better gathering of data over vast regions with higher precision in terms of time and space. These technologies are increasingly becoming accessible to individual farmers, aiding in monitoring, awareness, and informed decision making. (Preza Fontes et al., 2019). Recent studies introduce RS techniques as rapid, low-cost, and non-destructive assessment, for the estimation of various soil properties, including SOC among others (Angelopoulou, 2019). RS techniques are based on aerial and satellite imagery. These can be used to estimate SOC at the landscape level (Kweon et al., 2013). Due to the relationship between electromagnetic radiation and the soil's complex chemical bonds, RS imagery has huge

potential for creating soil profile maps (Biney et al., 2021). In the 1980s, satellites were used for comprehensive SOC assessment (Frazier & Cheng, 1989).

The operation of the VNIR-SWIR used for RS applications is based on the principles of energy-matter interaction. The electromagnetic radiation radiated at the surface of the soil is reflected at different wavelengths, resulting in a spectrum that determines the fraction of radiation that occurs. This spectrum encodes information to obtain qualitative and soil properties. VNIR–SWIR spectroscopy is based on characteristic vibrations of chemical bonds in molecules. In particular, electronic transitions in the visible region (400–700 nm) form broad absorption bands associated with chromophores that affect the soil colour, while in the NIR–SWIR (700–2500 nm) weak overtones and combinations of these vibrations occur due to stretching and bending of the N-H, O-H, and C-H bonds. Ben-Dor (2011) evaluated the NIR measurements and concluded that the OH groups have strong absorption characteristics in the range of 1400-1900 nm, mostly due to the soil water content, hydroxyl content, and clay content. It has also been observed that the reflectance of soils at certain wavelengths can be correlated with organic components (cellulose, lignin, starch), and provides valuable qualitative and quantitative information.

Based on the principles of energy-matter interaction, a material can reflect, absorb, scatter, and emit electromagnetic radiation in a distinctive manner that depends on its molecular composition and shape, resulting in a unique spectral signature. A sensor can measure the reflection of an object over a wide area of a wavelength, which provides information about its components (Angelopoulou et al., 2020). There are gradients within fields that have different colours. These gradients vary with soil type and are related to the soil's ability to retain nutrients and moisture. Red, Green and Blue (RGB), near-infrared spectroscopy (NIRS), and other sensors can quantify these properties throughout the soil profile (Murray, 2017).

RS methods for data collection can be divided into two categories: airborne (UAV) and spaceborne (e.g., Sentinel-2A). At broad spatial scales, these technologies are being increasingly used for soil mapping and yield forecasting in the recent decades (Khanal et al., 2020). It has been observed that the prediction accuracy decreases from UAV to satellite platforms (Guio Blanco et al., 2018).

## **(1) Unmanned Aerial Vehicle (UAV)**

UAV hyperspectral imaging has the ability to spatially assess soil conditions to obtain a more accurate imaging of observational parameters in agricultural areas (Whitehead et al., 2014). From a single flight they can produce information on a large area. It provides hyperspectral data and helps to classify a site according to its soil diversity (Stevens et al., 2008). Aircrafts have the ability to carry excellent payloads, providing the ability to attach a wide spectral range of hyperspectral sensors. In addition, while having the added benefit of operating on high cloud coverage, ventilated mounted sensors make the particular measurement time window more flexible, giving it the ability to select the best flight conditions (Villa et al., 2016).

## **(2) Spaceborne**

Based on the demonstrated background of interactions between the specific chemical bonds of the soil and electromagnetic radiation, spaceborne remote sensing is very likely to be used as a tool to create spatial maps of the upper soil horizon (Biney et al., 2021). Applications based on hyperspectral data became mainstream a few years later when the Hyperion spaceborne system came online (Angelopoulou et al., 2019). Indeed, their use was limited to ground monitoring due to (i) required atmospheric, geometric, and radiometric data corrections, (ii) simultaneous geological observations, (iii) the difficulty of locating large bare ground areas in a single image and (iv) obstacles associated with vegetation cover (Angelopoulou et al., 2019). At present, SOC estimation and classification based on spatial data is subject to significant changes. The major milestone in the RS community is the United States Geological Survey (USGS) policy change that allows free accessing of Landsat data. In addition, Sentinel-2 is oriented to the advent of the Big Earth observation data era provided by European Space Agency (ESA), led by free and open super-spectral imagery and the advent of large fleets of small satellites (e.g., Planet Cubesats) (Angelopoulou et al., 2019).

### **(2.1) Sentinel-2A**

Two satellites that were launched in 2015 and 2017 are part of the Copernicus Sentinel-2 mission and are in the same sun-synchronous orbit. Sentinel-2A is a European optical imaging satellite launched on 23 June 2015 and later Sentinel 2B was launched on 07 March 2017. The satellite has a high-resolution multispectral broadband image with 13

spectral bands. The spectral range is from vis-NIR to SWIR. Table 1 shows the Sentinel-2A spectral bands information.

Table 1: Summary of Sentinel-2A spectral bands

<b>Band</b>	<b>Name</b>	<b>Spatial Resolution (m)</b>	<b>Central wavelength (nm)</b>	<b>Bandwidth (nm)</b>
1	Coastal Aerosol	60	443	21
2	Blue	10	492	66
3	Green	10	560	36
4	Red	10	665	31
5	Vegetation Red Edge 5	20	704	15
6	Vegetation Red Edge 6	20	741	15
7	Vegetation Red Edge 7	20	783	20
8	Near-Infrared (NIR)	10	833	106
8a	Narrow NIR	20	865	21
9	Water Vapor Absorption Window	60	945	20
10	Shortwave Infrared - Cirrus	60	1374	31
11	Shortwave Infrared 1 (SWIR 1)	20	1614	91
12	Shortwave Infrared 1 (SWIR 2)	20	2202	175

### (3) Indices

RS indices are tools that use satellite imagery to assess crops and soil, helping in monitoring and managing these resources for sustainable agriculture (Martos et al., 2021). According to Jin et al., (2017) soil optical properties can be altered by components such as soil water content, mineral composition, and organic matter concentration. Biney et al., (2021) used nine derived spectral indices which have been applied to the Sentinel-2A and UAV datasets as independent variables, in order to improve the datasets capacity to predict. The added spectral indices were Colour Index (CI), Normalized Differences Vegetation Index (NDVI), Infrared Percentage Vegetation Index (IPVI), Normalized Difference Red Edge (NDRE), Soil Adjusted Vegetation Index (SAVI), Vegetation (V), Green Normalized Difference Vegetation Index (GNDVI), Difference Vegetation Index

(DVI), and Brightness Index (BI). Below the detailed overview of the indices is provided. Table 2 lists the spectral indices equation.

**Normalized Difference Vegetation Index (NDVI)** is the furthestmost popular and extensively used vegetation index for monitoring the Earth's vegetation cover(Matsushita et al., 2007). NDVI varies between -1.0 and +1.0.

**Normalized Difference Red Edge (NDRE)** is a spectral index that is built as a blend of several bands: NIR spectrum and a band that uses a narrow spectral range between visible Red and NIR. While analyzing NDRE images, the color palette is typically red to green, where red is bare soil, pale yellow to pale green is unhealthy plants or vegetation that is early in its maturation, and lastly, green is a healthy canopy.

**Soil Adjusted Vegetation Index (SAVI)** a measure of vegetation that attempts to reduce the impact of soil brightness by applying a correction factor. It is commonly used in areas with low vegetation cover, particularly in arid regions.

**Green Normalized Difference Vegetation Index (GNDVI)** is a method of measuring plant photosynthesis and is a widely-used vegetation index for assessing the absorption of water and nitrogen in the plant canopy.

**Difference Vegetation Index (DVI)** is an index which distinguishes between soil and vegetation, but it does not account for the difference between reflectance and radiance caused by atmospheric effects or shadows.

**Brightness Index (BI)** is sensitive to the brightness of soils. High soil brightness is linked with soil humidity and the presence of salts in the soil.

**Color Index (CI)** is to identify and distinguish between different types of land use and land cover.

**Infrared Percentage Vegetation Index (IPVI)** is linearly equivalent to the NDVI, but have the advantage of a fully non-negative range.

**Vegetation Index (VI)** is indirectly assessing soil characteristics such as soil moisture as they are linked to vegetation health.

Table 2: Spectral indices equation

Index	Definition based on UAV	Definition based on Sentinel-2A	References
CI	$\frac{Red - Green}{Red + Green}$	$\frac{B4 - B3}{B4 + B3}$	Pałas & Zawadzki, 2020
NDVI	$\frac{NIR - Red}{NIR + Red}$	$\frac{B8 - B4}{B8 + B4}$	Wilson & Sader, 2002
IPVI	$\frac{1}{2}(NDVI + 1)$	$\frac{1}{2}(NDVI + 1)$	Crippen, 1990
NDRE	$\frac{NIR - Red\ Edge}{NIR + Red\ Edge}$	$\frac{B8 - B5}{B8 + B5}$	Barnes et al., 2000
SAVI	$\frac{(NIR - Red) * (1 + L)}{NIR - Red + L}$	$\frac{(B8 - B4) * (1 + L)}{B8 - B4 + L}$	Pałas & Zawadzki, 2020
GNDVI	$\frac{NIR - Green}{NIR + Green}$	$\frac{B8 - B3}{B8 + B3}$	Gitelson et al., 1996
DVI	$NIR - Red$	$B8 - B4$	Naji, 2018
BI	$\frac{\sqrt{(Red * Red) + (Green * Green)}}{2}$	$\frac{\sqrt{(B4 * B4) + (B3 * B3)}}{2}$	Escadafal, 1989
VI	$\frac{NIR}{Red}$	$\frac{B8}{B4}$	Roncagliolo et al., 2012

## 1.2 Modelling

### a) Machine learning

Machine learning (ML) technique enables the prediction of a dependent variable using a variety of independent variables. Many of the techniques employed in digital SOC mapping are based on linear relationships. Recent developments in computer science have facilitated the utilization of machine learning methods to map SOC (Judge, 2007). Lamichhane et al., (2019) reported that in recent years there has been a clear transition from linear models to ML models for SOC predictive mapping. Digital SOC mapping is another area that is growing in favour for ML approaches because they can handle complicated, non-linear interactions between soil characteristics and predictor factors (Drake et al., 2006).

### b) Random forest

ML methods like RF are used for classification and regression (Gambill et al., 2016). Breiman (2001) designed the RF algorithm to improve the accuracy of regression and classification. The algorithm consists of a collection of decision trees, often known as a

forest or an ensemble (Guio Blanco et al., 2018). By using predictor values, the decision trees are randomly generated (Segal, 2004). The method utilizes a regression-based voting strategy (Cootes et al., n.d.). RF has the major advantage of being able to precisely explain the compound relationships between the independent and dependent variables (Emadi et al., 2020).

Liu et al., (2006) & Biney et al., (2021) use the RF algorithm to estimate SOC by using proximal and RS data. They claim that traditionally, regression trees have been susceptible to overfitting and are sensitive to outliers. According to John et al., (2020) RF regression method produces better SOC results than a cubist model, Artificial Neural Network (ANN), Support Vector Machine (SVM), and Multiple Linear Regression (MLR). According to Sothe et al., 2022, for soil mapping RF showed the highest predictive accuracy and robustness regarding model parameters.

Scientists used various measurements to determine the precision of the model's predictions. Wang et al., (2021) used the Root Mean Squared Error (RMSE) and higher coefficient of determination ( $R^2$ ) in his research. RMSE and  $R^2$  are two common metrics used to evaluate the performance of a RF prediction model. RMSE is a measure of the average error between the actual and the predicted values. It represents the standard deviation of the residuals. A lower RMSE indicates a better fit between the predicted and actual values, and a higher RMSE indicates a worse fit. It is a measurement of the percentage of variance in the dependent variable that can be described by the independent variables.  $R^2$  ranges from 0 to 1, with a value of 1 indicating that the model perfectly fits the data and a value of 0 indicating that the model does not explain any variance in the data (Farooq et al., 2022).

The research gap in soil organic carbon estimation using proximal and remote sensing technologies at the field-scale level could be related to several areas. These areas include the lack of studies that compare the accuracy of different RS and proximal sensing technologies; limited research on the impact of different soil management practices on SOC estimation; the need for more studies on the integration of ML techniques and statistical models; limited research on generalization across different soil types and environmental conditions, and the need for more validation studies to assess the accuracy and reliability of soil organic carbon prediction models. Below section describes about the research objective.

## **2. Research objective**

The main objective is the SOC estimation using proximal and RS at the field-scale level. It is done by using soil samples, Veris MSP3, UAV and Sentinel-2A dataset. The goal is to compare and to find out the optimal SOC estimation approach. The research objective and tasks for this thesis is as follows:

“Soil organic carbon estimation using proximal and remote sensing at field-scale level.”

Tasks:

1. Suitability assessment of proximal (Veris MSP3) and remote sensing data (UAV, Sentinel-2A) to estimate SOC.
2. Creation of SOC maps from proximal and remote sensed data at the field scale.
3. Evaluation of heterogeneity pattern of the SOC distribution at the field scale.
4. Comparison of proximal and remote sensed SOC estimations with ground truth measurements.

## II. MATERIALS AND METHODS

The following sections describe the workflow, provides an overview of the included data, ground truth data selection, preprocessing of Veris MSP3, UAV and Sentinel-2A, random forest model and evaluation.

### 1. Workflow

The workflow described in Figure 3 has three sections: types of data input, data preparation and modelling. Section 2 describes the input data and section 3 explains the preprocessing of data. Section 4 and 5 describes the ground truth data selection and modelling respectively.

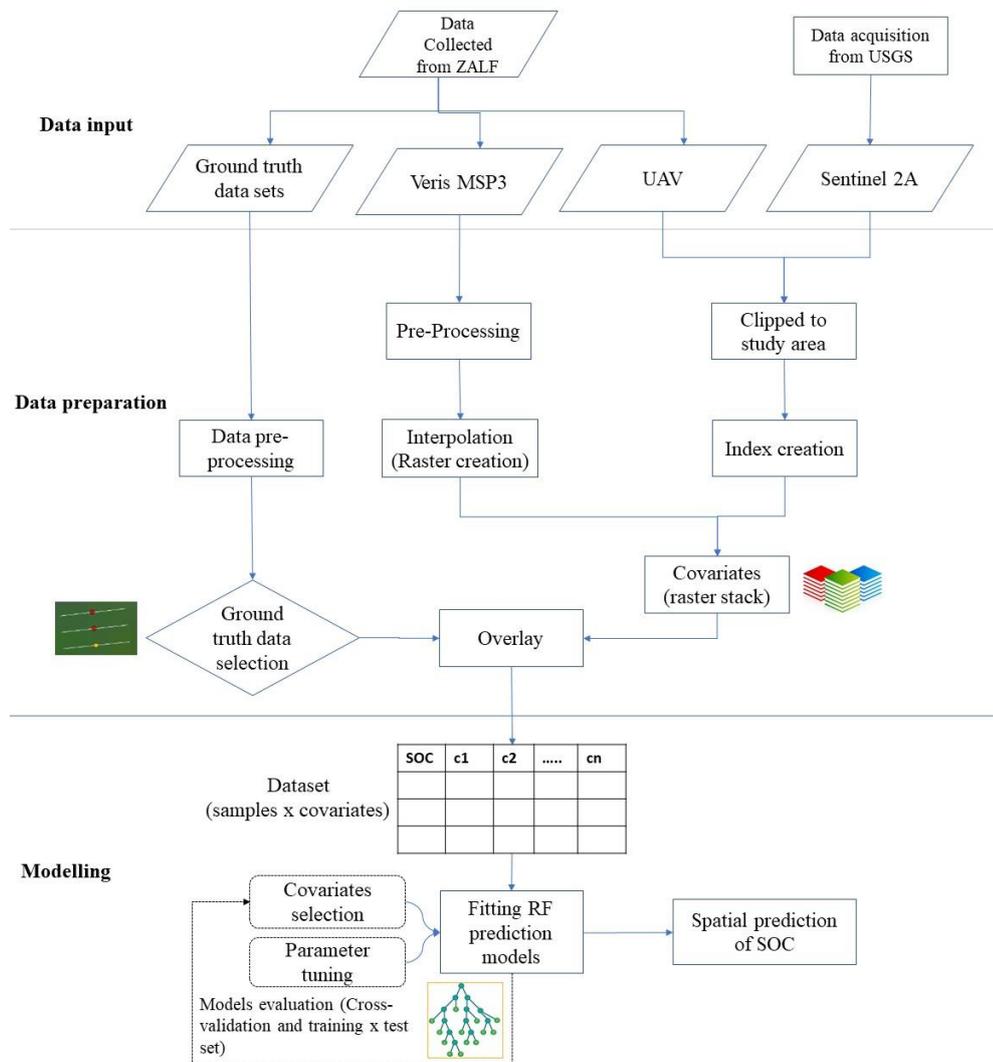


Figure 3: Methodological flowchart for modelling and mapping SOC

## 2. Description of input data

The data were collected through three different methods: (i) manual soil sampling and subsequent laboratory analysis, (ii) proximal sensing using Veris Mobile Sensor Platform (MSP3), and (iii) remote sensing using UAV and Sentinel-2A satellite imagery. Manual soil sampling data included three sampling campaigns. An initial sampling campaign was done on 10 December 2019, a second on June 2020 for Veris sensor data calibration and the third on 16 November 2020. Figure 5 picturise the timeline of data collection. In the following sections, a detailed overview of the data is provided.

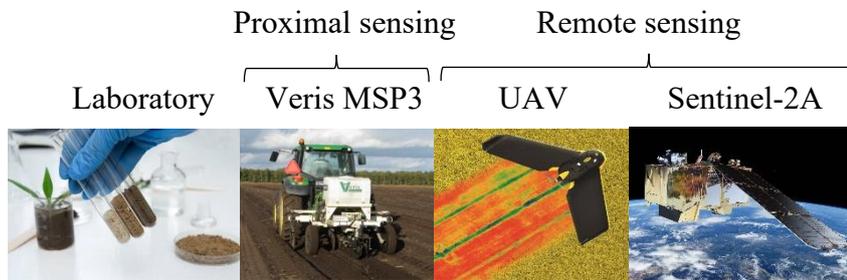


Figure 4: Soil spectra measured from different platform

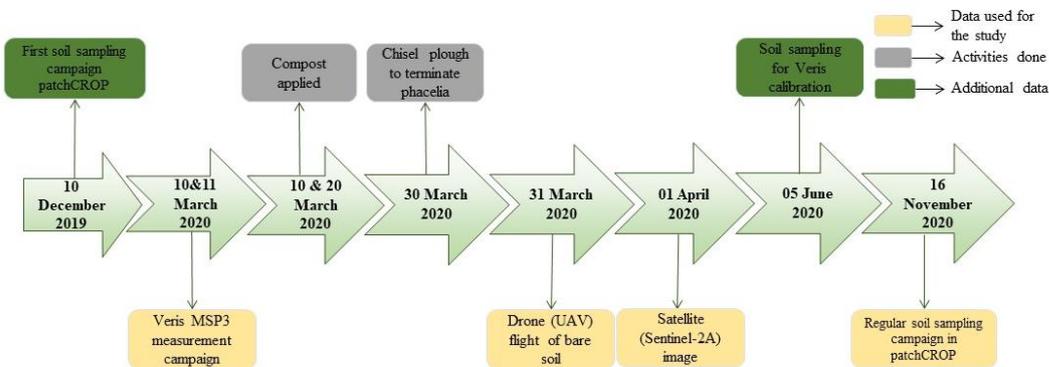


Figure 5: Timeline of data

### 2.1 Manual soil sampling data

Soil data collection was carried out for different purposes even before the implementation of the patchCROP experiment. In this study, three sets of manual soil samplings (Figure 6) were available. From these three datasets, SOC was analysed at ZALF's central laboratory and one sub-set of data was used for the ground truthing. The method used to determine SOC was according to DIN ISO 10694 using a RC 612 Leco (LECO

Corporation, St. Joseph, MI, USA). The section below provides more information about the manual soil samplings. Table 3 describes the three soil samplings.

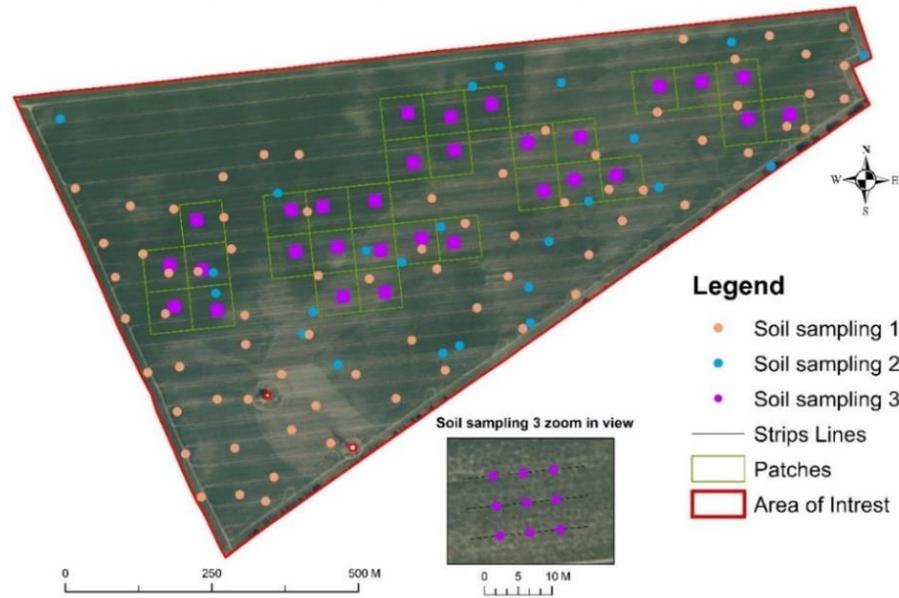


Figure 6: Soil sampling points

Table 3: Soil sampling detailed overview

Description	Date	Number of samples	Depth	Purpose
Soil sampling 1	10 December 2019	72 (25)	0 - 0.25m	Compost requisite analysis
Soil sampling 2	05 June 2020	10	0 - 0.3m	Veris data calibration
Soil sampling 3	16 November 2020	270 (90)	0 - 0.3m	patchCROP experiment regular sampling

(72(25), 270(90): The total collected 72 and 270 samples mixed together and final samples analysed in the laboratory is 25 and 90 respectively).

## 2.2 Proximal sensing data, Veris MSP3

Veris MSP3 was used to generate high-resolution soil data and subsequent SOC prediction. The data were collected on 10<sup>th</sup> and 11<sup>th</sup> of March 2020. During the calibration event in the field, the weather was partly cloudy and partly sunny with well-moistened soil. The field was driven along a permanent traffic track with an initial distance of 9 meters and an offset of 4.5 meters. The process was repeated with an offset of 18 meters, and then on 11.3 meters. A similar process was carried out with an offset of only 9 meters.

Figure 7 shows the point data collected using Veris MSP3. Table 4 shows the descriptive statistics of data collected using Veris MSP3.

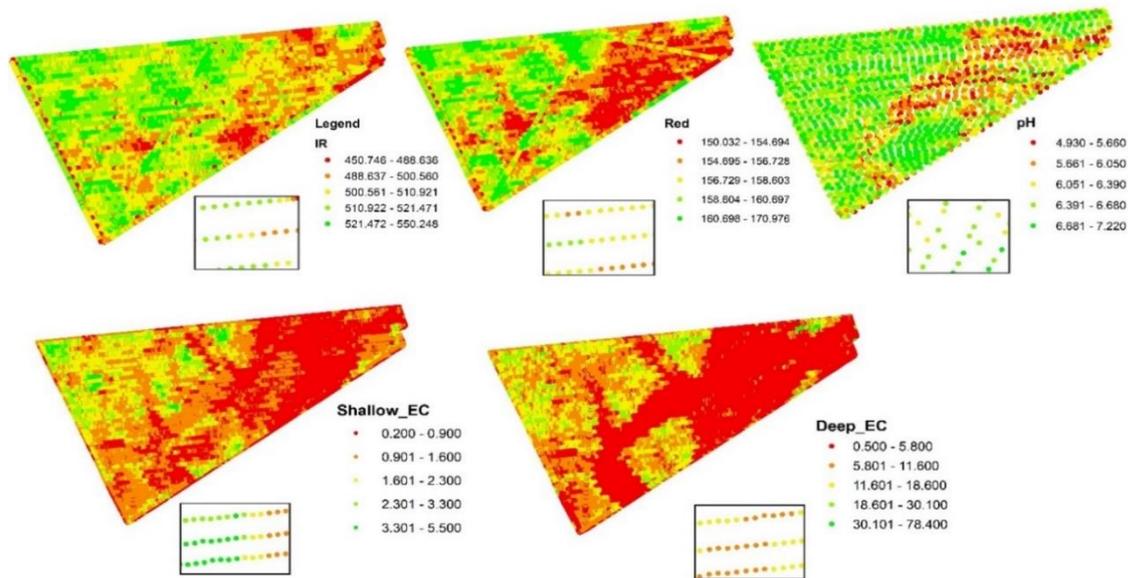


Figure 7: Veris MSP3 data collected on 10<sup>th</sup> & 11<sup>th</sup> March 2020  
(IR-Infrared, pH-potential of hydrogen, EC-Electrical conductivity)

Table 4: Descriptive statistics of Veris MSP3 variables

	Red	IR.	pH	Shallow EC	Deep EC
<b>Min.</b>	147.9	450.7	4.93	0.10	0.30
<b>1st Qu.:</b>	155.3	499.3	6.14	0.80	2.80
<b>Median</b>	157.2	508	6.48	1.20	6.0
<b>Mean</b>	157.3	507.3	6.379	1.31	7.86
<b>3rd Qu</b>	159.2	516.3	6.67	1.70	11.50
<b>Max.</b>	176.7	550.2	7.22	5.70	104.70

### 2.3 Remote sensing data

The following subsections describe both applications and the preprocessing of UAV and Sentinel-2A.

#### a) UAV

An UAV flew above the patchCROP experiment on 31 March, 2020. The images were captured while flying at a height of 70 meters., with a spatial resolution of 12.6 cm/pixel for multispectral and 2.83 cm/pixel for RGB images. The used UAV was the "RS eBee" (SenseFly Ltd.), which was equipped with the multispectral camera "Parrot Sequoia" and the 3D mapping RGB camera "senseFly SODA." (Figure 8). For each flight, the

multispectral camera took four spectral images with a wavelength of 530-570nm (green), 640-680nm (red), 730-740nm (red edge), and 770-810nm (NIR). Images were geotagged for orthomosaic processing using "Pix4D Mapper software".



Figure 8: RGB imagery from UAV on 31 March 2020

### **b) Sentinel-2A**

This study used images from Sentinel-2A taken on 01 April 2020, downloaded from the USGS Earth Explorer website. Level 2A Sentinel-2 imagery was used in this study (Figure 9).

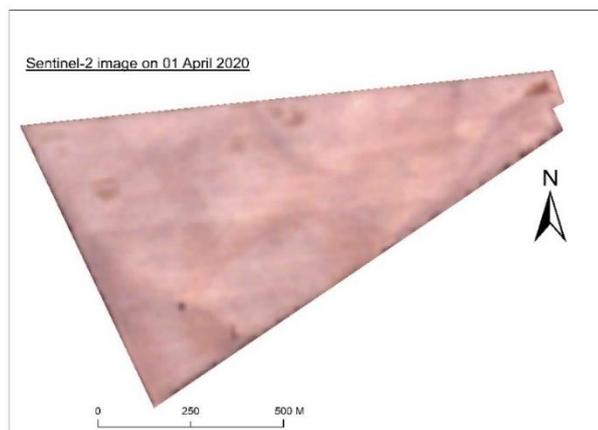


Figure 9: RGB imagery from Sentinel-2A on 01 April 2020

## **3. Preparation of data**

### **3.1 Preprocessing of soil samplings**

For laboratory analysis of the first soil sampling campaign, 72 samples were collected and reduced to 25 composite samples. The mixed sample consist of 2 to 5 single sampling

points taken in the entire field. These samples were analysed for SOC content in the commercial lab of Eurofins (JenaBios laboratory) and reanalysed in ZALF's central laboratory.

Soil sampling 2 was carried out for Veris data calibration, 10 samples were collected (Figure 6). These samples were analysed for SOC content at IGZ and reanalysed in ZALF's central laboratory.

Soil sampling 3 was carried out in the soil quadrant (18x18m) of each patch by dividing the quadrant into three strips. In each strip, a composite sample of three auger points was mixed (total of 90 points) for subsequent laboratory analysis at ZALF's central laboratory. All of the samples were in an accuracy of 4.5m distance and all three soil sampling campaigns obtained SOC values measured with the same lab method as sampling 1+2 used archived samples for reanalysis to exclude laboratory bias.

### **3.2 Preprocessing of Veris MSP3 data**

The data collected by the Veris system are stored in two different point shape files. To find the correlation with the ground truth data, these datasets were converted to raster format by using the kriging interpolation. Figure 11 is showing the images after kriging. Kriging is used 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 (Shekhar & Xiong, 2008). From the raster files, the sensor values at the 90 reference sampling points were extracted.

### **3.3 Preprocessing of UAV imagery**

The orthomosaic images of four bands (green, red edge, NIR, and red) were loaded to the R script, and further image processing was done using the R script and ArcMap. The study area is cropped from the orthomosaic images using the Area of Interest (AOI) shape file and indices were calculated. Figure 12 provides the results after indices calculation.

### **3.4 Preprocessing of Sentinel-2A**

The Sentinel 2A data clipped to the study area using shape file and selected 8 bands for further process. For this study, nine calculated spectral indices (Figure 13) were created using the 8 bands and applied to the Sentinel-2A dataset as independent variables to improve the predictive ability of the datasets.

#### 4. Ground truth data selection

To obtain more accurate results, more ground truthing data are required as dependent variable. From the three available datasets, the third manual sampling had a greater number of points and diversity in values compared to the other two datasets. Soil sampling 1 was conducted before the implementation of the patchCROP experiment and the number of samples was less compared to soil sampling 3. Soil sampling 2 was conducted for the Veris sensor data calibration and the number of samplings were also limited. Therefore, the manual soil sampling campaign 3 was used as training and testing data set to obtain accurate and precise results in this study.

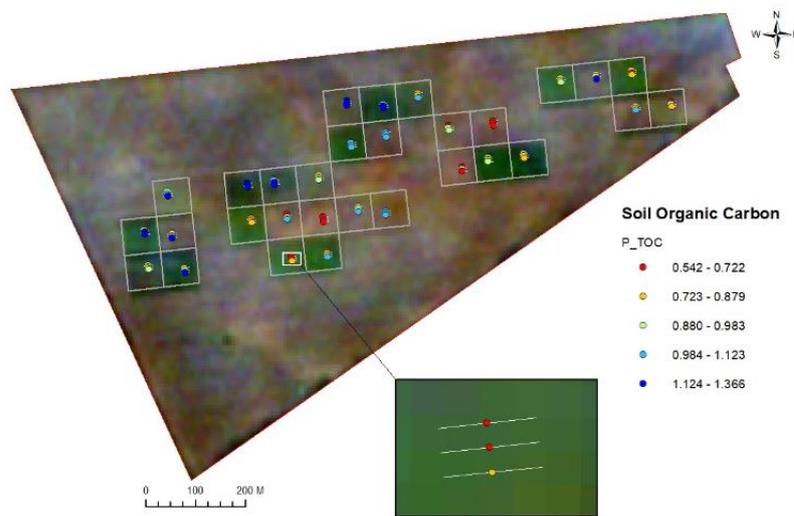


Figure 10: Soil sampling points of sampling campaign conducted on 16 November 2020, (Background imagery is from Sentinel 2A on 04 November 2020.)

#### 5. Modelling and prediction assessment

Data preprocessing and RF modelling were carried out using R Script and ArcGIS. The following libraries are used for the study: - Caret, randomForest, Raster, rgdal, RStoolbox, ggplot. RF model were used to predict the SOC from the three datasets. According to Sothe et al., (2022) RF showed the highest predictive accuracy and robustness regarding model parameters. Ten-fold cross-validation was performed on the training set, which constituted 80% of the samples. Additionally, the testing set which accounted for the remaining 20% of the samples, was also used to evaluate the model's performance. For each dataset, the RF model was different. The prediction accuracy was evaluated by index of determination ( $R^2_{CV}$ ), the Root Means Square Error of Prediction (RMSEP<sub>cv</sub>) of the ten-fold cross-validation. Prior to evaluating the predictive models, the normality of the distribution of the SOC contents was examined

### III. RESULTS AND DISCUSSION

#### 1. Data inputs

##### a) Veris MSP3

Figure 11 represents the preprocessed results from Veris MSP3. After the interpolation of points, IR result is ranging from 470 to 530 and it measured the reflectance or absorption of light by soil particles. From the preprocessed results, a topographical pattern is clear on IR, pH and Deep EC.

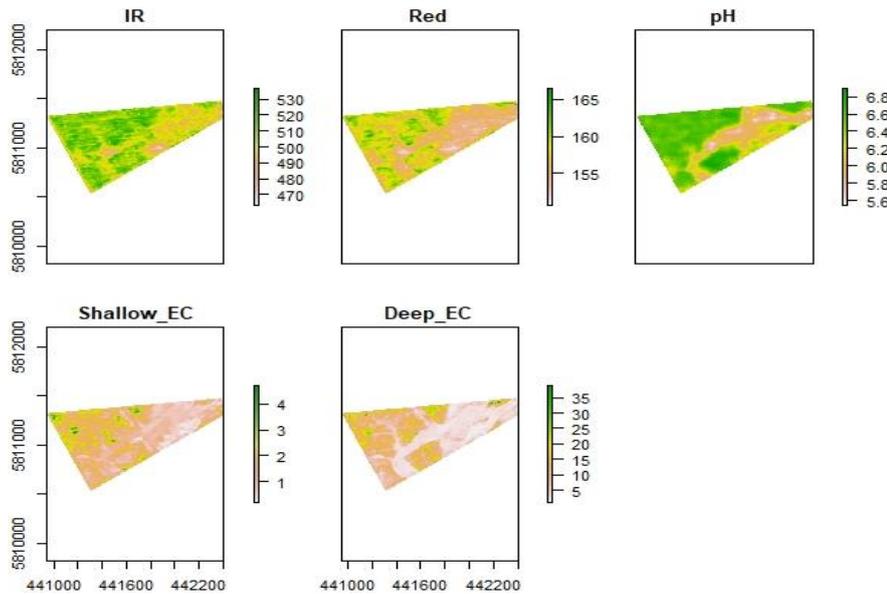


Figure 11: Variables measured by Veris MSP3 in raster form  
(EC – electrical conductivity)

##### b) UAV

Figure 12 shows the preprocessed results from UAV. The orthomosaic images have trajectory lines throughout the study area. After the problem analysis with the data collection team and Pix4D team, it was concluded that this might be due to solar radiation. If the solar radiation has changed significantly within the two flights, this stripe pattern can appear. Another reason could be the greater time difference between the neighbouring airlines on the western part as compared to the eastern part. If the cloud cover has changed between the start and the end of the flight, this will be visible in the striped pattern. In the eastern part, this is less pronounced as the time difference between neighbouring airlines is smaller. Two adjacent lanes pointing in the same direction (two lanes northbound, two lanes southbound, etc.). This could amplify the effect of the images changing when flying

towards the sun and when flying away from the sun. Hence, UAV data had errors during data collection.

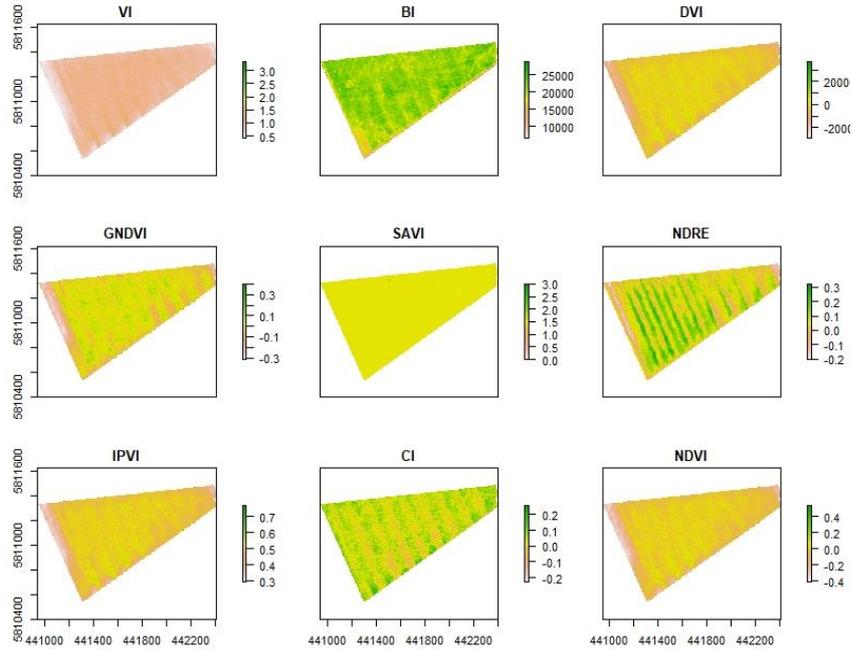


Figure 12: Spectral Indices from UAV

*(VI - Vegetation Index, BI - Brightness Index, GNDVI - Green Normalized Difference Vegetation Index, DVI - Difference Vegetation Index, SAVI - Soil Adjusted Vegetation Index, NDRE - Normalized Difference Red Edge, IPVI - Infrared Percentage Vegetation Index, CI - Color Index, NDVI - Normalized Difference Vegetation Index)*

### c) Sentinel-2A

Figure 13 shows the preprocessed results from Sentinel-2A. After indices calculation these indices are stacked together for retrieving the values using soil sampling 3. The BI, GNDVI and CI are clearly showing the patches in the study field.

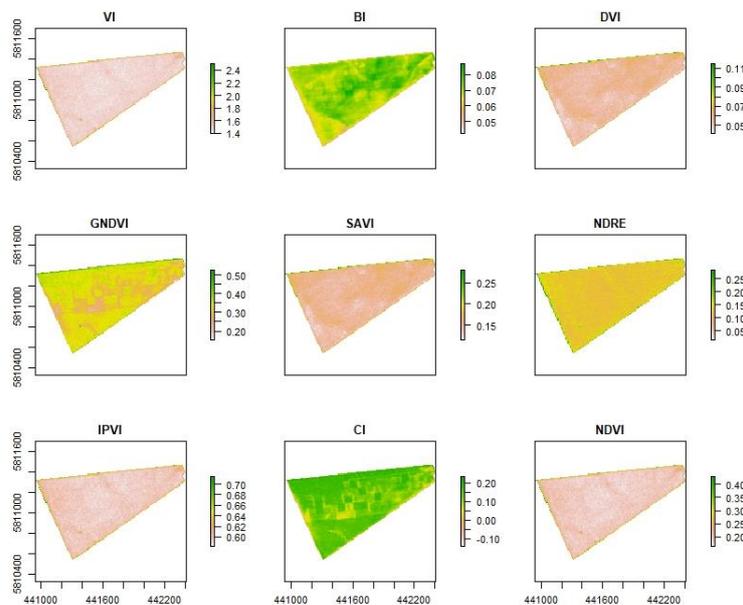


Figure 13: Spectral indices from Sentinel-2A

**d) Suitability assessment of proximal (Veris MSP3) and remote sensing data (UAV, Sentinel-2A) to estimate SOC.**

After the preprocessing of Veris MSP3, UAV and Sentinel-2A, the UAV data was excluded for modelling due to the error. Veris MSP3 is a proximal sensing technology that measures soil electrical conductivity, which is influenced by factors such as soil texture, soil moisture, and SOC. Sentinel-2A can provide spatially explicit data on soil reflectance properties, which can be related to soil properties such as SOC. The advantage of remote sensing techniques is their ability to cover large areas quickly and non-invasively, allowing for efficient monitoring of SOC over large regions. In this study, proximal and remote sensing data were used to estimate SOC.

**2. Data outputs**

**a) Ground truth data**

Figure 14 is a density histogram, boxplot and a statistics summary of SOC content in the third soil sampling campaign within the study area comprising mean value, median, standard deviation (SD), kurtosis, skewness, coefficient of variation (CV), minimum, maximum, and standard error (SE). The median value of SOC % was found to be 0.9245, which indicates that the distribution of SOC % is slightly skewed to the right with a skewness value of 0.0542. The distribution is almost normal, as evidenced by the kurtosis value of 0.5857, which suggests that the tails of the distribution are not very heavy compared to a normal distribution. The range was found to be 0.8240, which means that the minimum and maximum values were 0.5420 and 1.3660, respectively.

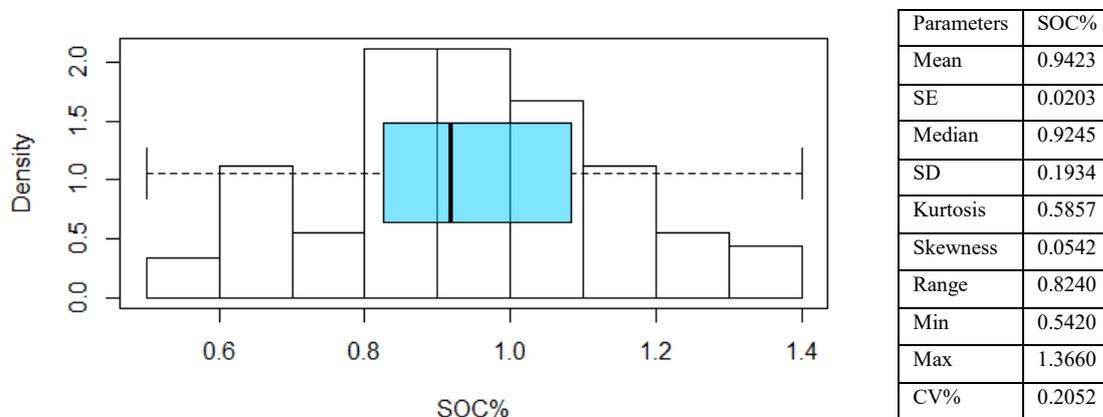


Figure 14: Density histogram, boxplot and statistics summary of SOC content (n=90)

### 3. Model structure

To make the RF model, the spectral bands obtained from Veris MSP3 and Sentinel-2A, including the determined spectral indices were each linked to the SOC determined in the laboratory (soil sampling 3) using collected soil samples from the field. Total of 90 soil samples, spectral bands and indices are used in this model. The RF regression model used the randomForest R Script library. 80% of the data for training the model and 20% for testing. RF model used 100 trees and optimised the hyperparameters. The model to predict SOC was chosen by selecting the all covariates in this study. Biney et al., (2021) used all covariates for SOC prediction in his research. RF regression is performed to understand initial important scores, the features with lowest importance scores. Figure 15 and 16 shows the covariate importance obtained by RF model. Based on the ten-fold cross-validation prediction error, multiple RF models were fitted repeatedly using all covariates.

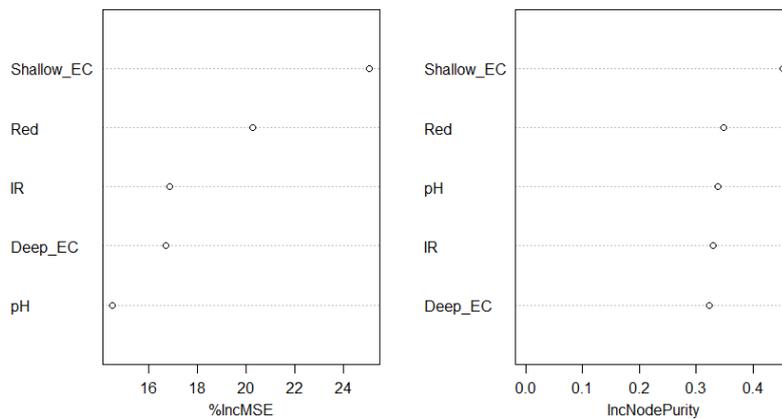


Figure 15: Covariate importance (ranking of predictors) from RF model fitting Veris MSP3 (%IncMSE: Mean Decrease Accuracy)

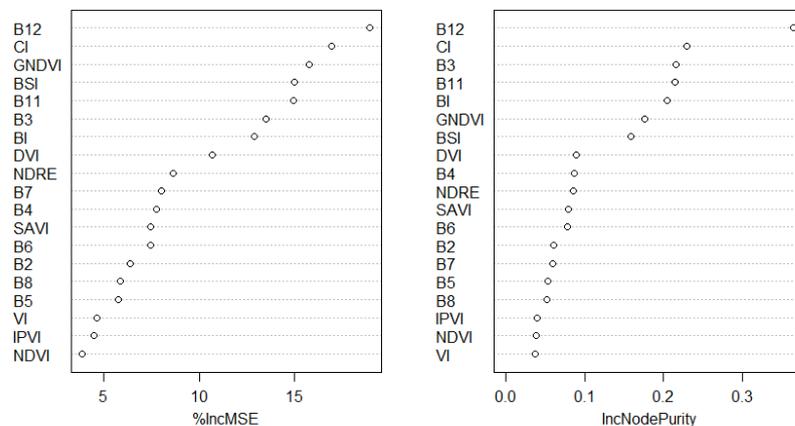


Figure 16: Covariate importance (ranking of predictors) from RF model fitting Sentinel-2A

**4. Creation of SOC maps from proximal and remote sensed data at the field scale.**

Figure 17 and 18 shows the SOC prediction using Veris MSP3 and Sentinel-2A. SOC contents ranged from 0.64% to 1.26 % from Veris MSP3 result and 0.59 % to 1.28% from Sentinel-2A. SOC distribution is varied for Veris MSP3 and Sentinel-2A.

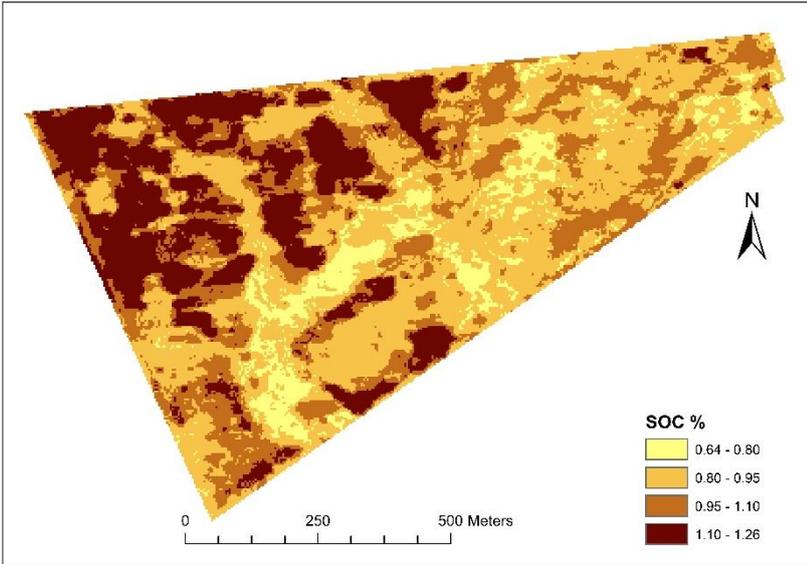


Figure 17: Spatial SOC distribution maps on study area based on prediction using Veris MSP3 dataset

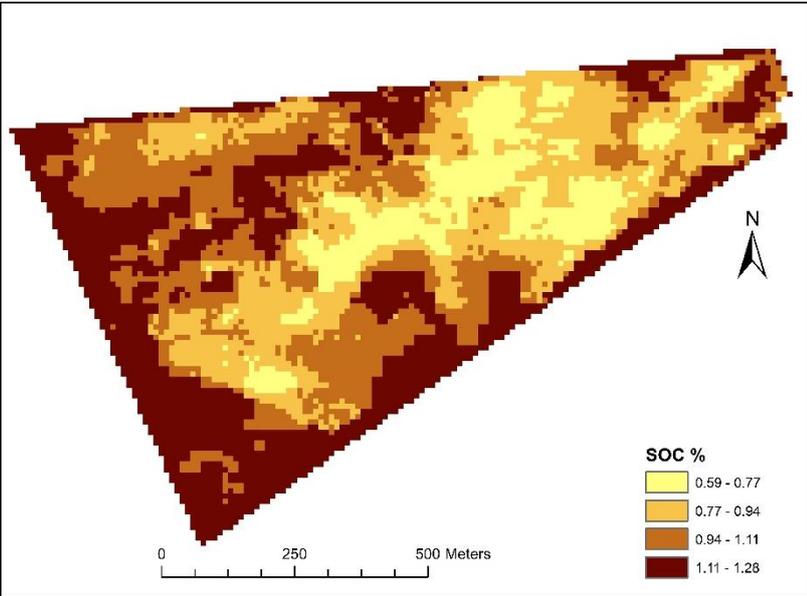


Figure 18: Spatial SOC distribution maps on study area based on prediction using Sentinel-2A dataset

## 5. Evaluation of heterogeneity pattern of the SOC distribution at the field scale.

Figure 19 and 20 shows SOC % in patchCROP study area. The study area included 30 patches in a 30x30 meter. For Veris MSP3, the overall distribution of SOC percentage across the study area ranged from 0.64% to 1.26% with a mean of 0.95%. For Sentinel-2A, the overall distribution of SOC percentage across the study area ranged from 0.59% to 1.28% with a mean of 0.91%.

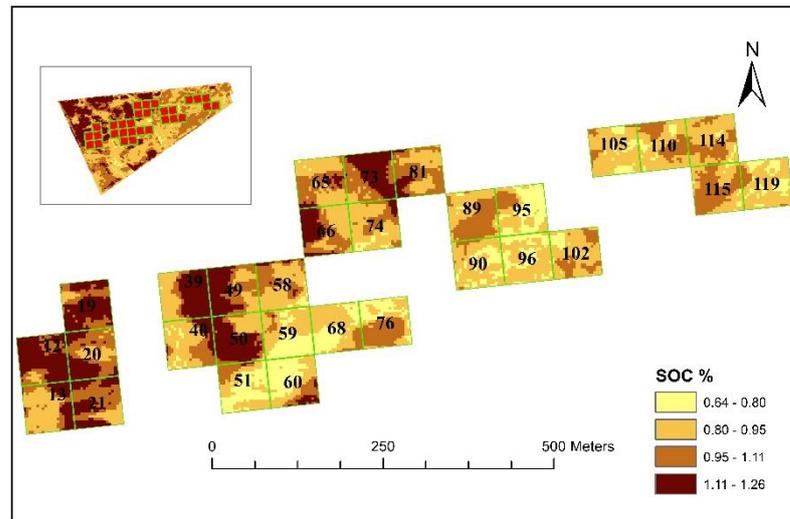


Figure 19: Spatial SOC distribution maps on patchCROP study area based on prediction using Veris MSP3

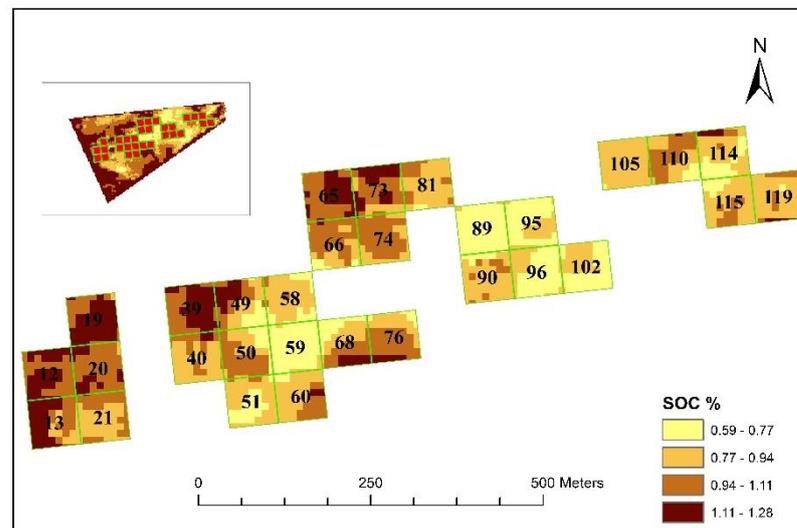


Figure 20: Spatial SOC distribution maps on patchCROP study area based on prediction using Sentinel-2A

Figure 21, 22 and 23 represents the boxplot of SOC percentage from Veris MSP3, Sentinel 2A and soil sampling respectively. X-axis represents the patch ID and Y-axis represents the SOC%. Boxplots of SOC % by patch showed considerable heterogeneity across different patches. For example, patch 19 and patch 59 is showing considerable

amount of SOC change in boxplot and also on the SOC distribution map. Similarly, from Veris MSP3 and Sentinel 2A SOC predictions, patches located on the south-west side have higher SOC% than central and north-east part. The results suggest that the heterogeneity of SOC % across different patches in the study area might be related to topography, the type of crops cultivated. Patches located on south-west may have higher SOC percentages due to differences in soil moisture and temperature. These patterns of heterogeneity may have importance in crop production.

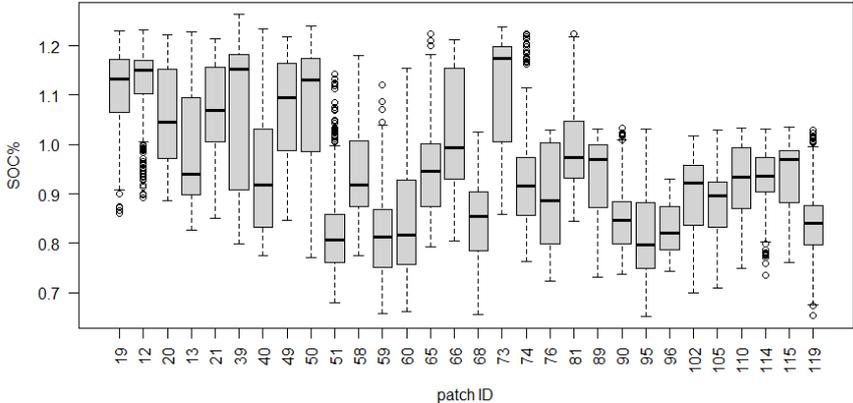


Figure 21: Boxplot of SOC patches from Veris MSP3 at the patch scale

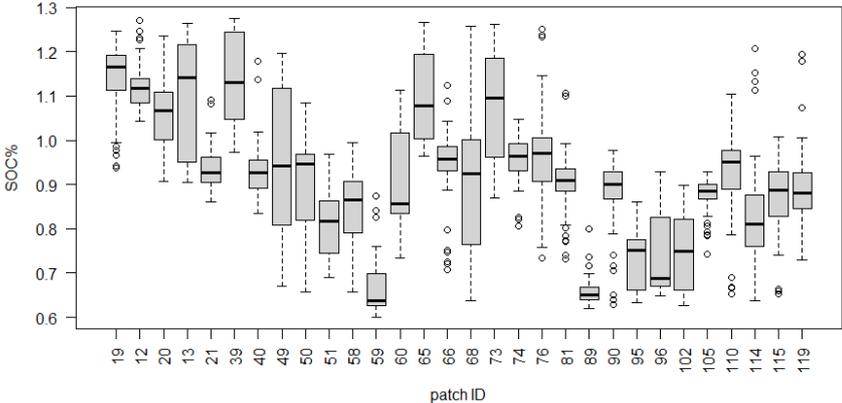


Figure 22: Boxplot of SOC patches from Sentinel-2A at the patch scale

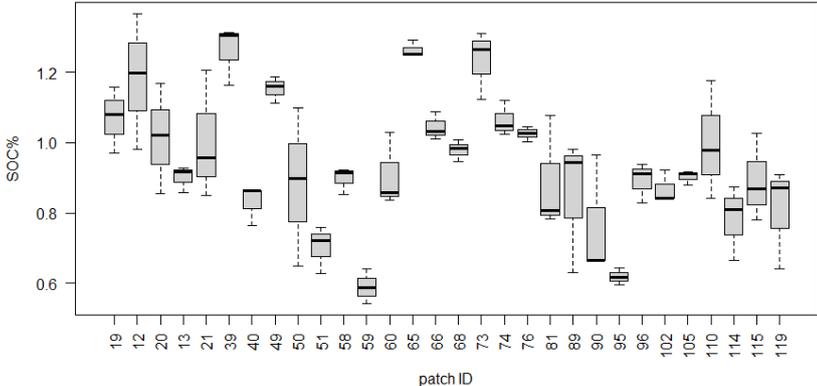


Figure 23: Boxplot of SOC patches from Soil sampling 3 at the patch scale

## **6. Comparison of proximal and remote sensed SOC estimations with ground truth measurements.**

According to Biney et al., (2021), when data are collected using various methods, sampling techniques, sample preparation prior to analysis, instrument requirements, analytical approaches, and algorithms, SOC prediction performance can be highly variable. In this study, comparison with the Sentinel-2A, the proximal sensing Veris MSP3 show the highest prediction as expected (RMSE = 0.11,  $R^2 = 0.68$ ) (Figure 24 and 25). Although the RMSE and  $R^2$  value for this study was not as high as reported elsewhere, it is comparable to Vaudour et al., (2016) studies.

In this study, the accuracy of SOC predictions using Sentinel-2A was the lowest compared with the Veris MSP3 data set. The low covariate importance of all Sentinel-2A bands and derived indices utilized with SOC may be one of the possible reasons of its model performance. The RF model can be influenced by the smaller number of training samples, date change and activities done in the field between the Sentinel-2A imagery and the soil data. The Veris MSP3 data were collected on the 10<sup>th</sup> and 11<sup>th</sup> of March 2020, Sentinel-2A data were from 01 April 2020, and soil sampling were from 16 November 2020. The time period between Veris MSP3 and Sentinel-2A data collection was to ensure bare soil for Sentinel-2A. According to Bartholomeus et al., (2011), existence of vegetation can sometimes affect the spectral reflectance, therefore soil properties prediction accuracy could be affected. During the time period of proximal sensing, the study area was covered with the cover crop and compost was applied and incorporated in the study area before crops were planted in the end of March 2020.

The relatively low image resolution of Sentinel-2A might also be a challenging factor for prediction. Even though, compared to other method Sentinel-2A data set can be desirable due to its wide spatial coverage and short revisiting time. Additionally, Berger et al., (2012) mentioned in his study the enormously frequent data streams produced by satellite sensors can ensure that soil monitoring and mapping methods for greater areas may be built properly, quickly, and efficiently.

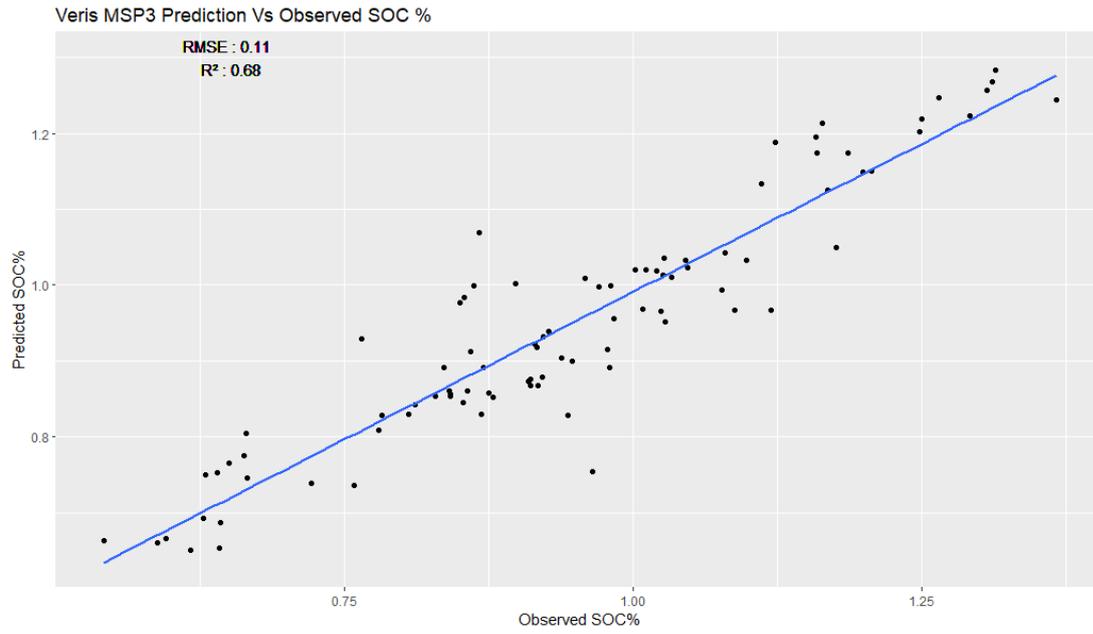


Figure 24: Spatial SOC distribution maps on study area based on prediction using Veris MSP3. Two statistical indicators (RMSE and R2 are shown)

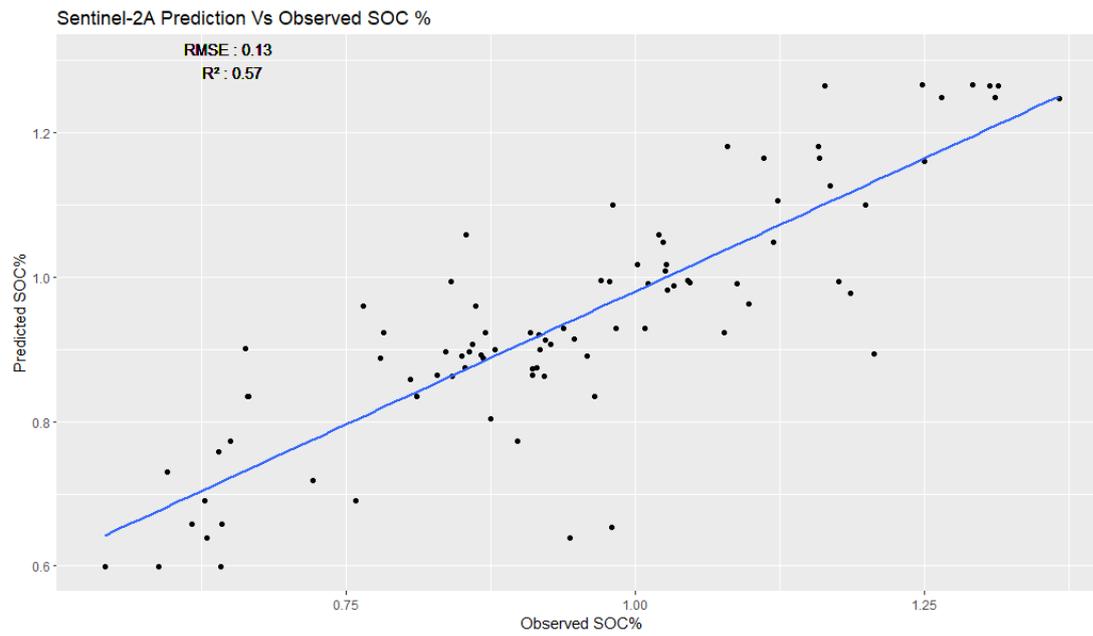


Figure 25: Scatter plots between the observed SOC values and predicted SOC values by RF model by using Sentinel-2A. Two statistical indicators (RMSE and R2 are shown)

#### **IV. CONCLUSION**

The objective of the thesis was to compare the proximal (Veris MSP3) and RS (UAV and Sentinel-2A) technologies to estimate SOC at the field scale. This study compared and explored the ability to predict SOC in a field with SOC content using Veris MSP3, UAV and Sentinel-2A with spectral indices. For the prediction accuracy of SOC, proximal sensed (Veris MSP3) data obtained better model fitting, compared to RS (Sentinel-2A) data. After preprocessing UAV datasets were excluded from the modelling due to the data errors. The study indicates that both proximal and remote sensing technologies have their advantages when it comes to comparing and contrasting the two forms of measurement. For instance, Sentinel-2A offers a larger spatial coverage, while Veris MSP3 has the advantage of a reduced distance between the sensor and the soil surface, which can contribute to a more comprehensive retrieval of soil spectra. Proximal sensing will not substitute the use of UAV or satellite imagery for larger scale assessments but will greatly contribute to local management at small to medium scales. However, the comparison of the actual performance was challenging due to the time difference in ground truth data collection and the field activities done including compost application in the study field which may also have affected the changes in estimated and analysed SOC content. The prediction accuracy was low with RMSE value 0.11 and 0.13,  $R^2$  value 0.68 and 0.57 for Veris MSP3 and Sentinel-2A, respectively. The poor model performance may relate to the availability of limited training data. This leads to the conclusion that successful SOC prediction with proximal and remotely sensed technologies requires more soil samplings for training and data collection should be done on the same time.

#### **Future Work**

The limitations to this research are probably the use of a limited number of soil samples and time of data collection. The limited number of soil sampling and time difference in data collection may affect the RF model accuracy. More research is required to explore the potential of proximal and RS technologies for estimating SOC in fields with low SOC content. This could involve using various spectral indices, using different ML algorithms, and analysing both high and low SOC content fields to better understand the actual differences.

## V. ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Luis Miranda and Dr. Kathrin Grahmann, mentors Prof. Dr. Herald Kaechele, Dr. Evelyn Wallor and Dr. Sebastian Vogel, for their invaluable guidance, support, and encouragement throughout my research. Their expertise, constructive criticism, and insightful suggestions were crucial to the successful completion of this thesis.

I would also like to thank patchCROP project team members at Leibniz-Centre for Agricultural Landscape Research (ZALF) e.V. University for providing the necessary resources and facilities that enabled me to carry out my research.

I extend my gratitude to the study participants who willingly dedicated their time and effort to take part in this research. Without their support, this research would not have been achievable.

I would like to convey my sincere gratefulness to my family and friends for their constant inspiration, love, and unwavering support throughout my academic journey.

Finally, I would like to acknowledge the contributions of my colleagues and classmates, who provided constructive feedback, valuable insights, and engaging discussions that helped to shape the direction of my research and enhance the quality of this thesis.

## VI. REFERENCES

1. Angelopoulou, T., Balafoutis, A., Zalidis, G., & Bochtis, D. (2020). From laboratory to proximal sensing spectroscopy for soil organic carbon estimation-A review. *Sustainability (Switzerland)*, *12*(2), 1–24. <https://doi.org/10.3390/su12020443>
2. Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., & Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing*, *11*(6), 1–18. <https://doi.org/10.3390/rs11060676>
3. Bartholomeus, H., Kooistra, L., Stevens, A., van Leeuwen, M., van Wesemael, B., Ben-Dor, E., & Tychon, B. (2011). Soil Organic Carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *International Journal of Applied Earth Observation and Geoinformation*, *13*(1), 81–88. <https://doi.org/10.1016/j.jag.2010.06.009>
4. Berger, M., Moreno, J., Johannessen, J. A., Levelt, P. F., & Hanssen, R. F. (2012). ESA's sentinel missions in support of Earth system science. *Remote Sensing of Environment*, *120*, 84–90. <https://doi.org/10.1016/j.rse.2011.07.023>
5. Biney, J. K. M., Saberioon, M., Borůvka, L., Houška, J., Vašát, R., Agyeman, P. C., Coblinski, J. A., & Klement, A. (2021). Exploring the suitability of uas-based multispectral images for estimating soil organic carbon: Comparison with proximal soil sensing and spaceborne imagery. *Remote Sensing*, *13*(2), 1–19. <https://doi.org/10.3390/rs13020308>
6. Chaudhari, P. R., Ahire, D. V., Ahire, V. D., Chkravarty, M., & Maity, S. (2013). Soil Bulk Density as related to Soil Texture, Organic Matter Content and available total Nutrients of Coimbatore Soil. *International Journal of Scientific and Research Publications*, *3*(1), 2250–3153. [www.ijsrp.org](http://www.ijsrp.org)
7. Christy, C. D. (2008). Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and Electronics in Agriculture*, *61*(1), 10–19. <https://doi.org/10.1016/j.compag.2007.02.010>
8. Cootes, T. F., Ionita, M. C., Lindner, C., & Sauer, P. (n.d.). *Cootes\_Eccv12*. 1–14.
9. Donat, M., Geistert, J., Grahmann, K., Bloch, R., & Bellingrath-Kimura, S. D. (2022). Patch cropping- a new methodological approach to determine new field arrangements that increase the multifunctionality of agricultural landscapes.

- Computers and Electronics in Agriculture*, 197(February), 106894.  
<https://doi.org/10.1016/j.compag.2022.106894>
10. Drake, J. M., Randin, C., & Guisan, A. (2006). Modelling ecological niches with support vector machines. *Journal of Applied Ecology*, 43(3), 424–432.  
<https://doi.org/10.1111/j.1365-2664.2006.01141.x>
  11. Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., & Scholten, T. (2020). Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran. *Remote Sensing*, 12(14).  
<https://doi.org/10.3390/rs12142234>
  12. Farooq, I., Bangroo, S. A., Bashir, O., Shah, T. I., Malik, A. A., Iqbal, A. M., Mahdi, S. S., Wani, O. A., Nazir, N., & Biswas, A. (2022). Comparison of Random Forest and Kriging Models for Soil Organic Carbon Mapping in the Himalayan Region of Kashmir. *Land*, 11(12), 1–15.  
<https://doi.org/10.3390/land11122180>
  13. Fleming, K. L., Heermann, D. F., & Westfall, D. G. (2000). *for Management Zone Delineation. 2002*, 1581–1587.
  14. Frazier, B. E., & Cheng, Y. (1989). Remote sensing of soils in the Eastern Palouse region with landsat thematic mapper. *Remote Sensing of Environment*, 28(C), 317–325. [https://doi.org/10.1016/0034-4257\(89\)90123-5](https://doi.org/10.1016/0034-4257(89)90123-5)
  15. G. Kweon, E. Lund, and C. M., & Veris. (2012). *International Conference on Precision Agriculture. 2012. Indianapolis IN. 71526*.
  16. Gambill, D. R., Wall, W. A., Fulton, A. J., & Howard, H. R. (2016). Predicting USCS soil classification from soil property variables using Random Forest. *Journal of Terramechanics*, 65, 85–92.  
<https://doi.org/10.1016/j.jterra.2016.03.006>
  17. Gholizadeh, A., Žižala, D., Saberioon, M., & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218(September), 89–103.  
<https://doi.org/10.1016/j.rse.2018.09.015>
  18. Godfray, H. C. J., Crute, I. R., Haddad, L., Muir, J. F., Nisbett, N., Lawrence, D., Pretty, J., Robinson, S., Toulmin, C., & Whiteley, R. (2010). The future of the global food system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1554), 2769–2777. <https://doi.org/10.1098/rstb.2010.0180>
  19. Guio Blanco, C. M., Brito Gomez, V. M., Crespo, P., & Ließ, M. (2018). Spatial

- prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. *Geoderma*, 316(July 2017), 100–114. <https://doi.org/10.1016/j.geoderma.2017.12.002>
20. Jin, X., Song, K., Du, J., Liu, H., & Wen, Z. (2017). Comparison of different satellite bands and vegetation indices for estimation of soil organic matter based on simulated spectral configuration. *Agricultural and Forest Meteorology*, 244–245(May), 57–71. <https://doi.org/10.1016/j.agrformet.2017.05.018>
  21. John, K., Isong, I. A., Kebonye, N. M., Ayito, E. O., Agyeman, P. C., & Afu, S. M. (2020). Using machine learning algorithms to estimate soil organic carbon variability with environmental variables and soil nutrient indicators in an alluvial soil. *Land*, 9(12), 1–20. <https://doi.org/10.3390/land9120487>
  22. Judge, J. (2007). Microwave Remote Sensing of Soil. *Advances*, 50(5), 1645–1649.
  23. Khanal, S., Kushal, K. C., Fulton, J. P., Shearer, S., & Ozkan, E. (2020). Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sensing*, 12(22), 1–29. <https://doi.org/10.3390/rs12223783>
  24. Kuang, B., Mahmood, H. S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M., & van Henten, E. J. (2012). Sensing soil properties in the laboratory, in situ, and on-line. A review. In *Advances in Agronomy* (1st ed., Vol. 114). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-394275-3.00003-1>
  25. Kweon, G., Lund, E., & Maxton, C. (2013). Soil organic matter and cation-exchange capacity sensing with on-the-go electrical conductivity and optical sensors. *Geoderma*, 199, 80–89. <https://doi.org/10.1016/j.geoderma.2012.11.001>
  26. Kweon, G., & Maxton, C. (2013). Soil organic matter sensing with an on-the-go optical sensor. *Biosystems Engineering*, 115(1), 66–81. <https://doi.org/10.1016/j.biosystemseng.2013.02.004>
  27. Ladoni, M., Bahrami, H. A., Alavipanah, S. K., & Norouzi, A. A. (2010). Estimating soil organic carbon from soil reflectance: A review. *Precision Agriculture*, 11(1), 82–99. <https://doi.org/10.1007/s11119-009-9123-3>
  28. Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, 352(January), 395–413. <https://doi.org/10.1016/j.geoderma.2019.05.031>
  29. Liu, D., Wang, Z., Zhang, B., Song, K., Li, X., Li, J., Li, F., & Duan, H. (2006).

- Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. *Agriculture, Ecosystems and Environment*, 113(1–4), 73–81. <https://doi.org/10.1016/j.agee.2005.09.006>
30. Martos, V., Ahmad, A., Cartujo, P., & Ordoñez, J. (2021). Ensuring agricultural sustainability through remote sensing in the era of agriculture 5.0. *Applied Sciences (Switzerland)*, 11(13). <https://doi.org/10.3390/app11135911>
  31. Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to topographic effects: A case study in high-density cypress forest. *Sensors*, 7(11), 2636–2651. <https://doi.org/10.3390/s7112636>
  32. Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping - A review. *Geoderma*, 162(1–2), 1–19. <https://doi.org/10.1016/j.geoderma.2010.12.018>
  33. Nelson, D. W., & Sommers, L. E. (2018). Total carbon, organic carbon, and organic matter. *Methods of Soil Analysis, Part 3: Chemical Methods*, 5, 961–1010. <https://doi.org/10.2136/sssabookser5.3.c34>
  34. Preza Fontes, G., Bhattarai, R., Christianson, L. E., & Pittelkow, C. M. (2019). Combining Environmental Monitoring and Remote Sensing Technologies to Evaluate Cropping System Nitrogen Dynamics at the Field-Scale. *Frontiers in Sustainable Food Systems*, 3(February), 1–13. <https://doi.org/10.3389/fsufs.2019.00008>
  35. Rochette, P., & Bertrand, N. (2007). Soil-surface gas emissions. In *Soil Sampling and Methods of Analysis: Second Edition*. <https://doi.org/10.1201/9781420005271.ch65>
  36. Rossel, R. A. V., & Adamchuk, V. I. (2013). Proximal soil sensing. *Precision Agriculture for Sustainability and Environmental Protection*, 9780203128(November), 99–118. <https://doi.org/10.4324/9780203128329>
  37. Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A. B., McKenzie, N. J., Mendonca Santos, M. L., Minasny, B., Montanarella, L., Okoth, P., Palm, C. A., Sachs, J. D., Shepherd, K. D., Vågen, T.-G., Vanlauwe, B., Walsh, M. G., ... Zhang, G.-L. (2009). Environmental science. Digital soil map of the world. *Science (New York, N.Y.)*, 325(5941), 680–681. <https://doi.org/10.1126/science.1175084>
  38. Segal, M. R. (2004). Machine Learning Benchmarks and Random Forest

- Regression Publication Date Machine Learning Benchmarks and Random Forest Regression. *Center for Bioinformatics and Molecular Biostatistics*, 15. <https://escholarship.org/uc/item/35x3v9t4>
39. Shekhar, S., & Xiong, H. (2008). Esri. *Encyclopedia of GIS*, 290–290. [https://doi.org/10.1007/978-0-387-35973-1\\_369](https://doi.org/10.1007/978-0-387-35973-1_369)
  40. Sothe, C., Gonsamo, A., Arabian, J., & Snider, J. (2022). Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. *Geoderma*, 405, 115402. <https://doi.org/10.1016/j.geoderma.2021.115402>
  41. Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., & Ben-Dor, E. (2008). Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. *Geoderma*, 144(1–2), 395–404. <https://doi.org/10.1016/j.geoderma.2007.12.009>
  42. Vaudour, E., Gilliot, J. M., Bel, L., Lefevre, J., & Chehdi, K. (2016). Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. *International Journal of Applied Earth Observation and Geoinformation*, 49, 24–38. <https://doi.org/10.1016/j.jag.2016.01.005>
  43. Villa, T., Gonzalez, F., Miljevic, B., Ristovski, Z. D., & Morawska, L. (2016). An overview of small unmanned aerial vehicles for air quality measurements: Present applications and future perspectives. *Sensors (Switzerland)*, 16(7), 12–20. <https://doi.org/10.3390/s16071072>
  44. Vogel, S., Bönecke, E., Kling, C., Kramer, E., Lück, K., Philipp, G., Rühlmann, J., Schröter, I., & Gebbers, R. (2021). Direct prediction of site-specific lime requirement of arable fields using the base neutralizing capacity and a multi-sensor platform for on-the-go soil mapping. *Precision Agriculture*, 0123456789. <https://doi.org/10.1007/s11119-021-09830-x>
  45. Waiser, T. H., Morgan, C. L. S., Brown, D. J., & Hallmark, C. T. (2007). In Situ Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy. *Soil Science Society of America Journal*, 71(2), 389–396. <https://doi.org/10.2136/sssaj2006.0211>
  46. Wang, X., Han, J., Wang, X., Yao, H., & Zhang, L. (2021). Estimating Soil Organic Matter Content Using Sentinel-2 Imagery by Machine Learning in Shanghai. *IEEE Access*, 9, 78215–78225.

<https://doi.org/10.1109/ACCESS.2021.3080689>

47. Whitehead, K., Hugenholtz, C. H., Myshak, S., Brown, O., Leclair, A., Tamminga, A., Barchyn, T. E., Moorman, B., & Eaton, B. (2014). Remote sensing of the environment with small unmanned aircraft systems (Uass), part 2: Scientific and commercial applications. *Journal of Unmanned Vehicle Systems*, 2(3), 86–102. <https://doi.org/10.1139/juvs-2014-0007>

## VII. APPENDICES

### Repository Link

[https://github.com/akhil01216/Master\\_Thesis.git](https://github.com/akhil01216/Master_Thesis.git). This is the link to the repository to get the codes used in this master thesis.

### Historical images of the study area from Google Earth



Figure 26: Historical images from Google Earth

### List of Figures

Figure 1: Location of the study area .....	10
Figure 2: Veris MSP3 Unit .....	12
Figure 3: Methodological flowchart for modelling and mapping SOC .....	21
Figure 4: Soil spectra measured from different platform .....	22
Figure 5: Timeline of data .....	22
Figure 6: Soil sampling points .....	23
Figure 7: Veris MSP3 data collected on 10 <sup>th</sup> &11 <sup>th</sup> March 2020 .....	24
Figure 8: RGB imagery from UAV on 31 March 2020.....	25
Figure 9: RGB imagery from Sentinel-2A on 01 April 2020.....	25

Figure 10: Soil sampling points of sampling campaign conducted on 16 November 2020, (Background imagery is from Sentinel 2A on 04 November 2020.) .....	27
Figure 11: Variables measured by Veris MSP3 in raster form .....	28
Figure 12: Spectral Indices from UAV .....	29
Figure 13: Spectral indices from Sentinel-2A.....	29
Figure 14: Density histogram, boxplot and statistics summary of SOC content (n=90) 30	
Figure 15: Covariate importance (ranking of predictors) from RF model fitting Veris MSP3 (%IncMSE: Mean Decrease Accuracy) .....	31
Figure 16: Covariate importance (ranking of predictors) from RF model fitting Sentinel-2A .....	31
Figure 17: Spatial SOC distribution maps on study area based on prediction using Veris MSP3 dataset.....	32
Figure 18: Spatial SOC distribution maps on study area based on prediction using Sentinel-2A dataset .....	32
Figure 19: Spatial SOC distribution maps on patchCROP study area based on prediction using Veris MSP3.....	33
Figure 20: Spatial SOC distribution maps on patchCROP study area based on prediction using Sentinel-2A.....	33
Figure 21: Boxplot of SOC patches from Veris MSP3 at the patch scale .....	34
Figure 22: Boxplot of SOC patches from Sentinel-2A the patch scale.....	34
Figure 23: Boxplot of SOC patches from Soil sampling 3 at the patch scale .....	34
Figure 24: Spatial SOC distribution maps on study area based on prediction using Veris MSP3. Two statistical indicators (RMSE and R2 are shown) .....	36
Figure 25: Scatter plots between the observed SOC values and predicted SOC values by RF model by using Sentinel-2A. Two statistical indicators (RMSE and R2 are shown)36	
Figure 26: Historical images from Google Earth .....	45

## List of Tables

Table 1: Summary of Sentinel-2A spectral bands .....	16
Table 2: Spectral indices equation .....	18
Table 3: Soil sampling detailed overview (72(25): The total collected 72 and 270 samples mixed together and final samples analysed in the laboratory is 25 and 90 respectively). .....	23
Table 4: Descriptive statistics of Veris MSP3 variables (IR-Infrared, pH-potential of hydrogen, EC-Electrical conductivity) .....	24

## List of Abbreviations

ANN	Artificial Neural Network
AOI	Area of Interest
BI	Brightness Index
CI	Color Index
CV	Coefficient of Variation
DVI	Difference Vegetation Index
EC	Electric Conductivity
ESA	European Space Agency
GLASOD	Global Assessment of Soil Degradation
GNDVI	Green Normalized Difference Vegetation Index
GNSS	Global navigation satellite system
GPS	Global Positioning System
IPVI	Infrared Percentage Vegetation Index
ISEs	Ion Selective Electrode
ISRIC	International Soil Reference and Information Centre
LED	Light Emitting Diode
MIR	Mid-Infrared
ML	Machine Learning

MLR	Multiple Linear Regression
MSP3	Multi Sensor Platform 3
NDRE	Normalized Difference Red Edge
NDVI	Normalized Difference Vegetation Index
$R^2$	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
RS	Remote Sensing
SAVI	Soil Adjusted Vegetation Index
SD	Standard Deviation
SOC	Soil Organic Carbon
SVM	Support Vector Machine
SWIR	Short-Wave Infrared
UAV	Unmanned Aerial Vehicle
UNEP	United Nations Environment Programme
USGS	United States Geological Survey
VI	Vegetation Index
VNIR	Visible and Near-Infrared
ZALF	Leibniz-Centre for Agricultural Landscape Research

Wyrażam zgodę na udostępnienie mojej pracy w czytelniach Biblioteki SGGW w tym w Archiwum Prac Dyplomowych SGGW

I agree to share my work in the reading rooms of the SGGW Library, including the SGGW Theses Archive.

Ich erteile meine Zustimmung zur Veröffentlichung meiner Arbeit in der Bibliothek der SGGW (Warschauer Naturwissenschaftliche Universität), einschließlich des Archivs der Diplomarbeiten.

.....  
*(czytelny podpis autora pracy)*

*(legible signature of the author)*

*(lesbare Unterschrift des Autors der Arbeit)*

