Head Office: Università degli Studi di Padova

Department

DAFNAE - Department of Agronomy, Food, Natural resources, Animals and Environment

PhD COURSE CROP SCIENCE

SERIES 34° Cycle

DEVELOPING A DECISION SUPPORT SYSTEM FOR SUSTAINABLE

CULTIVATION OF DURUM WHEAT CV. "SENATORE CAPPELLI"

THROUGH PROXIMAL AND REMOTE SENSING

Coordinator: Prof. Claudio Bonghi

Supervisor: Prof. Michele Pisante

Co-Supervisor: Prof. Antonio Berti

PhD Student: Matteo Petito

Matricola: 1193819

ABSTRACT

The Durum wheat (*Triticum durum* Desf.) *cv*. "Senatore Cappelli" is a cultivar selected by Nazareno Strampelli in 1915. The *cv*. "Senatore Cappelli" is characterized by very tall plants, low production, and high grain protein content. In the last years, it has met with considerable interest from the market for its grain quality and nutraceutical properties. Future European agricultural production will need to be at the same time more productive and more sustainable. Site-specific management (SSM) can promote a new sustainable European agriculture. SSM, also known as Precision Agriculture (PA), is an agronomic strategy that manages within-field variability. Several methods are often used to measure spatial field variability, but these have several limitations in large-scale applications.

The main goal of this Thesis was to develop a Decision Support System (DSS) for the sustainable cultivation of Durum wheat *cv*. "Senatore Cappelli".

This DSS can be divided into two modules: a) Digital Soil Mapping Module, which uses rapid methods to assess field variability and map the same soil properties. b) Crop Suitability module where the information obtained from the first module are used to explain yield variability.

Alternative methods at the apparent soil electrical conductivity (ECa) were evaluated in the DSM module. Geospatial ECa mapping is a sensor technology that has a significant role in spatial field characterization. However, this method has several limitations; therefore, the rapid expansion of Precision Agriculture required an alternative, faster way to map spatial field variability.

Multi-temporal images with bare soil pixels (SYSI) can be a valid alternative to ECa, understood as a proxy of soil variability. In particular, the red band (B4) of SYSI and the bare soil index SOCI (Soil Organic Carbon Index) can be used as a rapid and economically proxy of soil variability.

Moreover, in the DSM module, rapid methods for Soil Organic Carbon mapping were assessed.

The study and management of the Soil Organic Carbon variability are essential for a new sustainable agriculture. The rapid expansion of Precision Agriculture requires rapid and inexpensive methods for SOC mapping. Bare soil indices provided by SYSI can explain SOC variability and could be used as auxiliary information in sampling optimization. Moreover, SYSI allowed a reduction of the sampling scheme by about 75% with a consequent considerable efficiency in terms of time and costs..

Finally, a preliminary suitability analysis of *cv*. "Senatore Cappelli" was carried out in the Crop Suitability module, using topographical features, soil data, and bare soil images by SYSI.

Soil and topography properties explain only 13% of with-in grain yield variability, and SYSI alone produced accurate predictions and confirmed that it is a good way to quickly obtain information on soil variability and suitability for the cultivation of a specific crop or cultivar.

This method could be a vital impulse for Precision Agriculture adoption and diffusion. Therefore, this result could open the way to new accurate and cheaper services based on data and informed decisions.

Index

INTRODUCTION

The Durum wheat (*Triticum durum* Desf.) *cv.* "Senatore Cappelli" is a cultivar selected by Nazareno Strampelli. The cultivar "Senatore Cappelli", is a pure line belonging to the Mediterranean type extracted from the North African population Jean Retifah in 1915. The *cv.* "Senatore Cappelli" is characterized by very tall plants, low production, and high grain protein content (Dinelli et al. 2013; Giunta et al. 2019; Pagnani et al. 2020). Until the advent of the Green Revolution in late 1960, *cv.* "Senatore Cappelli" was the most important variety cultivated in the Mediterranean area (Mefleh et al. 2018). In the last years, *cv.* "Senatore Cappelli" has met with considerable interest from the market for its grain quality (Acquistucci et al. 2020; Bianco et al. 2019; Mascia et al. 2014; Mefleh et al. 2018; Rocco et al. 2019) and nutraceutical properties (Ianiro et al. 2019).

Future European agricultural production will need to be at the same time more productive and more sustainable (Cillis et al. 2017; Farooq et al. 2019; Pisante et al. 2012; Pretty 1997). Site-specific management (SSM) can promote a new sustainable European agriculture. SSM, also known as Precision Agriculture (PA), is an agronomic strategy that manages within-field variability (Sudduth et al. 2015). SSM focuses on the influence of spatial variability in field varibility on crop yield (Plant 2001). The premises behind the site-specific farming are (Plant 2001): a) significant within-field spatial variability exists in factors that influence crop yield, b) causes of this variability can be identified and measured, c) the information from these measurements can modify crop management to increase economic and environmental sustainability.

Therefore, SSM may be able to optimize production, but this is related to the accuracy of the assessment of field variability. Several methods are often used to measure spatial field variability, but these have several limitations in large-scale applications. The rapid expansion of PA required an alternative, faster way to map spatial field variability.

The main goal of this Thesis was to develop a Decision Support System (DSS) for the sustainable cultivation of Durum wheat *cv.* "Senatore Cappelli". A DSS is a tool that helps the user to make decisions. When a decision is agronomic, like when related to site-specific management of a crop,

the definition of the decision space may be much more complicated. The nature of the decision itself may depend on the availability of data. Therefore, the basic idea of this project was to create tools capable of intercepting field variabilty quickly and economically.

This DSS can be divided into two modules: a) Digital Soil Mapping Module, which uses rapid methods to assess field variability and map the same soil properties. b) Crop Suitability module where the information obtained from the first module are used to explain yield variability.

The main objectives of this Thesis are:

1) Development of rapid tools to map within-field spatial soil variability through remotely sensed data.

2) Development of rapid tools to map Soil Organic Carbon variability and sampling optimization.

3) Evaluation and estimation of the Effect of soil properties and topographic properties on yield variability.
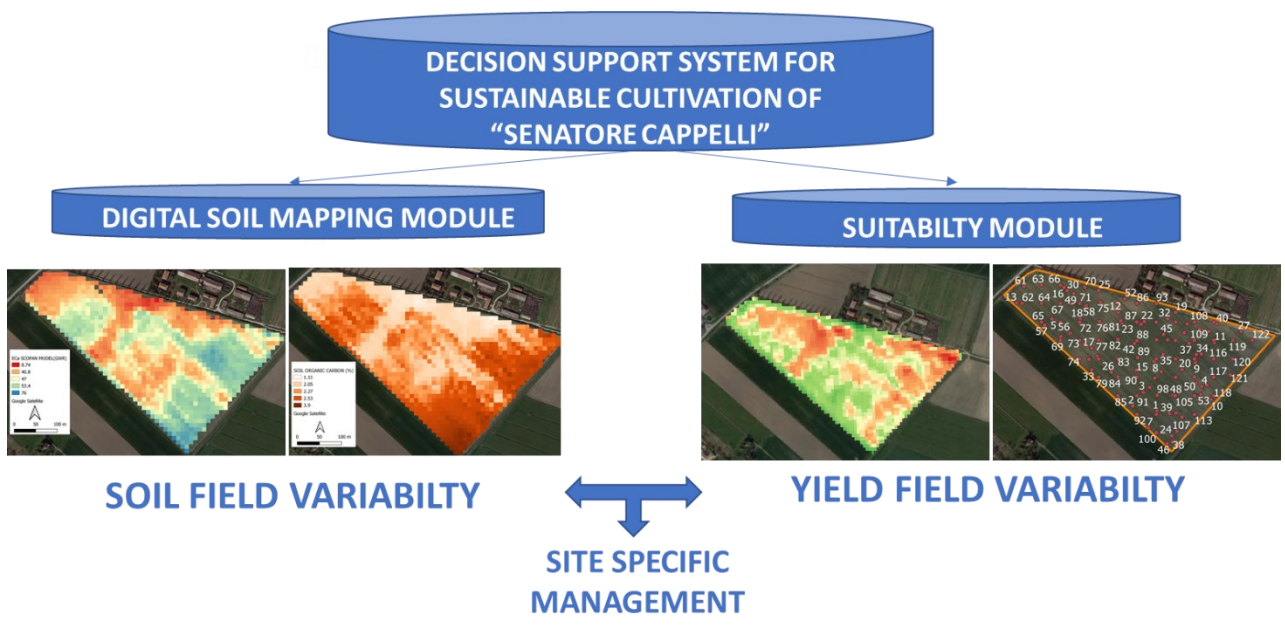
Figure 1. Decision Support System for sustainable cultivation of Durum wheat *cv.* "Senatore Cappelli".

The Thesis is structured in three chapters:

In Chapter I, different methods of measuring spatial variability are examined, in particular, a prediction model for Soil Electrical Conductivity using remote sensing data was developed. Finally, the multitemporal bare soil image (SYSI) Demattê et al. (2018) was evaluated to intercept and spatialize within-field variability.

In Chapter II: rapid methods to Soil Organic Carbon (SOC) mapping and their limits, have been examined. In particular, a model for Soil organic prediction using multitemporal bare soil image was developed. Then a tool for soil sampling optimization and definition of the minimal number of sample points was created.

In Chapter III evaluated the interaction between yield and topographic features and soil properties. This PhD project provides a basis for the site-specific management of Durum wheat *cv.* "Senatore Cappelli".

Finally, conclusion and final remarks presents the main traits arising from this Thesis, Chapter IV.

References

Dinelli, G.; Marotti, I.; di Silvestro, R.; Bosi, S.; Bregola, V.; Accorsi, M.; di Loreto, A.; Benedettelli, S.; Ghiselli, L.; Catizone, P. Agronomic, Nutritional and Nutraceutical Aspects of Durum Wheat *(Triticum Durum Desf.) Cultivars under Low Input Agricultural Management. Italian Journal of Agronomy* **2013**, *8*, 12, doi:10.4081/ija.2013.e12.

Giunta, F.; Pruneddu, G.; Motzo, R. Grain Yield and Grain Protein of Old and Modern Durum Wheat Cultivars Grown under Different Cropping Systems. *Field Crops Research* **2019**, *230*, 107–120, doi:10.1016/J.FCR.2018.10.012.

Pagnani, G.; Galieni, A.; Stagnari, F.; Pellegrini, M.; del Gallo, M.; Pisante, M. Open Field Inoculation with PGPR as a Strategy to Manage Fertilization of Ancient Triticum Genotypes. *Biology and Fertility of Soils* **2020**, *56*, 111–124, doi:10.1007/s00374-019-01407-1.

Mefleh, M.; Conte, P.; Fadda, C.; Giunta, F.; Piga, A.; Hassoun, G.; Motzo, R. From Ancient to Old and Modern Durum Wheat Varieties: Interaction among Cultivar Traits, Management, and Technological Quality. **2018**, doi:10.1002/jsfa.9388.

Rocco, M.; Tartaglia, M.; Izzo, F.P.; Varricchio, E.; Arena, S.; Scaloni, A.; Marra, M. Comparative Proteomic Analysis of Durum Wheat Shoots from Modern and Ancient Cultivars. *Plant Physiology and Biochemistry* **2019**, *135*, 253–262, doi:10.1016/j.plaphy.2018.12.010.

Acquistucci, R.; Melini, V.; Galli, V. Durum Wheat Grain and Pasta from Locally-Grown Crops: A Case-Study on Saragolla (Triticum Turgidum Ssp. Turanicum) and Senatore Cappelli (Triticum Turgidum Ssp. Durum) Wheats. *Emirates Journal of Food and Agriculture* **2020**, 47, doi:10.9755/ejfa.2020.v32.i1.2058.

Mascia, I.; Fadda, C.; Dostálek, P.; Olšovská, J.; Caro, A. del Preliminary Characterization of an Italian Craft Durum Wheat Beer. **2014**, doi:10.1002/jib.176.

Bianco, A.; Fancello, F.; Balmas, V.; Dettori, M.; Motroni, A.; Zara, G.; Budroni, M. Microbial Communities and Malt Quality of Durum Wheat Used in Brewing. **2019**, doi:10.1002/jib.555.

Ianiro, G.; Rizzatti, G.; Napoli, M.; Matteo, M.; Rinninella, E.; Mora, V.; Fanali, C.; Leonetti, A.; Benedettelli, S.; Mele, M.; et al. A Durum Wheat Variety-Based Product Is Effective in Reducing Symptoms in Patients with Non-Celiac Gluten Sensitivity: A Double-Blind Randomized Cross-Over Trial. *Nutrients* **2019**, *11*, doi:10.3390/nu11040712.

Pretty, J.N. The Sustainable Intensification of Agriculture. *Natural Resources Forum* **1997**, *21*, 247–256, doi:10.1111/j.1477-8947.1997.tb00699.x.

Farooq, M.; Rehman, A.; Pisante, M. Sustainable Agriculture and Food Security. In *Innovations in Sustainable Agriculture*; Springer International Publishing: Cham, 2019; pp. 3–24.

Pisante, M.; Stagnari, F.; Grant, C.A. Agricultural Innovations for Sustainable Crop Production Intensification. *Italian Journal of Agronomy* **2012**, *7*, 40, doi:10.4081/ija.2012.e40.

Cillis, D.; Pezzuolo, A.; Marinello, F.; Basso, B.; Colonna, N.; Furlan, L.; Sartori, L. Conservative Precision Agriculture: An Assessment of Technical Feasibility and Energy Efficiency within the LIFE+ AGRICARE Project. *Advances in Animal Biosciences* **2017**, *8*, 439–443, doi:10.1017/S204047001700019X.

Sudduth, K.A.; Hummel, J.W.; Birrell, S.J. Sensors for Site-Specific Management. In; 2015; pp. 183–210.

Plant, R.E. Site-Specific Management: The Application of Information Technology to Crop Production. *Computers and Electronics in Agriculture* **2001**, *30*, 9–29, doi:10.1016/S0168-1699(00)00152-6.

Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A Powerful Data Mining Procedure to Retrieve Soil Spectral Reflectance from Satellite Images. *Remote Sensing of Environment* **2018**, *212*, 161–175, doi:10.1016/j.rse.2018.04.047.

CHAPTER I: Rapid tools to map within-field spatial soil variability through remotely sensed data

1.   Introduction

Future European agricultural production will need to be at the same time more productive and more sustainable (Cillis et al. 2017; Farooq et al. 2019; Pisante et al. 2012; Pretty 1997; Sudduth et al. 2015)]. Sitespecific management (SSM) can promote a new sustainable European agriculture. SSM, also known as Precision Agriculture (PA), is an agronomic strategy that manages within-field variability (Sudduth et al. 2015). SSM focuses on the influence of spatial variability in field properties on crop yield (Plant 2001). The premises behind the site-specific farming are (Plant 2001): a) significant within-field spatial variability exists in factors that influence crop yield, b) causes of this variability can be identified and measured, c) the information from these measurements can modify crop management to increase eco-nomic and environmental sustainability. Therefore, SSM may be able to optimize production, but this is related to the accuracy of the assessment of field variability. Several methods measure spatial field variability, the best-known are Yield Monitoring, Topographic factor assessment, and apparent soil Electrical Conductivity (ECa) mapping.

Yield maps play an essential role in SSM. The spatial distribution of yield depends by the interaction between the soil characteristics, position in the landscape, weather, and man-agement. Historical maps from at least four growing seasons allow understanding the spatial distribution of yield at field scale. The fields can be divided into stable and unstable portions based on the yield temporal variability (Blackmore 2000). In the stable portions of a field, where yield has not changed from year to year, yields are related to soil properties and landscape position (Maestrini and Basso 2018). Whereas yield is unstable in the area from year to year, yields strictly depend on weather (Maestrini and Basso 2018). Therefore, the stable zone may be managed using a site-specific management strategy, indicating where varying cropping inputs are needed based upon spatial patterns of crop productivity.

Crop growth and yield are also affected by topography. Field topography changed soil water availability and influenced soil particle redistribution organic matter and soil nutrients(Jiang and Thelen n.d.; Kravchenko and Bullock n.d.; Mulder et al. 2011). Topography has been reported to be related to the yield of many crops, therefore Digital elevation model (DEM) is used to characterize sites and determine soil properties such as soil organic matter and soil moisture (Kühn et al. 2009).

The measurement of ECa is a tool for spatial mapping variation of the soil's physico-chemical properties (Corwin and Lesch 2005a). Apparent ECa is a measure of the bulk conductivity of the soil. It can be measured using spatial-geophysical techniques such as electrical resistivity (ER) and electromagnetic induction (EMI). ECa is related to clay content, organic matter (OM), cation exchange capacity (CEC), and water content in non-salty soils. Geospatial ECa mapping is a sensor technology that has a significant role in spatial field characterization (Corwin and Scudiero 2020). Several studies have used ECa survey to define spatially some soil properties such as soil texture, salinity, nutrients, organic matter, and others soil properties. (Barca et al. 2019; Buttafuoco et al. 2017; Castrignanò et al. 2012; Corwin and Lesch 2005b; Sudduth et al. n.d.). Moreover, geo-referenced ECa measurements have been correlated with yield maps with mixed results (Corwin and Lesch 2005b), stressing the importance of the adoption of proper measurements protocols. When properly implemented, ECa mapping have been successfully used to map soil variability in precision farming due to its rapid and cheap(Barca et al. 2015; Castrignanò et al. 2018; Corwin and Lesch 2005b; Scudiero et al. 2013; Terrón et al. 2011). ECa is a complex measurement that requires interpretation and expertise. Moreover, ECa survey required specific mobile measurement equipment(Corwin and Lesch 2005a). Usually, expert operators can map an 18-ha field in one day 8-h workday at a 10-m spacing. Moreover, ECa survey can be affected by secondary influences, which may cause biased measurement. These secondary influences are positional offset effects, field edge effects, metal, surface topography end elevation, surface roughness, irrigation management, compaction due to farm vehicles, and ambient temperature on instrumentation (Corwin and Scudiero 2020). However, the coulter-electrodes of ER need to pene-trate the soil to a few centimeters, and this can be a disadvantage in stony soils, typical of Mediterranean regions (Terrón et al. 2011). Furthermore, survey by ER should be done either prior to tillage or when the field has been uniformly tilled (Terrón et al. 2011). Therefore, the geospatial measurement of ECa is a good surrogate of soil spatial variability, but the rapid expansion of PA required an alternative, faster way to map spatial field variability, and Satellite bare soil images can be a valid alternative (Weiss et al. 2020).

Satellite bare soil images are significant among the tools available for soil properties assessment. They allow indirect information collection on soil properties(Castaldi et al. 2016, 2019; Poggio and Gimona 2017)economically and over time. An image obtained from a single period will not have bare soil for the entire area. Studying field variability at a large scale requires tools capable of furnishing bare soil images in different study areas.

In recent studies have been developed new methodologies to obtain bare soil images, such as the Geospatial Soil Sensing System (GEOS3) Demattê et al. (2018), born on Landsat satellite and subsequently developed using Sentinel 2(Silvero et al. 2021). GEOS3 allows creating a single

synthetic bare soil image (Synthetic Soil Image, SYSI) using multi-temporal images. Several studies have used SYSI to map topsoil properties(Fongaro et al. 2018; Mendes et al. 2019; Safanelli et al. 2020a).

This study aims a) to predict Electrical Conductivity as a proxy of spatial variation of soil properties at field scale using SYSI and other environmental variables. b) Assess the multitemporal bare soil image (SYSI) potential to intercept and spatialize within-field variability.

2.   Materials and Methods

2.1.   Study area

The study area is located in the deltaic plain of Po River, near the town of Massa Fiscaglia (44° 79' N; 11°98' E), in Northern Italy. The study site area is composed by four fields (ID 3, ID 5, ID 16, ID 19) as reported in Figure 1, with a total surface of approximately 40 ha. Area and elevation, and slope for each field are reported in Table 1.

The climate of this area is temperate. According to ERA 5 LAND (Hersbach et al. 2020) in the last 40 years (1981- 2021), the yearly average precipitation is 770 mm; yearly average air temperature is 14.61 °C, yearly average minimum air temperature 0.43 °C, and yearly average maximum air temperature 28.79 °C.

ID Field	Hectares	Elevation (m a.s.l.)	Slope (%)
3	13.01	-0.99(mean)	1.11 (mean)
		-1.94 (min)	0.03 (min)
		-0.26 (max)	9.03 (max)
		0.37 (sd)	1.19 (sd)
5	6.39	-1.91(mean)	0.72 (mean)
		-2.43 (min)	0.05(min)
		-1.24 (max)	2.76(max)
		0.20(sd)	0.37(sd)
16	8.10	-1.08(mean)	1.06 (mean)
		-1.78 (min)	0.03 (min)
		-0.43 (max)	4.28 (max)
		0.29 (sd)	0.71 (sd)
19	12.05	-0.80 (mean)	0.84 (mean)
		-2.11 (min)	0.04 (min)
		1.28 (max)	4.32 (max)
		0.87 (sd)	0.44 (sd)

Table 1. Description of the study area.

Figure 2. Study area.

The ECa survey (Fig. 2) was conducted in September 2019 with Veris I scan soil electrical conductivity sensor (Veris Technologies Inc., Salina, KS, USA). The Veris cart was attached to a tractor and was pulled over the fields.

It is known that smaller soil particles (clay size particles) conduct more current than larger ones (silty or sand). Therefore, one pair of colter-electrodes measured the voltage drop. The Veris I scan generates two sets of data: topsoil data (EC_{30}), weighted depth (0-30 cm), and deep data from 0 to 90 cm (EC_{90}). In this study, only (EC_{30}) was considered.

A Soil sampling was carried to determine the predominant factor influencing spatial ECa, in field 19, in August 2020. A grid of 122 points, 61 according to a regular grid and 61 chosen randomly (Fig.3), was used to take the sample from topsoil (0-30 cm).

The coordinates of each sampling site were determined with a GPS model SP20 handheld GNSS (Spectra Geospatial) and a private laboratory carried out the soil analysis.

The samples were air-dried and analyzed using the Robinson pipette for particle size distribution. Organic Carbon was determined by dichromate oxidation,(Walkley and Black 1934) .
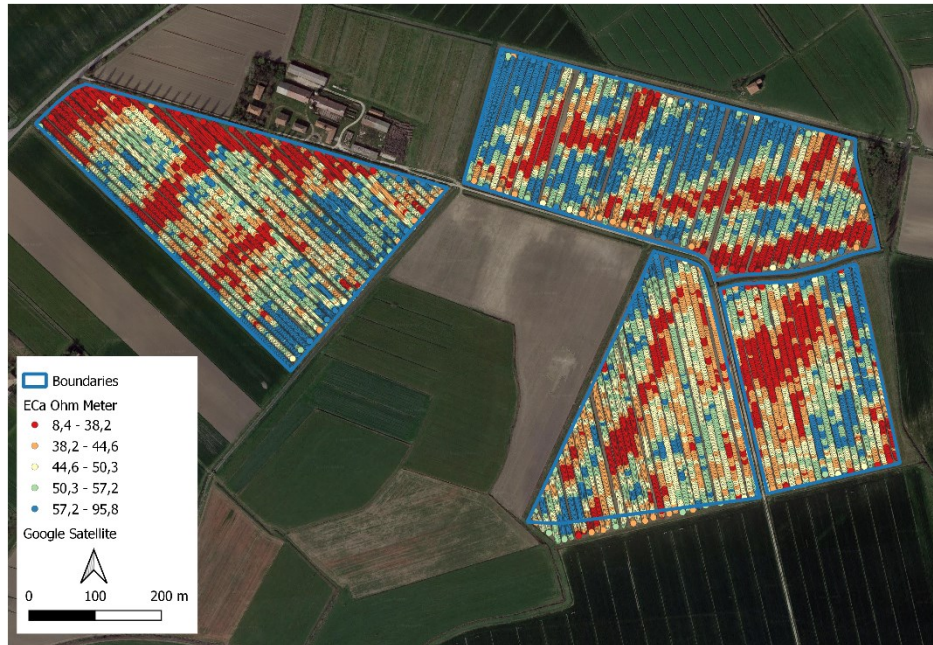
Figure 3. ECa survey.
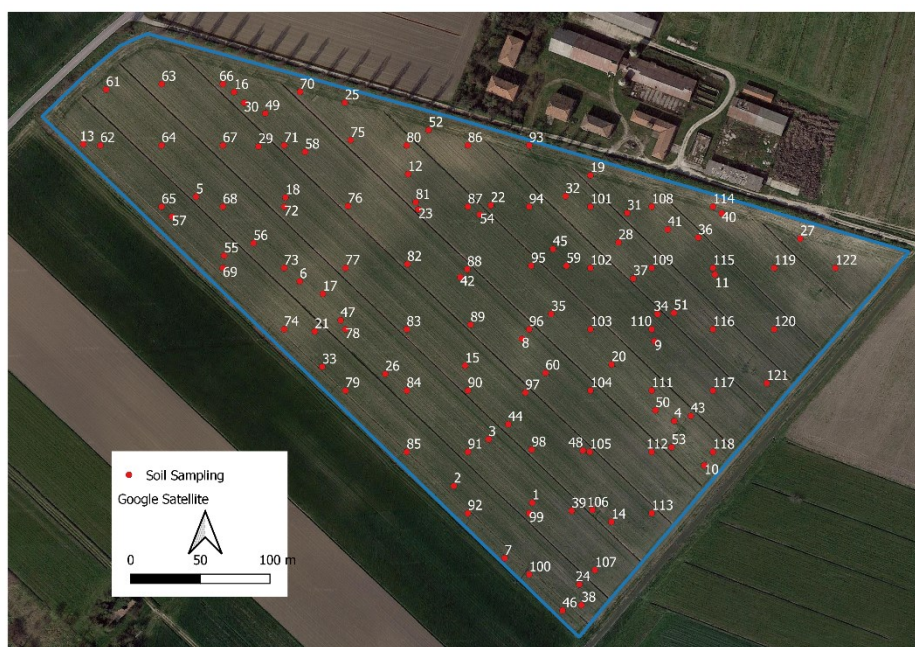
Figure 4. Soil sampling grid.

2.2. SCORPAN model

With the aim of predicting the ECa through remote sensing data, in field 19, the SCORPAN model (McBratney et al. 2003) was used. It is widely used in DSM. This conceptual model provides a framework for quantitative mapping of soil.

SCORPAN model, based on the seven predictive factors as a generalization of Jenny's five factors (Jenny 1941)

$$Sa=f (s, c, o, r, p, a, n)$$

Where Sa is soil attribute, s: soil properties; c: climate; o: vegetation; r: relief; p: lithology; a: age; n: spatial position; f: math function, only s, o, r, p, n factors were used in this work.

2.2.1. Soil properties (s) and lithology (p)

A multi-temporal bare soil image named SYSI was obtained by applying the GEOS3 using Sentinel-2 satellite data(Silvero et al. 2021) on the Google Earth Engine platform (GEE) Gorelick et al. (2017). One hundred fourteen images from 2018-2021 with at least 80% of the area without clouds and shadows were used. The GEOS3 method consists of (1) obtaining the images, (2) masking pixels that were non-bare soil, and (3) calculating the median reflectance of the masked images.

After obtaining Sentinel-2 images, NDVI (Normalized Difference Vegetation Index) and NBR2 (Normalized Burn Ratio 2) were calculated and used to mask the images.

Following the GEOS method, the range between 0 and 0.25 for NDVI was considered bare soil, and the values above this threshold were flagged as N.A., NBR2 values > 0.075 were also flagged as N.A. and corresponded to sites covered with straw or burned. For each image, quality masked was used to exclude areas with clouds and shadows. Finally, the masked images were ordered by date and the median reflectance was calculated. Form SYSI VIS-NIR bands (Table 2) were extracted (fig.4) with a spatial pixel resolution of 10 meters, and then several bare soil indices were calculated as reported in Table 3.

Sentinel -2 Bands	Central wavelength (nm)
Band 2 (B2) - Blue	490
Band 3 (B3) - Green	560
Band 4 (B4) - Red	665
Band 8 (B8) - NIR	842

Table 1. Sentinel -2 Band characteristics.

Index	Formula	Soil properties (reference)
Brightness Index (B.I.)	$(B4^2 + B3^2 + B2^2/3)^{0.5}$	Average reflectance (Mathieu et al. 1998a)
Saturation Index (SI)	(B4-B2)/(B4+B2)	Magnitude (Mathieu et al. 1998a) Spectra slope (Mathieu et al. 1998a)
Hue Index (HI)	(2 B4-B3-B2)/(B3-B2)	Primary colours (Mathieu et al. 1998a)
Coloration Index (CI)	(B4-B3)/(B4+B3)	Soil Colour (Mathieu et al. 1998a)
Soil Background Line (SBL)	(B8 –(B4 1.24))	
SOCI (Soil Organic Carbon Index)	(B2/(B4 B3))	Organic Carbon (Thaler et al. 2019)
Grain Size Index (GSI)	(B4-B2)/(B4+B3+B2)	Soil Texture (Xiao et al. 2006)
Iron Oxide Ratio (IOR)	(B4/B2)	Iron Oxide soil content (Segal 1982)
Ferric Iron 3+ Index	(B4/B3)	Iron 3+ soil content (Rowan and Mars 2003)
Salt Index (SIndx)	$\sqrt{(B4\ B2)}$	Salt soil content (Douaoui et al. 2006)
Normalize Salt Index (NSI)	(B4-B8)/(B4+B8)	Salt soil content (Douaoui et al. 2006)
Salinity Index 2	$\sqrt{(B3 + B4 + B8)^2}$	Salt Soil content (Douaoui et al. 2006)

Table 2. Bare soil indices.

2.2.2. Organism-Vegetation (o)

The organism is related to vegetation. Therefore, a multitemporal analysis of Sentinel-2 data on the GEE platform was carried out. Fifty-two images from 2019-2020 with at least 80% of the area without clouds and shadows were used.

NDVI and MSAVI (Modified Soil Adjustment Vegetation Index) vegetation indices (Table 4) were calculated for each image.

Finally, all images' minimum, median end maximum pixels values were extracted.

Vegetation Index	Formula
Normalized vegetation Index	$(B4^2 + B3^2 + B2^2/3)^{0.5}$
Modified Soil Adjustment Vegetation Index	(B4-B2)/(B4+B2)

Table 3. Vegetation indices.

2.2.3. Regarding Relief- Topografic (r)

DEM provided by Emilia Romagna Region (https://geoportale.regione.emilia-romagna.it/catalogo/dati-cartografici) with a pixel spatial resolution of 5 meters, was resampled at the exact resolution of VIS-NIR bands of Sentinel-2 Satellite (10 meters) using the algorithm Nearest Neighbor in Qgis. Using the resampled DEM, several topographic indices were calculated (Table 5) using Qgis software.

Topographic Indices
Slope-
Tangential Curvature
Longitudinal Curvature
Total Curvature
Profile Curvature
General Curvature
Plan Curvature
Minimal Curvature
Maximal Curvature
Aspect-
Hillshade
Elevation

Table 4. Topographic Indices.

Concerning Spatial Position (n), latitude and longitude maps were produces in a 10-m resolution as spatial position predictor.

2.2.4. Mathematical function (f)

In this study, the performance of Multiple Linear Regression (MLR), Geographically Weighted Regression (GWR), Regression Kriging (RK), and Multivariate Adaptive Regression Spline (MARS) were compared.

2.2.4.1.　　　Multiple Linear Regression (MLR)

MLR is widely used in DSM; it extends simple linear regression. Multiple Linear regression can be written as:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon_i$$

Where y is the predicted ECa, k is the number of SCORPAN factors, βi is regression coefficient, x_i is SCORPAN factor and ε_i represents the error term. MLR was carried out in R environment.

2.2.4.2.　　　Geographically Weighted Regression (GWR)

GWR is an extension of the traditional regression (Brunsdon et al. 2010). In traditional regression, it is assumed that the model coefficients are constant across the study area, but this is not necessarily true in many situations. Therefore, in the GWR, the regression coefficients are specific to a location rather than global estimates.

The GWR methods can be represented as:

$$y(x, y) = \beta_0(x, y) + \sum_{i=1}^{k} \beta_i(x, y)\, x_i + \varepsilon_i(x, y)$$

Where $y(x, y)$ is the predicted ECa at site i, (x,y) are the coordinates for the i site, k is the number of SCORPAN factors, βi is regression coefficient, x_i is SCORPAN factor at the site i and $\boldsymbol{\varepsilon_i}$ Represents the error term.

The bandwidth chosen in GWR is a key affecting the regression results. The Adaptive (AICc) methods were used to choose the best bandwidth in this study. For adaptive kernel size, the weight of each point can be calculated using the Gaussian function. GWR was carried out in R environment using GWmodel library.

2.2.4.3.　　　Regression Kriging (RK)

RK is one of the most popular spatial interpolation techniques in digital soil mapping (Keskin and Grunwald 2018). RK is a combination of the traditional MLR and kriging. In this study, the residual generated from MLR were kriged and then added to the predicted drift. The model is expressed as:

$$y_{mlr\,(x,y)} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon_i$$

$$y_{rk} = y_{mlr(x,y)} + \sum_{i=1}^{k} w_i(x,y)\, \varepsilon_i(x,y)$$

Where y_{mlr} is the predicted ECa at site i using MLR, y_{rk} is the predicted ECa at site i using the RK. model, $w_i(x,y)$ are weights determined by the covariance function and ε_i Represents the error term. Regression Kriging was carried out in R environment using Automap library.

2.2.4.4.   Multivariate Adaptive Regression Spline (MARS)

MARS is a non-parametric model. MARS combines the classical linear regression, the mathematical construction of splines, the binary recursive partitioning, and brute search, and intelligent algorithms to produce a model to predict the value of a target variable from a set of independent variables. MARS general expression can be written as follows:

$$y = f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x)$$

Where y is the value predicted by the model using a function $f(x)$, which can be decomposed into an initial constant β_0 and a sum of M terms, each of them are formed by a coefficient β_m and basic functions $h_m(x)$. The basic function (BF) represents the information included in one or more independent variables and is selected step-by-step.

2.3.   Model evaluation

Using Caret library in R environment, all the ECa sample sites were divided into two classes for calibration (80% n=98) and validation (20% n= 24).
The agreement between predictions and observations is generally carried out using Pearson Correlation or Spearman indices. According to (Steichen and Cox 2002) Lin's concordance coefficient is preferred to quantify the goodness of model adaptation. Lin's concordance coefficient provides a measure of overall accuracy that takes into account both bias correction and precision. The formula as follows:

$$\rho_c = \frac{2\, r\, s_x\, s_y}{s_Y^2 + s_X^2 + (m_y - m_x)^2}$$

Where r is the Pearson's coefficient; $s_x\ s_y$ are the standard deviation of the true and predicted values; m_y and m_x are the means; and s_Y^2 and s_x^2 are the variances of the true and predicted values.

Therefore Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

$$RMSE = \sqrt{(\sum_{i=1}^{n}[D(x_i)] - D(x_i)]^2}$$

$$MAE = \frac{\sum_{i=1}^{n}[D(x_i)] - D(x_i)]}{n}$$

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{n}\left|\frac{(D(x_i) - D*(x_i))}{D(x_i)}\right|$$

D (xi) is the measurement of ECa, *D (x_i)* is the predicted ECa, and n is the number of validation sites, respectively.

2.4.   Preprocessing

Pearson correlation analysis was conducted to select the environmental covariates significantly correlating with ECa. Then a model for ECa was constructed by stepwise multiple regression using Akaike information criterion (AIC) as a selection criterion.

After removing collinearity using multiple stepwise regression, the remaining variables were selected for modeling.

The ECa sample points of the fields 3, 5, 16, 19 were interpolated using Inverse Distance weighting (IDW) in Qgis with a spatial pixel resolution of 10 meters (Fig. 6 -12).

Then ECa interpolated maps were aligned with Bare soil indices using Qgis.

Finally, Pearson correlation between ECa maps and SYSI images was conducted to assess SYSI correlation between SYSI and interpolated ECa (ECa(idw)).
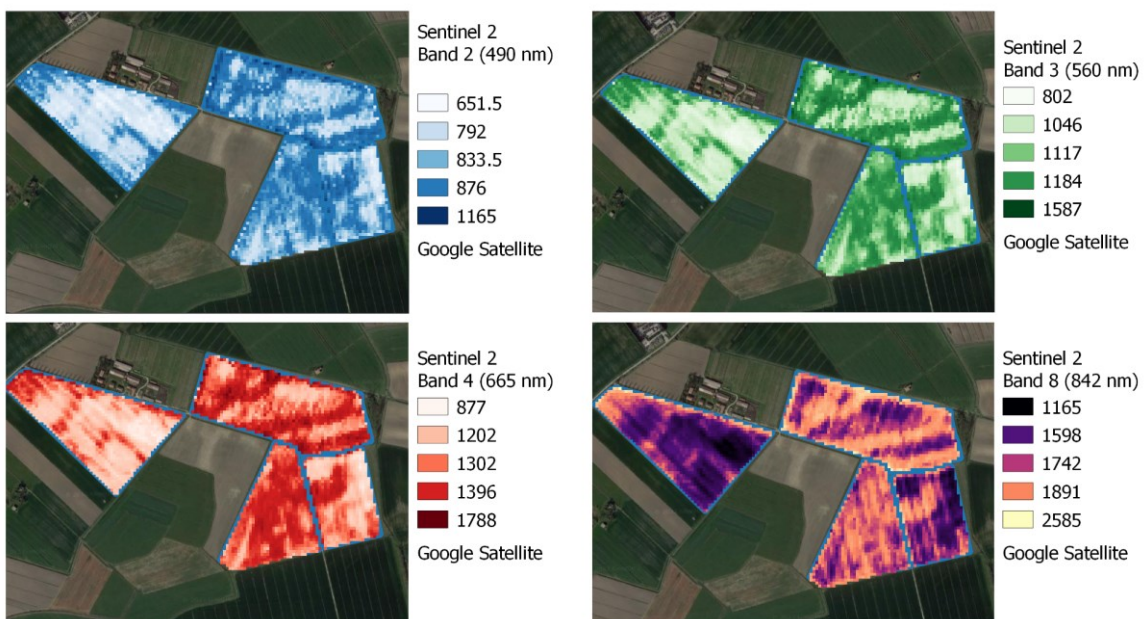
Figure 5. SYSI VIS-NIR bands.

3. Results and discussion

3.1. Descriptive statistics of ECa

In field 19, the ECa ranged from 15.94 to 77.72-ohm meter ($\Omega \cdot m$), with mean and standard deviation (SD) 46.74 and 11.28, respectively. The ECa data approximately follow a normal distribution (Skewness= -0.02, kurtosis= 0.15). The Coefficient of Variation (CV) (0.24) shows a low variability in the dataset.

In field 3, the minimum value was 15.35, and the maximum value was 93.54. The mean was 49.98, and the standard deviation 13.00. The skewness was -0.02 and kurtosis -0.54. The variability in the dataset was low CV = 0.26.

In field 5 ranged from 18.54 to 76.38-ohm meter ($\Omega \cdot m$), mean 46.94 and standard deviation 10.76. The skewness and kurtosis were 0.12, -0.39. CV was 0.23.

In field 16, the minimum value was 25.56, and the maximum value was 68.13. The mean was 46.86, and the standard deviation 7.79. The skewness was 0.07 and kurtosis -0.03. CV was 0.17.

3.2. Results of the sampling soil

The 122 sample points were classified according to USDA Textural soil classification, as seen in Fig 5. The soil textural classes were: (Cl) Clay, (ClLo) Clay Loam, (Lo) Loam, (SiCl) Silty Clay, (SiClLo) Silty Clay Loam, (SiLo) Silty loam.
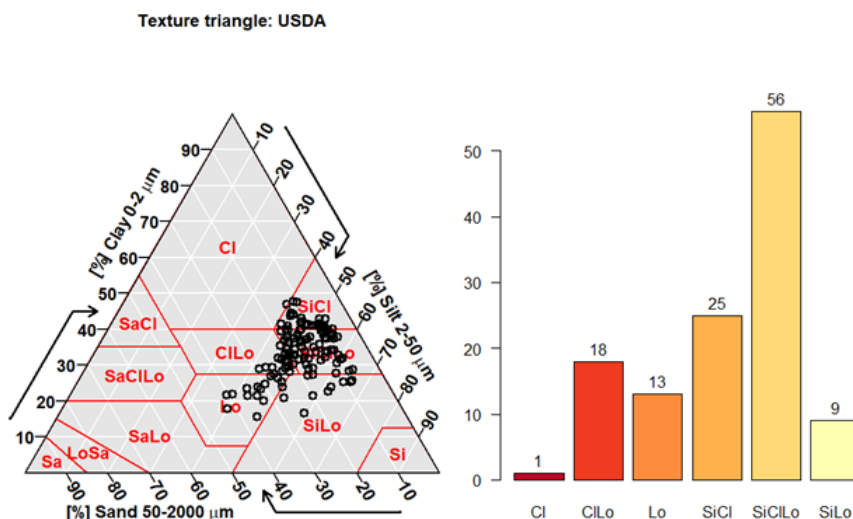
Figure 6. Textural soil classification.

3.3. Correlation between ECa(idw) and soil

The Pearson correlation coefficient (r) among ECa(idw) with soil properties was highly significant ($p < 0.1$) except with silt ($p < 0.1$), as reported in Table 6.

Clay and SOC were strongly correlated to ECa(idw) $r \geq 0.72$. The sand was negatively related r - 0.50, while a low correlation was found between Silt and ECa(idw).

These results confirmed the hypothesis that soil electrical conductivity is related to Clay and OM in no-salt soil, according to literature (Corwin and Lesch 2005a, 2005b; Corwin and Scudiero 2020; Friedman 2005; Sudduth et al. 2005.).

According to (Corwin and Lesch 2005b)in ECa measurement, three parallel pathways contribute 1) a solid-liquid pathway, primarily via exchangeable cation associated with clay minerals, 2) a liquid phase pathway via salts contained in the soil water occupying the large pores, and finally 3) solid pathway via soil particles that are in direct contact with another.

Consequently, clay content and type, cation exchange capacity, and O.M. influence ECa measurements. Therefore, ECa in this study area was a good proxy of soil variability.

	r	*p*
Clay- ECa(idw)	0.73	0.00
Sand- ECa(idw)	-0.50	0.00
SOC- ECa(idw)	0.72	0.00
Silt ECa(idw)	-0.17	0.07

Table 5. Correlation between ECa(idw) and soil.
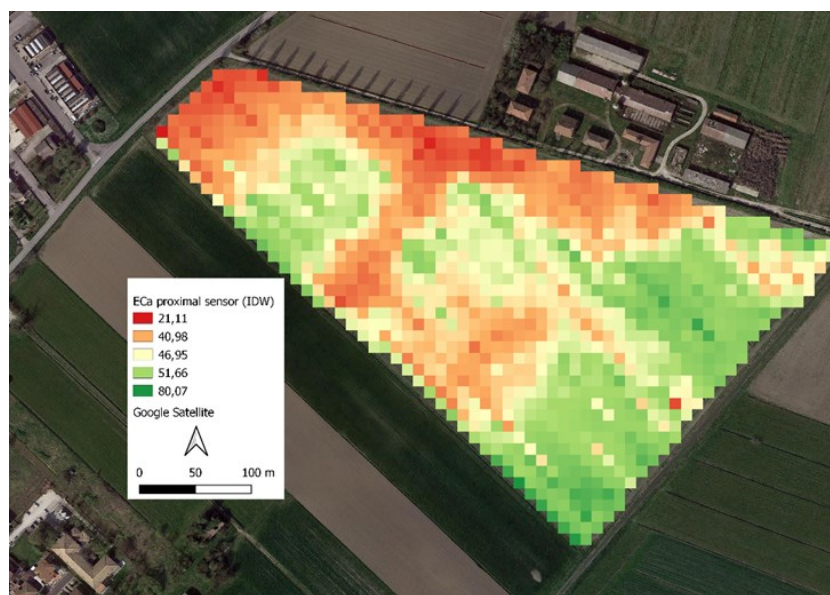
Figure 6. ECa interpolated using IDW.

3.4. Correlation between ECa(idw) and Topographic factors

Regarding topographic factors, Table 7 shows weak correlation ($r \leq 0.1$) for Slope (-0.08), Tangential Curvature (0.07), Longitudinal Curvature (0.08), Profile Curvature (0.07), General Curvature (0.08), Minimal Curvature (0.08), Maximal Curvature (0.07) and Aspect (0.07). The topographic factors of the field (19) could indirectly affect texture and organic matter (Jiang and Thelen n.d.; Kravchenko and Bullock n.d.). Therefore, topographic factors showed a low correlation with ECa, mainly because of their low field variability (Mulder et al. 2011).

Finally, Hillshade (0.28) showed the best correlation over other topographic variables.

It is known that hillshade strongly affects microclimate, related to solar radiation and soil water content. Besides Clay and O.M., soil water content is one of the main factors affecting ECa measurement (Corwin and Plant 2005).

3.5. Correlation between ECa(idw) and Vegetation Indices

Concerns Vegetation factors (Table 7), NDVI_mean, OSAVI_mean, MSAVI_mean, have a significant correlation ($p < 0.05$) with ECa(idw).

In particular, the Vegetation Indices (VI) were negatively correlated with ECa(idw). Generally, VI has a positive correlation with ECa. That means that in portions of the field with low values of ECa (Poorly of the OM and Clay), vegetation grows less than the portion of the field with a high level of ECa.

In this study, the negative correlation can be explained due to the behavior of the Durum wheat (*Triticum durum* Desf.) *cv*. "Senatore Cappelli" (Chapter 3).

It is characterized by very tall plants and frequent phenomena of lodged grain. Moreover, soil and topography properties explain only 13% of with-in yield variability (Chapter 3).

3.6. Correlation between ECa(idw) and Soil and Lithology factors

All Soil and Lithology factors had a significant correlation ($p < 0.05$) with ECa(idw) as reported in Table 9 and 10.

B2 (-0.29) had shown the lowest correlation between bands of SYSI. While B4 was highly correlated with ECa(idw) (-0.62) followed by B3 (- 0.60), B8 (- 0.56).

VIS-NIR SYSI bands were related to soil properties, as reported in table 11. In particular, Clay was highly correlated with some VIS -NIR SYSI bands (B3 r - 0.70, B4 r - 0.74, B8 r - 0.70) while Silt, Sand, and OC showed low correlations.

These results confirmed that SYSI images could be a good proxy of soil properties according to literature (Demattê et al. 2004; Demattê et al. 2018; Mendes et al. 2019; Safanelli et al. 2020a; Silvero et al. 2021). Soil reflectance is related to soil color (Douaoui et al. 2006); a study in Europe (Safanelli et al. 2020b) showed that the median reflectance estimated by the full SYSI had a higher correlation with the reference spectra collected in the laboratory. Soil color is related to the visible portion of the electromagnetic spectrum (Sovdat et al. 2019). Soil color provides an indirect measure of several soil properties (Mathieu et al. 1998a).

Moreover, soil colors is an essential component of the Visual Soil Assessment (VSA) Shepherd et al. (2008).

VSA properties provide a semi-quantitative and cost-effective method to assess and monitoring soil properties and soil quality (Shepherd et al. 2008).

Soil color is related to the soil organic matter content, iron (Fe) oxidation state, and manganese (Mn). In general, the darker the color is, the greater is the amount of organic matter in the soil (Shepherd et al. 2008).

Soil color is related to soil drainage, soil aeration degree, and, consequently, soil texture (Shepherd et al. 2008). Brown, yellow-brown, reddish-brown, and red soil indicate well-aerated, well-drained conditions where Fe (iron) and Mn (manganese) occur in the oxidized form of ferric (Fe^{3+}) and manganic (Mn^{3+}) oxides. Grey-blue colors can indicate that the soil is poorly drained and poorly aerated for long periods, reducing Fe and Mn to ferrous (Fe^{2+}) and manganic (Mn^{2+}) oxides (Shepherd et al. 2008).

Therefore, in field 19, as reported in table 9, ECa was correlated to bare indices, related to soil color, in particular to S.I. (*r* -0.67) and B.I. (*r* -0.60).

In addition, ECa was correlated to GSI (-0.67). It is related to topsoil grain size distribution, which positively correlates with sand (Xiao et al. 2006).

ECa showed a positive correlation with SOCI (0.68), which is related to soil organic carbon (Thaler et al. 2019) and IOR (0.65). IOR is related to iron soil content (Segal 1982).

Finally, ECa was correlated to SIindx2 (*r* -0.60), latter is related to Soil electrical conductivity (Douaoui et al. 2006).

	r	*p*
Slope-ECa(idw)	-0.08	0.00
Tangential Curv - ECa(idw)	0.07	0.01
Longitudinal Curv.– ECa	0.08	0.00
Total Curvature- ECa(idw)		
Profile Curvature - ECa(idw)	0.07	0.00
General Curvature ECa(idw)	0.08	0.00
Plan Curvature- ECa(idw)		
Minimal Curv. - ECa(idw)	0.08	0.00
Maximal Curv. - ECa(idw)	0.07	0.02
Aspect- ECa(idw)	0.07	0.02
Hillshade -ECa(idw)	0.28	0.00
Elevation-ECa(idw)		

Table 6. Correlation between ECa(idw) and topographic factors.

	r	*p*
NDVImean- ECa(idw)	-0.20	0.00
NDVImax- ECa(idw)		
OSAVImean– ECa(idw)	-0.20	0.00
OSAVImax- ECa(idw)		
MSAVImean - ECa(idw)	-0.19	0.00
MSAVImax-ECa(idw)		

Table 7. Correlation between ECa(idw) and vegetation indices.

	r	*p*
SI - ECa(idw)	-0.67	0.00
SBL-ECa(idw)	-0.21	0.00
HI- ECa(idw)	0.12	0.00
CI- ECa(idw)	-0.41	0.00
BI- ECa(idw)	-0.60	0.00
NSI- ECa(idw)	0.13	0.00
SIindx- ECa(idw)	-0.45	0.00
SIindx2- ECa(idw)	-0.60	0.00
GSI- ECa(idw)	-0.67	0.00
SOCI- ECa(idw)	0.68	0.00
FE3- ECa(idw)	-0.41	0.00
IOR- ECa(idw)	0.65	0.00

Table 8. Correlation between ECa(idw) and bare soil indices.

	r	*p*
B2- ECa(idw)	-0.29	0.00
B3- ECa(idw)	-0.60	0.00
B4– ECa(idw)	-0.62	0.00
B8- ECa(idw)	-0.56	0.00

Table 9. Correlation between ECa(idw) and Sysi Vis-NIR bands.

	r	*p*
Clay- B2	-0.46	0.00
Clay- B3	-0.70	0.00
Clay-B4	-0.74	0.00
Clay – B8	-0.70	0.00
Silt - B2	0.41	0.00
Silt – B3	0.40	0.00
Silt- B4	0.38	0.00
Silt -B8	0.42	0.00
Sand - B2		
Sand – B3	0.30	0.00
Sand – B4	0.34	0.00
Sand – B8	0.27	0.00
O.C. – B2		
OC - B3	-0.47	0.00
OC – B4	-0.52	0.00
OC – B8	-0.50	0.00

Table 10. Correlation between soil and Sysi Vis-NIR bands.

3.7. Environmental covariates for regression models

The correlations between environmental covariates were strong, which may cause the collinearity problem. These factors were removed by stepwise regression. Therefore, forward stepwise multiple regression was then performed, and only environmental covariates in which contribution to regression was significant (p < 0.10) were retained in the regression equation. The covariates selected were:

$$ECa \sim Maximal\ curvature + hillshade + B2 + SBL + SIindx2 + GSI + NDVI\ meaN + latitude + longitude$$

3.8. Comparison of the performance of models

The performance of four models was evaluated on the training data for the coefficient of determination, while the Lin's concordance coefficient and RMSE, MAE, MAPE on the test data. Evaluation results of the estimation are summarized in Table 12.

Model	R^2	r	Lin's concordance	RMSE	MAE	MAPE
MLR	0.57	0.69	0.96	8.21	6.62	15.29
GWR	0.83	0.91	0.99	4.80	3.73	8.53
RK	0.74	0.86	0.99	5.80	4.31	10.01
MARS	0.73	0.80	0.99	6.60	5.28	11.91

Table 11. Performance of models.

GWR produced a high determination coefficient (0.83) and correlation coefficient (0.91). GWR, RK, and MARS showed the same accuracy (0.99), but GWR generated low RMSE (4.80), MAE (3.73), and MAPE (8.53) values.

The best performance of GWR can be explained due to the relationships between topsoil ECa, and environmental covariates are not fixed across space as reported in Table 13.

These regression parameters estimated in the GWR analysis are not fixed across space, which means that the relationships between topsoil E.C. and environmental covariates are spatially non-stationary. This assumption is confirmed by several authors (Terrón et al. 2011; Wu et al. 2017; Yang et al. 2019).

	Min.	1st Qu	Median	3rd Qu	Max
Intercept	-369538.12	111302.94	274391.24	559681.15	1189105.66
Maximal curvature	-4696.57	- 5235.29	886.44	1714.24	15.01
Hillshade	-4.39	-0.55	0.65	1.98	9.00
B2	-1.65	-0.02	0.24	0.58	2.16
SBL	-0.34	0.00	0.06	0.14	0.44
SIindx2	-0.46	-0.13	-0.06	-0.00	0.35
GSI	-4330.61	186.35	514.39	1542.88	6015.46
NDVI mean	-281.30	88.17	-45.68	14.29	126.31
Latitude	-0.50	-0.04	-0.00	0.04	0.27
Longitude	-0.24	0.11	-0.05	-0.02	0.07

Table 12. Coefficients of GWR.

3.9. Predictive map of soil ECa

The map generated by GWR was chosen to show the spatial distribution of soil ECa (Fig.7).

Figure 8, highlights how ECa spatial distribution follows a slope–elevation gradient.

The high values of ECa were found in accumulation areas, while low values of ECa were found in areas with greater slopes.

This means that soil with a high content of Clay and SOC is located in concave areas of the field. The fact can explain that field topography changed soil water availability and influenced soil particle redistribution, organic matter and soil nutrients, as reported to [10–12].

The maps of the ECa original data (Fig. 6), interpolated by IDW and ECa maps by GWR (Fig. 7) showed similar spatial patterns, and also by GWR and slope-elevation gradient (Fig. 8). This was confirmed by the difference map (Fig. 9).

This map shows similar spatial patterns across the space (mean of raster: 0.75 ($\Omega \cdot m$)).

These results confirmed the first hypothesis of this study, that is, the ability of the SCORPAN model, to predict the spatial variation of soil properties at field scale, using SYSI and other environmental variables.
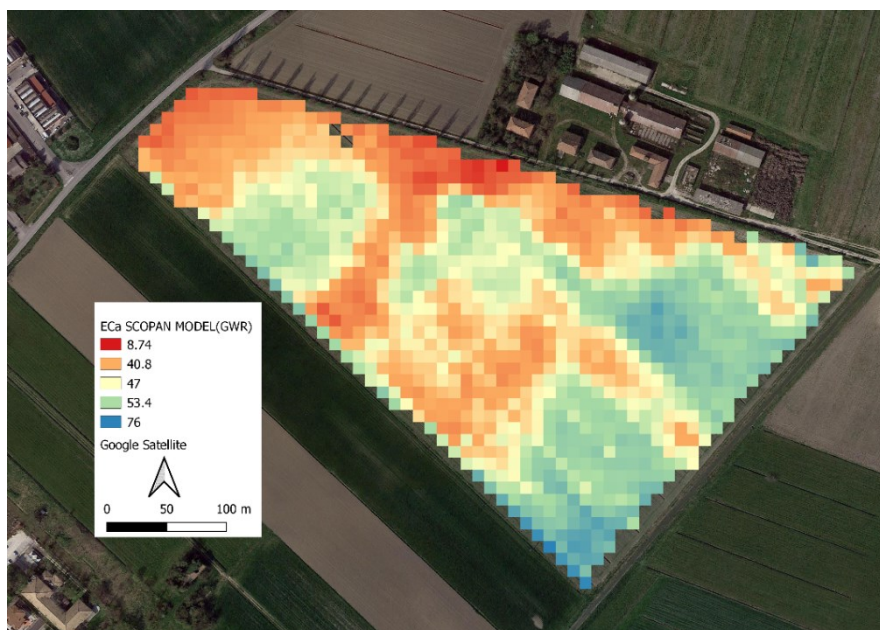
Figure 7. ECa map generated by GWR.
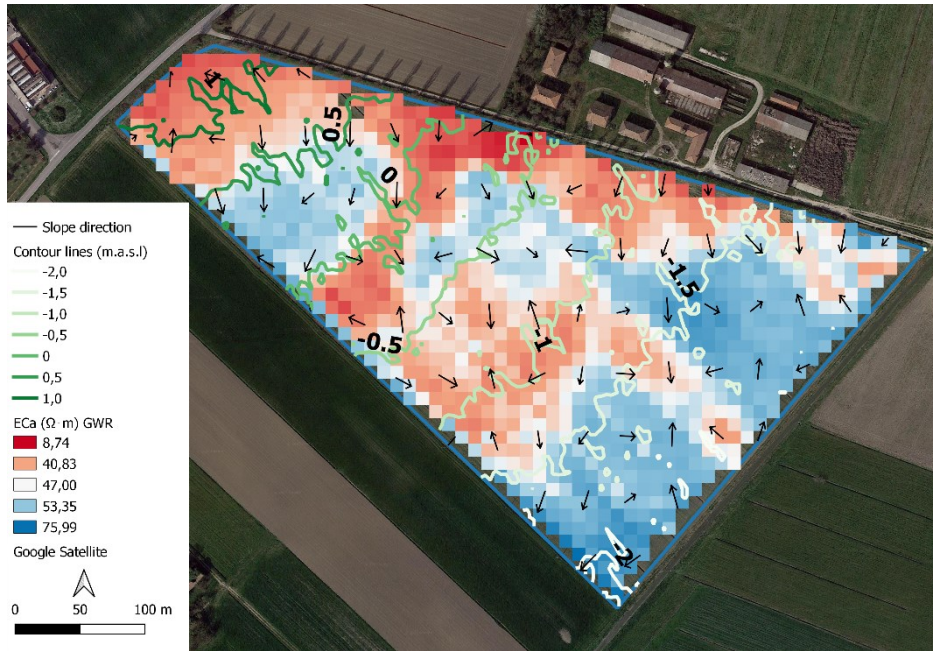
Figure 8. ECa map generated by GWR and slope-elevation gradient.

Figure 9. Difference pixel map.

3.10. SYSI and within-field variability

The second aim of this study was to assess the multitemporal bare soil image (SYSI) potential to intercept and spatialize within-field variability. The Pearson correlation between ECa interpolated and SYSI bands and bare soil indices for each field, is reported in tables 14, 15, 16, 17.

In this section, only indices with a correlation coefficient (r) greater than 0.60 have been taken into account.

Index	r	p
IOR- ECa(idw)	-0.63	0.00
SI- ECa(idw)	-0.63	0.00
GSI- ECa(idw)	-0.62	0.00
B4- ECa(idw)	-0.60	0.00
BI- ECa(idw)	-0.58	0.00
B8- ECa(idw)	-0.56	0.00
CI- ECa(idw)	-0.55	0.00
FE3+ -ECa(idw)	-0.55	0.00
SIindx2- ECa(idw)	-0.54	0.00
HI- ECa(idw)	-0.43	0.00
SIindx- ECa(idw)	-0.40	0.00
B3- ECa(idw)	-0.32	0.00
SBL-ECa(idw)	-0.16	0.00
B2- ECa(idw)	-0.08	0.00
NSI- ECa(idw)	0.09	0.00
SOCI- ECa(idw)	0.65	0.00

Table 13. Correlation between ECa(idw) and SYSI for Field 3.

Index	r	p
B8- ECa(idw)	-0.72	0.00
BI- ECa(idw)	-0.72	0.00
SIindx2- ECa(idw)	-0.72	0.00
B4- ECa(idw)	-0.71	0.00
GSI- ECa(idw)	-0.68	0.00
IOR - ECa(idw)	-0.68	0.00
SI- ECa(idw)	-0.68	0.00
SIindx-- ECa(idw)	-0.67	0.00
B3 -ECa(idw)	-0.65	0.00
FE3+- ECa(idw)	-0.60	0.00
CI- ECa(idw)	-0.59	0.00
B2- ECa(idw)	-0.50	0.00
HI- ECa(idw)	-0.34	0.00
SBL-ECa(idw)	-0.32	0.00
NSI- ECa(idw)	-0.25	0.00
NSI- ECa(idw)	-0.09	0.00
SOCI- ECa(idw)	0.65	0.00

Table 14. Correlation between ECa(idw) and SYSI for Field 5.

Index	r	p
B4- ECa(idw)	-0.64	0.00
BI- ECa(idw)	-0.63	0.00
SIindx2- ECa(idw)	-0.63	0.00
B8- ECa(idw)	-0.59	0.00
B3- ECa(idw)	-0.58	0.00
SIindx- ECa(idw)	-0.57	0.00
GSI- ECa(idw)	-0.49	0.00
IOR -ECa(idw)	-0.49	0.00
SI- ECa(idw)	-0.49	0.00
CI- ECa(idw)	-0.46	0.00
FE3+- ECa(idw)	-0.46	0.00
B2 - ECa(idw)	-0.38	0.00
HI-ECa(idw)	-0.27	0.00
NSI- ECa(idw)	-0.04	0.00
SBL- ECa(idw)	-0.02	0.00
SOCI- ECa(idw)	0.63	0.00

Table 15. Correlation between ECa(idw) and SISY for Field 16.

Index	r	p
GSI- ECa(idw)	-0.79	0.00
IOR- ECa(idw)	-0.79	0.00
SI- ECa(idw)	-0.79	0.00
B4- ECa(idw)	-0.63	0.00
BI- ECa(idw)	-0.63	0.00
B8- ECa(idw)	-0.62	0.00
SIindx2- ECa(idw)	-0.62	0.00
B3 -ECa(idw)	-0.57	0.00
CI- ECa(idw)	-0.55	0.00
FE3+- ECa(idw)	-0.54	0.00
SIindx- ECa(idw)	-0.46	0.00
SBL - ECa(idw)	-0.31	0.00
B2-ECa(idw)	-0.17	0.00
HI- ECa(idw)	-0.03	0.00
NSI- ECa(idw)	-0.02	0.00
SOCI- ECa(idw)	0.75	0.00

Table 16. Correlation between ECa(idw) and SYSI for Field 19.

SOCI (Fig. 10) showed a good correlation with ECa(idw) in all fields, r 0.63 -0.75. Confirming Its relationship with SOC(Thaler et al. 2019) and consequently with it s (Corwin and Lesch 2005a, 2005b; Corwin and Plant 2005; Corwin and Scudiero 2020).

B4 (Fig. 11) showed good correlations with ECa(idw) concerning VIS-NIR bands, ranging from 0.60 to 0.71. B4 was related to Clay and SOC as reported in Table 11, according to [29].

Figure 10. SOCI, Soil Organic Carbon Index.

Figure 11. Sentinel-2 Band 4.

Figure 12. ECa interpolated using IDW.

4.   Conclusions

Geospatial ECa mapping is a sensor technology that has a significant role in spatial field characterization. However, this method has several limitations; therefore, the rapid expansion of Precision Agriculture required an alternative, faster way to map spatial field variability. Multi-temporal images with bare soil pixels (SYSI), obtained from the time series of Sentinel-2 satellite, can be a valid alternative to ECa, understood as a proxy of soil variability. In particular, the red band (B4) of SYSI and the bare soil index SOCI (Soil Organic Carbon Index) showed interesting results. These results require further validations in other areas of study. This method could be a vital impulse for Precision Agriculture diffusion. Therefore, this result could open the way to new accurate and cheaper services.

5.   References

Pretty, J.N. The Sustainable Intensification of Agriculture. *Natural Resources Forum* **1997**, *21*, 247–256, doi:10.1111/j.1477-8947.1997.tb00699.x.

Farooq, M.; Rehman, A.; Pisante, M. Sustainable Agriculture and Food Security. In *Innovations in Sustainable Agriculture*; Springer International Publishing: Cham, 2019; pp. 3–24.

Pisante, M.; Stagnari, F.; Grant, C.A. Agricultural Innovations for Sustainable Crop Production Intensification. *Italian Journal of Agronomy* **2012**, *7*, 40, doi:10.4081/ija.2012.e40.

Cillis, D.; Pezzuolo, A.; Marinello, F.; Basso, B.; Colonna, N.; Furlan, L.; Sartori, L. Conservative Precision Agriculture: An Assessment of Technical Feasibility and Energy Efficiency within the LIFE+ AGRICARE Project. *Advances in Animal Biosciences* **2017**, *8*, doi:10.1017/S204047001700019X.

Sudduth, K.A.; Hummel, J.W.; Birrell, S.J. *Sensors for Site-Specific Management*; 1997;

Plant, R.E. Site-Specific Management: The Application of Information Technology to Crop Production. *Computers and Electronics in Agriculture* **2001**, *30*, 9–29, doi:10.1016/S0168-1699(00)00152-6.

Blackmore, S. The Interpretation of Trends from Multiple Yield Maps. *Computers and Electronics in Agriculture* **2000**, *26*, 37–51, doi:10.1016/S0168-1699(99)00075-7.

Maestrini, B.; Basso, B. Predicting Spatial Patterns of Within-Field Crop Yield Variability. *Field Crops Research* **2018**, *219*, 106–112, doi:10.1016/J.FCR.2018.01.028.

Kravchenko, A.N.; Bullock, D.G. *SPATIAL VARIABILITY Correlation of Corn and Soybean Grain Yield with Topography and Soil Properties*;

Jiang, P.; Thelen, K.D. *SITE-SPECIFIC ANALYSIS Effect of Soil and Topographic Properties on Crop Yield in a North-Central Corn-Soybean Cropping System*;

Mulder, V.L.; de Bruin, S.; Schaepman, M.E.; Mayr, T.R. The Use of Remote Sensing in Soil and Terrain Mapping — A Review. *Geoderma* **2011**, *162*, 1–19, doi:10.1016/J.GEODERMA.2010.12.018.

Kühn Á Wehrhan Á S Koszinski Á M Sommer, J.M.; Kühn, J.; Brenning, A.; Sommer, M. Interpretation of Electrical Conductivity Patterns by Soil Properties and Geological Maps for Precision Agriculture. *Precision Agric* **2009**, *10*, 490–507, doi:10.1007/s11119-008-9103-z.

Corwin, D.L.; Lesch, S.M. Apparent Soil Electrical Conductivity Measurements in Agriculture. *Computers and Electronics in Agriculture* **2005**, *46*, 11–43, doi:10.1016/j.compag.2004.10.005.

Corwin, D.L.; Scudiero, E. Field-Scale Apparent Soil Electrical Conductivity. *Soil Science Society of America Journal* **2020**, *84*, 1405–1441, doi:10.1002/saj2.20153.

Castrignanò, A.; Wong, M.T.F.; Stelluti, M.; de Benedetto, D.; Sollitto, D. Use of EMI, Gamma-Ray Emission and GPS Height as Multi-Sensor Data for Soil Characterisation. *Geoderma* **2012**, *175–176*, 78–89, doi:10.1016/J.GEODERMA.2012.01.013.

Corwin, D.L.; Lesch, S.M. Characterizing Soil Spatial Variability with Apparent Soil Electrical Conductivity: I. Survey Protocols. *Computers and Electronics in Agriculture* **2005**, *46*, 103–133, doi:10.1016/J.COMPAG.2004.11.002.

Sudduth, K.A.; Kitchen, N.R.; Wiebold, W.J.; Batchelor, W.D.; Bollero, G.A.; Bullock, D.G.; Clay, D.E.; Palm, H.L.; Pierce, F.J.; Schuler, R.T.; et al. Relating Apparent Electrical Conductivity to Soil Properties across the North-Central USA. *Computers and Electronics in Agriculture* **2005**, *46*, 263–283, doi:10.1016/J.COMPAG.2004.11.010.

Barca, E.; de Benedetto, D.; Stellacci, A.M. Contribution of EMI and GPR Proximal Sensing Data in Soil Water Content Assessment by Using Linear Mixed Effects Models and Geostatistical Approaches. **2019**, doi:10.1016/j.geoderma.2019.01.030.

Buttafuoco, G.; Castrignanò, A.; Cucci, G.; Lacolla, G.; Lucà, F. Geostatistical Modelling of Within-Field Soil and Yield Variability for Management Zones Delineation: A Case Study in a Durum Wheat Field. *Precision Agriculture* **2017**, *18*, 37–58, doi:10.1007/s11119-016-9462-9.

Castrignanò, A.; Buttafuoco, G.; Quarto, R.; Parisi, D.; Viscarra Rossel, R.A.; Terribile, F.; Langella, G.; Venezia, A. A Geostatistical Sensor Data Fusion Approach for Delineating Homogeneous Management Zones in Precision Agriculture. *CATENA* **2018**, *167*, doi:10.1016/j.catena.2018.05.011.

Barca, E.; Castrignanò, A.; Buttafuoco, G.; de Benedetto, D.; Passarella, G. Integration of Electromagnetic Induction Sensor Data in Soil Sampling Scheme Optimization Using Simulated Annealing. *Environmental Monitoring and Assessment* **2015**, *187*, 422, doi:10.1007/s10661-015-4570-y.

Scudiero, E.; Teatini, P.; Corwin, D.L.; Deiana, R.; Berti, A.; Morari, F. Delineation of Site-Specific Management Units in a Saline Region at the Venice Lagoon Margin, Italy, Using Soil Reflectance and Apparent Electrical Conductivity. *Computers and Electronics in Agriculture* **2013**, *99*, 54–64, doi:10.1016/J.COMPAG.2013.08.023.

Terrón, J.M.; da Silva, J.R.M.; Moral, F.J.; García-Ferrer, A. Soil Apparent Electrical Conductivity and Geographically Weighted Regression for Mapping Soil. *Precision Agriculture* **2011**, *12*, 750–761, doi:10.1007/s11119-011-9218-5.

Weiss, M.; Jacob, F.; Duveiller, G. Remote Sensing for Agricultural Applications: A Meta-Review. *Remote Sensing of Environment* **2020**, *236*, 111402, doi:10.1016/J.RSE.2019.111402.

25.	Castaldi, F.; Hueni, A.; Chabrillat, S.; Ward, K.; Buttafuoco, G.; Bomans, B.; Vreys, K.; Brell, M.; van Wesemael, B. Evaluating the Capability of the Sentinel 2 Data for Soil Organic Carbon Prediction in Croplands. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *147*, 267–282, doi:10.1016/J.ISPRSJPRS.2018.11.026.

Castaldi, F.; Palombo, A.; Santini, F.; Pascucci, S.; Pignatti, S.; Casa, R. Evaluation of the Potential of the Current and Forthcoming Multispectral and Hyperspectral Imagers to Estimate Soil

Texture and Organic Carbon. *Remote Sensing of Environment* **2016**, *179*, 54–65, doi:10.1016/J.RSE.2016.03.025.

Poggio, L.; Gimona, A. Assimilation of Optical and Radar Remote Sensing Data in 3D Mapping of Soil Properties over Large Areas. *Science of the Total Environment* **2017**, *579*, 1094–1110, doi:10.1016/j.scitotenv.2016.11.078.

Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A Powerful Data Mining Procedure to Retrieve Soil Spectral Reflectance from Satellite Images. *Remote Sensing of Environment* **2018**, *212*, 161–175, doi:10.1016/j.rse.2018.04.047.

Silvero, N.E.Q.; Demattê, J.A.M.; Amorim, M.T.A.; Santos, N.V. dos; Rizzo, R.; Safanelli, J.L.; Poppiel, R.R.; Mendes, W. de S.; Bonfatti, B.R. Soil Variability and Quantification Based on Sentinel-2 and Landsat-8 Bare Soil Images: A Comparison. *Remote Sensing of Environment* **2021**, *252*, 112117, doi:10.1016/j.rse.2020.112117.

Fongaro, C.T.; Demattê, J.A.M.; Rizzo, R.; Safanelli, J.L.; de Sousa Mendes, W.; Dotto, A.C.; Vicente, L.E.; Franceschini, M.H.D.; Ustin, S.L. Improvement of Clay and Sand Quantification Based on a Novel Approach with a Focus on Multispectral Satellite Images., doi:10.3390/rs10101555.

Mendes, W.D.S.; Medeiros Neto, G.; Demattê, J.A.M.; Gallo, B.C.; Rizzo, R.; Safanelli, J.L.; Fongaro, C.T. Is It Possible to Map Subsurface Soil Attributes by Satellite Spectral Transfer Models? **2019**, doi:10.1016/j.geoderma.2019.01.025.

Safanelli, J.L.; Chabrillat, S.; Ben-Dor, E.; Demattê, J.A.M. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. *Remote Sensing* **2020**, *12*, 1369, doi:10.3390/rs12091369.

Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society* **2020**, *146*, 1999–2049, doi:10.1002/qj.3803.

Walkley, A.; Black, A.; An Examination of the Degtjareff Method for Determining Soil Organica Matter, and a Proposed Modification of the Chromic Acid Titration Method. **1934**.

Jenny, H. Factors of Soil Formation. *Soil Science* **1941**, *52*, 415, doi:10.1097/00010694-194111000-00009.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27, doi:10.1016/J.RSE.2017.06.031.

Mathieu, R.; Pouget, M.; Cervelle, B.; Escadafal, R. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid

Environment. *Remote Sensing of Environment* **1998**, *66*, 17–28, doi:10.1016/S0034-4257(98)00030-3.

Thaler, E.A.; Larsen, I.J.; Yu, Q. A New Index for Remote Sensing of Soil Organic Carbon Based Solely on Visible Wavelengths. *Soil Science Society of America Journal* **2019**, *83*, 1443–1450, doi:10.2136/sssaj2018.09.0318.

Xiao, J.; Shen, Y.; Tateishi, R.; Bayaer, W. Development of Topsoil Grain Size Index for Monitoring Desertification in Arid Land Using Remote Sensing. *International Journal of Remote Sensing* **2006**, *27*, 2411–2422, doi:10.1080/01431160600554363.

Segal, D. Theoretical Basis for Differentiation of Ferric-Iron Bearing Minerals, Using Landsat MSS Data. *Proceedings of Symposium for Remote Sensing of Environment, 2nd Thematic Conference on Remote Sensing for Exploratory Geology, Fort Worth, T* **1982**, 949–951.

Rowan, L.C.; Mars, J.C. Lithologic Mapping in the Mountain Pass, California Area Using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Data. *Remote Sensing of Environment* **2003**, *84*, 350–366, doi:10.1016/S0034-4257(02)00127-X.

Douaoui, A.E.K.; Nicolas, H.; Walter, C. Detecting Salinity Hazards within a Semiarid Context by Means of Combining Soil and Remote-Sensing Data. *Geoderma* **2006**, *134*, 217–230, doi:10.1016/J.GEODERMA.2005.10.009.

Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* **2010**, *28*, 281–298, doi:10.1111/j.1538-4632.1996.tb00936.x.

Keskin, H.; Grunwald, S. Regression Kriging as a Workhorse in the Digital Soil Mapper's Toolbox. *Geoderma* **2018**, *326*, 22–41, doi:10.1016/J.GEODERMA.2018.04.004.

45.   Steichen RJRT, T.J.; Cox, N.J. *A Note on the Concordance Correlation Coefficient*; 2002; Vol. 2;.

Friedman, S.P. Soil Properties Influencing Apparent Electrical Conductivity: A Review. *Computers and Electronics in Agriculture* **2005**, *46*, 45–70, doi:10.1016/J.COMPAG.2004.11.001.

Sudduth, K.A.; Kitchen, N.R.; Bollero, G.A.; Bullock, D.G.; Wiebold, W.J. *Comparison of Electromagnetic Induction and Direct Sensing of Soil Electrical Conductivity*;

Corwin, D.L.; Plant, R.E. Applications of Apparent Soil Electrical Conductivity in Precision Agriculture. *Computers and Electronics in Agriculture* **2005**, *46*, 1–10, doi:10.1016/J.COMPAG.2004.10.004.

Demattê, J.A.M.; Campos, R.C.; Alves, M.C.; Fiorio, P.R.; Nanni, M.R. Visible–NIR Reflectance: A New Approach on Soil Evaluation. *Geoderma* **2004**, *121*, 95–112, doi:10.1016/J.GEODERMA.2003.09.012.

Sovdat, B.; Kadunc, M.; Batič, M.; Milčinski, G. Natural Color Representation of Sentinel-2 Data. *Remote Sensing of Environment* **2019**, *225*, 392–402, doi:10.1016/J.RSE.2019.01.036.

Shepherd, G.; Stagnari, F.; Pisante, M.; Benites, J. *Visual Soil Assesment (VSA) Field Guides*; FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS (FAO), Ed.; Rome, Italy, 2008; Vol. Annual Crops.;

Yang, S.H.; Liu, F.; Song, X.D.; Lu, Y.Y.; Li, D.C.; Zhao, Y.G.; Zhang, G.L. Mapping Topsoil Electrical Conductivity by a Mixed Geographically Weighted Regression Kriging: A Case Study in the Heihe River Basin, Northwest China. *Ecological Indicators* **2019**, *102*, 252–264, doi:10.1016/j.ecolind.2019.02.038.

Wu, C.; Liu, G.; Huang, C. Prediction of Soil Salinity in the Yellow River Delta Using Geographically Weighted Regression. *Archives of Agronomy and Soil Science* **2017**, *63*, 928–941, doi:10.1080/03650340.2016.1249475.

McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On digital soil mapping, *Geoderma, Volume 117, Issues 1–2*, **2003**, 3-52.

CHAPTER II: Soil Organic Carbon variability and sampling optimization

1.   Introduction

Soil organic carbon (SOC) concentration is a valuable soil property to guide agricultural applications of chemical inputs (Ladoni et al. n.d.). Organic carbon in soil serves many purposes in soil fertility and structure by improving water retention and infiltration, promoting soil organism growth, and by holding essential nutrients in the soil for healthy plant growth and production (Oades n.d.). The knowledge of soil physical and mechanical properties and the spatial variability of these properties are essential concepts in Precision Agriculture (PA) Iticha and Takele (2019).

Moreover, the soil system is a significant terrestrial carbon sink; estimates for the top meter of soil in the world range between 1200 and 2500 petagrams for organic carbon (Lal 2005) . Sustainable agriculture can contrast climate change, capturing and storing carbon emissions in soil (Farooq et al. 2019; Lugato et al. 2006; Pisante et al. 2012; Pisante et al. 2015). One mechanism that can facilitate the effective management of soil carbon is to treat it as a tradeable resource or commodity (de Gruijter et al. 2016). The development of carbon credit markets, that are accessible to private sectors would allow incentives such as government payments, tax credits, and emissions trading, which can aid in overcoming farmer reluctance to adopt management strategies that increase soil carbon (de Gruijter et al. 2016). There is an essential demand for fast and inexpensive methods for SOC mapping for these reasons.

Digital Soil Mapping (DSM) is widely used for mapping SOC (Iticha and Takele 2019). DSM aims to spatially predict soil properties by combining soil observation with auxiliary information such as remote sensing images (Costa et al. 2018; Kumar et al. 2013; Mishra et al. 2010; Wang et al. 2018; Zhang et al. 2011; Zhang et al. 2018). In particular, the main aim is to estimate a value of a pedological variable at an unstapled point (Szatmári et al. 2015). Therefore, sampling design is a critical stage of DSM (Brus 2019), and it can be divided into two strategies: design-based or model-based (Brus and de Gruijter 1997; Brus 2019). In a design-based approach, sampling locations are selected by probability sampling, and the statistical inference is based on the sampling design. In the design-based approach, units are selected by probability sampling. Estimates are based on the selection probabilities of the sampling units as determined by the sampling design (design-based inference) Brus (2019). In a model-based approach, a stochastic model is used in estimation. As the model already contains a random error term, a probability opens up the possibility of optimized non-probability sampling (Brus 2019).

A sampling of soil properties at within-field scales may involve considerable costs (travel, labor time, extraction of soil cores, laboratory analyses, etc.); therefore, for making PA cost-effective, these costs must be minimized. This economic constrain generally results in low sampling density, which is the primary recognized factor limiting PA efficacy.

One of the most used methods for sampling optimization is Spatial Simulated Annealing (SSA) van Groenigen (1997). SSA is an annealing-based algorithm to optimize sampling schemes, using an annealing-based algorithm to optimize sampling schemes on a continuous solution space for different quantitative optimization criteria using auxiliary information (van Groenigen 1997). Literature has shown that Soil Electrical Conductivity (ECa) Corwin and Plant (2005); Corwin and Scudiero (2020) can be a helpful tool for sampling optimization (Barca et al. 2015). ECa is a good proxy of soil variability, but as reported in chapter 1, ECa has several limitations. Therefore, the rapid expansion of PA required an alternative, faster way to map spatial field variability. Multi-temporal images with bare soil pixels (Synthetic Soil Image, SYSI) Demattê et al. (2018), obtained from the time series of Sentinel-2 satellite (Silvero et al. 2021), can be a valid alternative to ECa, understood as a proxy of soil variability(Safanelli et al. 2020a).

This study aims: a) to create a spatial model prediction of SOC in an agricultural field using some covariates provided by SYSI; b) to propose a method for optimizing the sampling scheme for the SOC assessment using the model provided from point as auxiliary information.

2.   Materials and Methods

2.1.   Study area

The study area is located in the deltaic plain of Po River, near Massa Fiscaglia (44° 79' N; 11°98' E), in Northern Italy. The study site area is composed of a field (ID 19), as reported in Figure 1, with an extension of approximately 12 ha.

The climate of this area is temperate. According to ERA 5 LAND (Hersbach et al. 2020), in the last 40 years (1981- 2021), the yearly average precipitation is 770 mm; yearly average air temperature is 14.61 °C, yearly average minimum air temperature 0.43 °C, and yearly average maximum air temperature 28.79 °C.

2.2. Soil Sampling

A Soil sampling was carried to determine the predominant soil properties in August 2020. A grid of 122 points, 61 according to a regular grid and 61 chosen randomly (Fig. 1), was used to take the sample from topsoil (0-30 cm). Soil sampling was carried out with no knowledge of soil variability. The coordinates of each sampling site were determined with a GPS model SP20 handheld GNSS (Spectra Geospatial), and a private laboratory carried out the soil analysis.

Organic Carbon was determined by dichromate oxidation (Walkley and Black 1934).
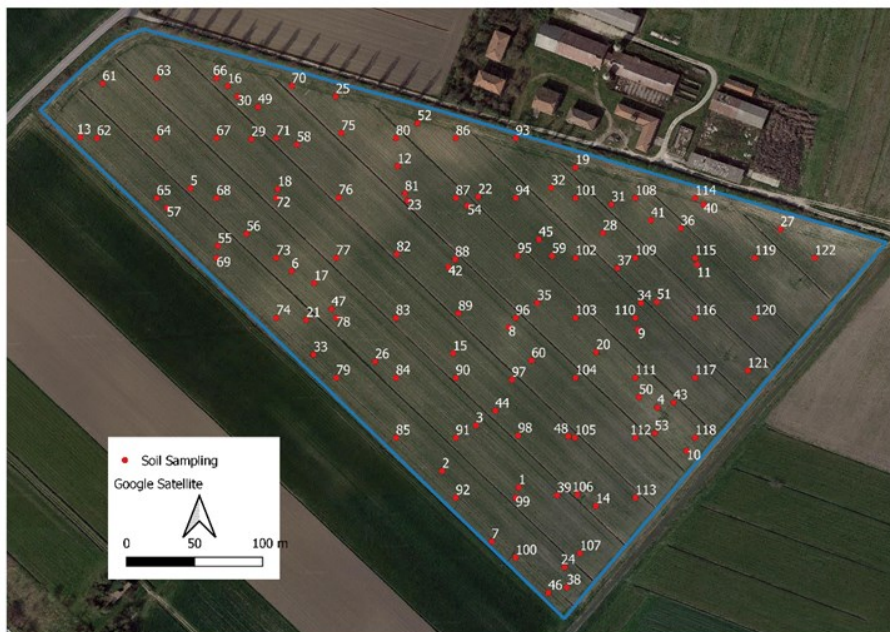
Figure 13. Soil sampling grid.

2.3. Satellite data and multi-temporal images

Sentinel2-MSI satellite (Drusch et al. 2012) data, as surface reflectance products, were used for this work. The Sentinel-2 MultiSpectral Instrument (S2-MSI) mission was launched by the European Space Agency (ESA) in 2015. Two satellites (2A and 2B) were placed in orbit, provide five -day revisit time. Sentinel2-MSI has spatial resolutions ranging from 10 to 60 meters and thirteen spectral bands that covet the visible, near-infrared, red-edge, shortwave-infrared and water and cirrus region. A multi-temporal bare soil image named SYSI was obtained by applying the GEOS-3 (Demattê et al. 2018) using Sentinel-2 satellite data (Silvero et al. 2021) on the Google Earth Engine platform (GEE) Gorelick et al. (2017).

One hundred fourteen images from 2018-2021 with at least 80% of the area without clouds and shadows were used. The GEOS3 method consists of (1) obtaining the images, (2) masking pixels that were non-bare soil, and (3) calculating the median reflectance of the masked images.

After obtaining Sentinel 2 images, NDVI (Normalized Difference Vegetation Index) and NBR2 (Normalized Burn Ratio 2) were calculated and used to mask the images.

Following the GEOS method, the range between 0 and 0.25 for NDVI was considered bare soil, and the values above this threshold were flagged as N.A.

NBR2 values > 0.075 were also flagged as N.A. and corresponded to sites covered with straw or burned. For each image, quality masked was used to exclude areas with clouds and shadows. Finally, the masked images were ordered by date and the median reflectance was calculated.

Form SYSI VIS-NIR bands (Table 1) were extracted with a spatial pixel resolution of 10 meters, and then several bare soil indices (Fig. 2) were calculated as reported in Table 2.

Sentinel -2 Bands	Central wavelength (nm)
Band 2 (B2) - Blue	490
Band 3 (B3) - Green	560
Band 4 (B4) - Red	665
Band 8 (B8) - NIR	842

Table 17. Sentinel-2 Bands.

Index	Formula	Soil properties (reference)
Brightness Index (BI)	$(B4^2 + B3^2 + B2^2/3)^{0.5}$	Average reflectance Magnitude (Mathieu et al. 1998b)
Saturation Index (SI)	(B4-B2)/(B4+B2)	Spectra slope (Mathieu et al. 1998b)
Hue Index (HI)	(2*B4-B3-B2)/(B3-B2)	Primary colours (Mathieu et al. 1998b)
Coloration Index (CI)	(B4-B3)/(B4+B3)	Soil Colour (Mathieu et al. 1998b)
Soil Background Line (SBL)	(B8 –(B4*1.24))	(Weiss et al. 2020)
SOCI (Soil Organic Carbon Index)	(B2/(B4*B3))	Organic Carbon(Thaler et al. 2019)
Grain Size Index (GSI)	(B4-B2)/(B4+B3+B2)	Soil Texture (Xiao et al. 2006)
Iron Oxide Ratio (IOR)	(B4/B2)	Irone Oxide soil content (Segal 1982)
Ferric Iron 3+ Index	(B4/B3)	Iron 3+ soil content (Rowan and Mars 2003)
Salt Index (SIndx)	$\sqrt{(B4*B2)}$	Salt soil content (Douaoui et al. 2006)
Normalize Salt Index (NSI)	(B4-B8)/(B4+B8)	Salt soil content (Douaoui et al. 2006)
Salinity Index 2	$\sqrt{(B3 + B4 + B8)^2}$	Salt Soil content (Douaoui et al. 2006)

Table 18. Bare soil indices.
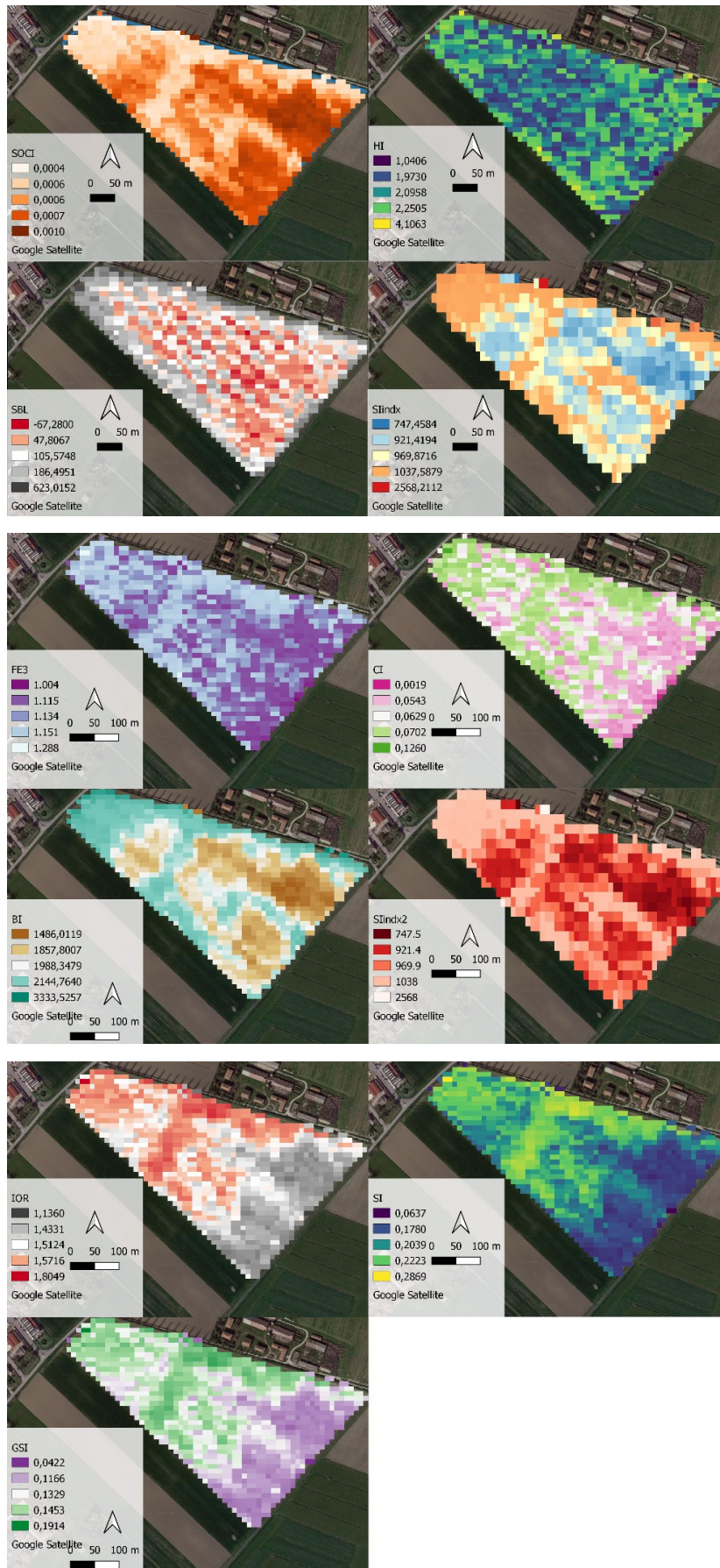
Figure 2. Maps of bare soil indices.

2.4.   SOC Prediction models

Focusing on the aim of SOC prediction, in this study, several statistical models were compared. In particular were evaluated the performance of Multiple Linear Regression (MLR), Geographically Weighted Regression (GWR), Generalized Additive Model (GAM) and Multivariate Adaptive Regression Spline (MARS), that are widely used in DSM (Lamichhane et al. 2019).

2.4.1. Multiple Linear Regression (MLR)

MLR extends simple linear regression. Multiple Linear Regression can be written as:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon_i$$

Where y is the predicted ECa, k is the number of covariates, β_i is regression coefficient, x_i is the covariate and ε_i represents the error term. MLR was carried out in R environment.

2.4.2. Geographically Weighted Regression (GWR)

GWR is an extension of the traditional regression (Brunsdon et al. 2010). In traditional regression, it is assumed that the model coefficients are constant across the study area, but this is not necessarily true in many situations. Therefore, in the GWR, the regression coefficients are specific to a location rather than global estimates. The GWR methods can be represented as:

$$y(x,y) = \beta_0(x,y) + \sum_{i=1}^{k} \beta_i(x,y)x_i + \varepsilon_i(x,y)$$

in which y(x, y) is the predicted SOC in the i-th site, (x,y) are the coordinates for the i site, k is the number of covariates, β_i is regression coefficient, x_i is the covariate at the site i and ε_i represents the error term. The bandwidth chosen in GWR is a key affecting the regression results. The Adaptive (AICc) methods were used to choose the best bandwidth in this study. For adaptive kernel size, the weight of each point can be calculated using the Gaussian function. GWR was carried out in R environment using GWmodel library.

2.4.3. Generalized Additive Model (GAM)

GAM (Hastie and Tibshirani 2017)is a generalization of linear regression model, in which the coefficients can be expanded as smooth functions of covariates. They are semi-parametric and can account for non-linear relationships between dependent variables and covariates.

$$E(Y|X_1, X_2, \ldots, X_P) = \propto + f_1 X_1 + f_2 X_2 + \cdots + f_p X_p$$

Where Y is the dependent variable, $X_1, X_2, \ldots, X_P$ reperesent the covariates and the f_i are the smoot functions. As for generalized linear models, the GAM approach specifies a distribution for the conditional mean $\mu(Y)$ along with a link function g relating the latter to an addictive function of the covariates.

$$g[\mu(Y)] = \propto + f_1 X_1 + f_2 X_2 + \cdots + f_p X_p$$

2.4.4. Multivariate Adaptive Regression Spline (MARS)

MARS is a non-parametric model. MARS combines the classical linear regression, the mathematical construction of splines, the binary recursive partitioning, and brute search, and intelligent algorithms to produce a model to predict the value of a target variable from a set of independent variables. MARS general expression can be written as follows:

$$y = f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x)$$

Where y is the value predicted by the model using a function f(x), which can be decomposed into an initial constant β_0 and a sum of M terms, each of them are formed by a coefficient β_m and basic functions h_m (x). The basic function (BF) represents the information included in one or more independent variables and is selected step-by-step.

2.5. Model evaluation

Using Caret library in R environment, all sample points were divided into two classes for calibration (80% n=98) and validation (20% n= 24).

The agreement between predictions and observations is generally carried out using Pearson Correlation or Spearman indices. According to Steichen and Cox (2002), Lin's concordance coefficient is preferred to quantify the goodness of model adaptation. Lin's concordance coefficient provides a measure of overall accuracy that takes into account both bias correction and precision. The formula as follows:

$$\rho_c = \frac{2 \, r \, s_x \, s_y}{s_Y^2 + s_X^2 + (m_{y-}m_x)^2}$$

Where r is the Pearson's coefficient; s_x s_y are the standard deviation of the true and predicted values; m_y and m_x are the means; and s_Y^2 and s_x^2 are the variances of the true and predicted values.

Furthermore, Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are used:

$$RMSE = \sqrt{(\sum_{i=1}^{n}[D(x_i)] - D(x_i)]^2}$$

$$MAE = \frac{\sum_{i=1}^{n}[D(x_i)] - D(x_i)]}{n}$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{n} \left| \frac{(D(x_i) - D*(x_i))}{D(x_i)} \right|$$

$D (x_i)$ is the measurement of ECa, $D (x_i)$ is the predicted ECa, and n is the number of validation sites, respectively.

2.6. Model Preprocessing

Pearson correlation analysis was conducted to select the environmental covariates significantly correlating with SOC. Then, a model for SOC was constructed by stepwise multiple regression using Akaike information criterion (AIC) as a selection criterion.

After removing collinearity using multiple stepwise regression, the remaining variables were selected for modelling. Finally, latitude and longitude maps were produces in a 10-m resolution as spatial position predictor.

2.7. Soil sampling reduction

A soil sampling reduction was carried out using the average ordinary kriging variance (MUKV) criteria (van Groenigen et al. 1999). Optimization of the sampling scheme allows reducing the number of sampling points without decreasing the accuracy of the investigated attribute (Barca et al. 2015). The objective of MUKV is to find the sample pattern with the smallest Mean Universal Kriging Variance. Theoretically, a sample scheme reduction needs to evaluate the MUKV for all combinations of sample points and select the one with the smallest value.

The soil sampling reduction was carried out using the function ssaOptim of the intamapInteractive R library. This function applies a Spatial Simulated Annealing search algorithm (Barca et al. 2015;

Corwin and Plant 2005; Demattê et al. 2018). Simulated annealing is an iterative, combinatorial optimization algorithm in which a sequence of combinations is generated by deriving a new combination from slightly and randomly changing the previous combination. Each time a new combination is generated, the quality measure (i.e., the MUKV) is evaluated and compared with the value of the previous combination. The new combination is accepted if the change has improved quality measures. SSA with MUKV makes it possible to assume a linear relationship between independent and dependent variables at observation locations using Regression Kriging (RK).

RK is one of the most popular spatial interpolation techniques in digital soil mapping (Brus and Heuvelink 2007; Keskin and Grunwald 2018). RK is a combination of the traditional MLR and kriging. RK can be expressed as:

$$y_{mlr\,(x,y)} = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon_i$$

$$y_{rk} = y_{mlr(x,y)} + \sum_{i=1}^{k} w_i(x,y)\,\varepsilon_i(x,y)$$

The covariates from the SOC prediction model (2.3 and 2.4) have been used in SSA. The ssaOptim allows removing sample points that provide redundant information, leaving only the point that reduces the MUKV. Then an iterative function was implemented in R environment to remove the stochasticity of SSA, integrating MUKV with the Moran's I (Anselin and Rey 2010). Moran's I allows the investigation of correlation within a single variable due to the spatial relationship amongst its observations. The weights (w_{ij}) are a function of the distance between each pair of observations of the variable under study (x_i; x_j). In its simplest form, weights will take values 1 for close neighbors, or otherwise 0 (Safanelli et al. 2020a).

Moran's I statistics is represented by the following equations:

$$I = \frac{N}{S_o} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{i,j},\, z_i z_j}{\sum_{j=1}^{N} z_1^2}$$

$$S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{i,j}$$

where z_i is the deviation of an attribute from its mean ($x_i - x$) and S_0 is the aggregate of all the spatial weights $\omega_{i,j}$. The null hypothesis for the test is that the data is randomly distributed, which means that the data does not have a spatial structure.

The function iteratively removes one point at a time, as long as the spatial structure exists, starting from the initial number of the sampling scheme. The removing operation stops when the null

hypothesis is accepted. This function is iterated for ten loops, then the sampling scheme optimized with the higher value of I Moran was selected. Finally, the remaining points were interpolated using RK, producing a new map (SOC Optimized Map – SOCom).

2.8.   Maps comparison

The similarity between SOC map and SOCom was evaluated using Cross-Correlogram. The cross-correlogram, which measures the correlation as a function of the distance between observations, is particularly well-suited to account for the inherently spatial characteristics of map representation. The analytical formulation of the cross-correlogram is the following:

$$r_{A,B}(h) = \frac{\sum_{i,j=1}^{N(h)} z_{i,jA} \, z_j{}', j'_B - \widehat{mA} \, \widehat{mB}}{\widehat{sA} \, \widehat{sB}}$$

To compute $r_{A,B}(h)$ the procedure is the following: from both the maps, all the couples whose locations are separated by the distance h are collected. Indices $\widehat{mA}$, $\widehat{mB}$ and $\widehat{sA} \, \widehat{sB}$ represent the mean and the standard deviation of mapped $z_{i,jA,}$ and $z_{i,jB}$ respectively. $N(h)$ is the total number of these pairs. For summarizing the results and comparing different outcomes, results of cross-correlation containing correlations at specified lag distances (0, 25 m, 50 m, 75 m, 100 m).

3.   Results and Discussion

3.1.   Descriptive statistics of SOC

Table 3 provides a statistical summary of the entire SOC data set of 120 observations after outliers removing. The observed SOC varies from 1.16 to 4.38 g/kg, which means significant SOC variations in the study area. The skewness value is 0.78, and kurtosis is 1.49, indicating that the SOC data are not normally distributed.

The Shapiro-Wilk normality test confirmed that hypothesis producing a significant coefficient (p-value = 0.00).

N	Min.	Max.	Mean	STD	Skew	Kurt
120	1.16	4.38	2.25	0.54	0.78	1.49

Table 19. Summary Statistics of SOM Sample Data.

N = Number of samples; Min. = Minimum; Max. = Maximum; STD = Standard Deviation;

Skew = Skewness; Kurt = Kurtosis.

3.2. Correlation between SOC and SYSI

The Pearson correlation coefficient (r) among SOC with VIS-NIR SYSI bands was highly significant ($p < 0.05$) except with B2, as reported in Table 4. According to the literature (Safanelli et al. 2020a; Silvero et al. 2021), these results confirmed that the SYSI image could be a good proxy of soil variability (Chapter 1). The longer time frame of the full SYSI provides a more stable median reflectance that is less affected by dynamic effects of the bare soils (Castellini et al. 2019) .

The most important spectral features for SOC prediction are located in the VIS region at 450, 590, and 664 nm, and very close to the S2 bands (Table 1) in this spectral range (490, 560 and 665 nm) (Drusch et al. 2012). The main chemical components in soils that interact with electromagnetic radiation across the Vis-NIR also termed "chromophores" are OH in free water and clay mineral lattice, organic matter, and nonclay minerals, such as iron oxides, carbonates, and salts (Safanelli et al. 2020a). A significant absorption centered on 665 nm can be attributed to organic matter. Therefore, soil reflectance decreases with organic matter and water content (Gorelick et al. 2017). Concerning bare soil indices (Table 5), in this section, only indices with a correlation coefficient (r) greater than 0.60 have been taken into account. Soil Organic Carbon Index (SOCI) shows a good correlation with SOC ($r = 0.60$). SOCI effectively estimates SOC because it tracks changes in the reflectance of the red and green wavelengths, which have been demonstrated to be relevant for SOC estimations (Thaler et al. 2019). Grain Size Index (GSI) is negatively correlated with SOC, it is related to fine sand content of topsoil (Xiao et al. 2006). A negative correlation can be explained considering that the reflectance decreases as the content of clay and silt grains increase, and inversely, the reflectance increases as the total fine sand content increases in the topsoil (Xiao et al. 2006). Moreover, SOC correlates with Iron Oxide Ratio (IOR) is related to soil iron oxide and phyllosilicates (Mathieu et al. 1998b). The correlation between the sorption of organic carbon and the content of iron oxide minerals indicates the importance of iron oxides in the accumulation and stabilization of organic carbon in soil [29]. Finally, SI, a bare soil index that is correlated to SOC: soil reflectance curves are typically monotonous and increasing in the visible range (Mathieu et al. 1998b). The SI measures the general slope of spectra, from red to blue wavelength or, in other words, its deviation relative to a flat spectra characteristic of a neutral color like grey (Mathieu et al. 1998b). It knows that soil color is related to the soil organic matter content; in general, the darker the color is, the greater is the amount of organic matter in the soil (Thaler et al. 2019) .

	r	p
SOC – B2		
SOC - B3	-0.47	0.00
SOC – B4	-0.52	0.00
SOC – B8	-0.50	0.00

Table 20. The Pearson correlation coefficient (*r*) among SOC and VIS-NIR SYSI bands.

	r	p
SOCI - SOC	0.60	0.00
HI-SOC		
NSI- SOC		
SBL- SOC	- 0.19	0.03
SIindx - SOC	- 0.32	0.00
FE3- SOC	- 0.51	0.00
CI- SOC	- 0.51	0.00
BI - SOC	-0.52	0.00
SIindx2- SOC	- 0.52	0.00
IOR- SOC	- 0.66	0.00
SI- SOC	- 0.66	0.00
GSI- SOC	- 0.66	0.00

Table 21. The Pearson correlation coefficient (*r*) among SOC and Bare Soil Indices.

3.3. Covariates for regression models

The correlations between environmental covariates were strong, which may cause the collinearity problem. These factors were removed by stepwise regression. Therefore, forward stepwise multiple regression was performed, and only environmental covariates in which contribution to regression was significant ($p < 0.10$) were retained in the regression equation. The covariates selected are:

$$SOC \sim Latitude + Longitude + HI + CI + BI + GSI$$

BI (Brightness Index), HI (Hue Index), and CI (Coloration Index) are colour indices. BI, gives a measure of the soil reflectance magnitude, HI together SI characterized its shape (Mathieu et al. 1998b). HI accounts for the relative proportion of the three primaries colours, red, green, and blue

(Mathieu et al. 1998b). In literature, BI is related to colour intensity while HI is related to the amount of colour, according to the Munsel system (Mathieu et al. 1998b).

CI corresponds to the normalized ratio of red and green bands. In this spectral range, soil reflectance curves are mainly affected by the absorption of iron oxides like goethite and hematite (Mathieu et al. 1998b). The correlation between SOC – soil colour and GSI, have already been explained in section 3.2.

3.4. Comparison of the performance of models

The performance of four models was evaluated on the training data for the coefficient of determination, while Lin's concordance coefficient and RMSE, MAE, MAPE on the test data. Evaluation results of the estimation are summarized in Table 6.

Model	R^2	r	Lin's accuracy	RMSE	MAE	MAPE
MLR	0.63	0.75	0.97	0.350	0.300	14.85
GWR	0.78	0.89	0.98	0.260	0.200	9.26
GAM	0.79	0.77	0.95	0.418	0.300	13.95
MARS	0.77	0.80	0.97	0.330	0.255	12.16

Table 22. Lin's concordance coefficient and RMSE, MAE, MAPE.

GAM produced the highest value of determination coefficient (0.79), slightly higher than GWR and GAM (0.79, 0.77), that means that using bare soil indices, the models can explain about 78% of the total variance of SOC. However, GWR shows the highest value of r (0.89) and low values of RMSE (0.26), MAE (0.25), and MAPE (9.26). Therefore, GWR was chosen as the most performing model. Using SYSI indices as a covariate for SOC prediction shows better performance than in literature (Hersbach et al. 2020) , probably due to the different scales of application.

Concerning the best performance of GWR, this result is according to other studies (Douaoui et al. 2006; Lamichhane et al. 2019; Rowan and Mars 2003; Segal 1982; Xiao et al. 2006). The GWR performance can be explained because the GWR makes predictions using weights by neighboring samples, whereas the MLR, GAM, and MARS are global models. Therefore, regression parameters estimated in the GWR analysis are not fixed across space (Fig. 3), which means that the relationships between topsoil SOC and some environmental covariates are spatially non-stationary (Table 7).

	Min.	1st Qu	Median	3rd Qu	Max
Intercept	-5,469.89	7,214.51	10,988.52	14,760.35	25,240.84
Latitude	-0.00	-0.00	-0.00	-0.00	-0.00
Longitude	-0.00	-0.00	-0.00	-0.00	-0.00
HI	-3.92	-1.80	-0.84	0.37	1.72
CI	-53.61	-9.52	21.76	45.73	115.69
BI	-0.00	-0.00	-0.00	-0.00	-0.00
GSI	-73.51	-30.26	-21.16	-7.05	22.66

Table 23. Regression parameters estimated in the GWR analysis.
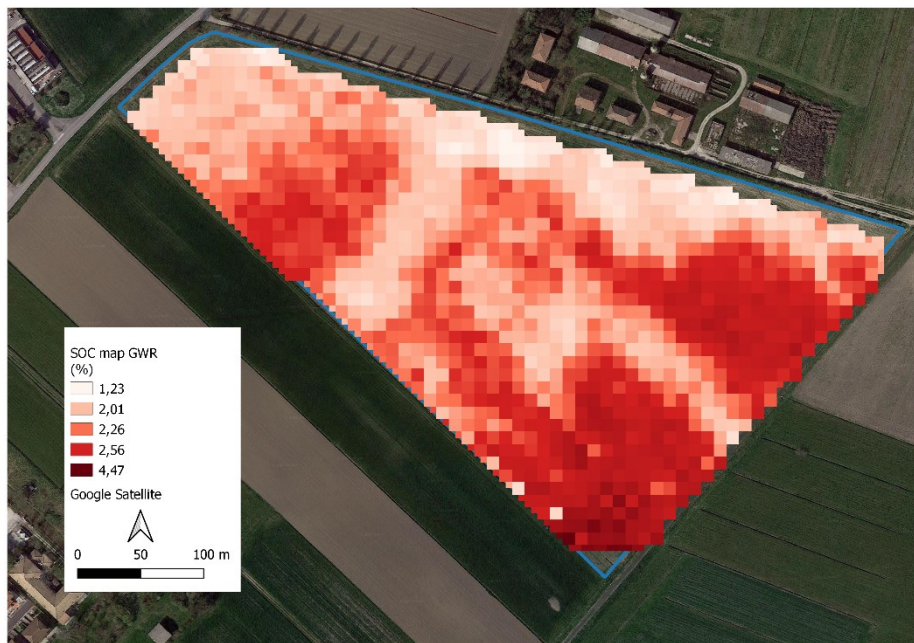
Figure 3. SOC map by GWR.

3.5. Soil sampling reduction

The results of sampling reduction are summarized in Table 8, and each loop is described from tables 9-18. Among the optimized sampling schemes, loop number 9 shows the highest value of Moran I (0.47). Therefore, it was chosen how to sample scheme (Fig. 4).

Loop	Remaing points	Moran I	p
1	25	0.38	0.00
2	30	0.43	0.00
3	35	0.33	0.00
4	35	0.40	0.00
5	35	0.26	0.02
6	30	0.31	0.00
7	20	0.47	0.01
8	20	0.32	0.01
9	35	0.47	0.00
10	30	0.29*	0.01

Table 24. Loop of sample scheme reducing.

Figure 4. Sampling scheme optimized.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.57	0.00
105	15	0.58	0.00
100	20	0.57	0.00
95	25	0.53	0.00
90	30	0.54	0.00
85	35	0.57	0.00
80	40	0.60	0.00
75	45	0.61	0.00
70	50	0.58	0.00
65	55	0.56	0.00
60	60	0.63	0.00
55	65	0.55	0.00
50	70	0.30	0.00
45	75	0.41	0.00
40	80	0.26	0.02
35	85	0.21	0.04
30	90	0.37	0.00
25	95	0.38	0.03
20	100	0.07	0.23

Table 25. Loop 1 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.57	0.00
105	15	0.56	0.00
100	20	0.53	0.00
95	25	0.58	0.00
90	30	0.57	0.00
85	35	0.57	0.00
80	40	0.57	0.00
75	45	0.59	0.00
70	50	0.57	0.00
65	55	0.63	0.00
60	60	0.44	0.00
55	65	0.38	0.00
50	70	0.32	0.00
45	75	0.34	0.00
40	80	0.39	0.00
35	85	0.36	0.01
30	90	0.42	0.00
25	95	0.15	0.17

Table 26. Loop 2 of sample scheme reducing.

Remaining points	Points deleted	Moran I	*p*
115	5	0.57	0.00
110	10	0.57	0.00
105	15	0.56	0.00
100	20	0.56	0.00
95	25	0.58	0.00
90	30	0.56	0.00
85	35	0.55	0.00
80	40	0.56	0.00
75	45	0.57	0.00
70	50	0.57	0.00
65	55	0.62	0.00
60	60	0.57	0.00
55	65	0.36	0.00
50	70	0.56	0.00
45	75	0.46	0.00
40	80	0.46	0.00
35	85	0.32	0.00
30	90	0.20	0.10

Table 27. Loop 3 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.57	0.00
105	15	0.57	0.00
100	20	0.55	0.00
95	25	0.54	0.00
90	30	0.57	0.00
85	35	0.58	0.00
80	40	0.58	0.00
75	45	0.55	0.00
70	50	0.53	0.00
65	55	0.60	0.00
60	60	0.61	0.00
55	65	0.32	0.00
50	70	0.39	0.00
45	75	0.48	0.00
40	80	0.41	0.00
35	85	0.40	0.00
30	90	0.27	0.09

Table 28. Loop 4 of sample scheme reducing.

Remaining points	Points deleted	Moran I	*p*
115	5	0.58	0.00
110	10	0.58	0.00
105	15	0.54	0.00
100	20	0.56	0.00
95	25	0.55	0.00
90	30	0.55	0.00
85	35	0.54	0.00
80	40	0.55	0.00
75	45	0.54	0.00
70	50	0.55	0.00
65	55	0.61	0.00
60	60	0.63	0.00
55	65	0.56	0.00
50	70	0.54	0.00
45	75	0.51	0.00
40	80	0.37	0.00
35	85	0.26	0.02
30	90	0.09	0.24

Table 29. Loop 5 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.56	0.00
105	15	0.58	0.00
100	20	0.57	0.00
95	25	0.58	0.00
90	30	0.59	0.00
85	35	0.55	0.00
80	40	0.63	0.00
75	45	0.49	0.00
70	50	0.56	0.00
65	55	0.61	0.00
60	60	0.49	0.00
55	65	0.50	0.00
50	70	0.34	0.00
45	75	0.53	0.00
40	80	0.36	0.00
35	85	0.31	0.00
30	90	0.10	0.25

Table 30. Loop 6 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.59	0.00
105	15	0.59	0.00
100	20	0.56	0.00
95	25	0.56	0.00
90	30	0.57	0.00
85	35	0.60	0.00
80	40	0.56	0.00
75	45	0.57	0.00
70	50	0.47	0.00
65	55	0.54	0.00
60	60	0.70	0.00
55	65	0.51	0.00
50	70	0.62	0.00
45	75	0.54	0.00
40	80	0.44	0.00
35	85	0.30	0.00
30	90	0.28	0.00
25	85	0.44	0.00
20	80	0.40	0.01
15	75	0.20	0.10

Table 31. Loop 7 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.59	0.00
105	15	0.59	0.00
100	20	0.56	0.00
95	25	0.56	0.00
90	30	0.57	0.00
85	35	0.60	0.00
80	40	0.56	0.00
75	45	0.57	0.00
70	50	0.47	0.00
65	55	0.54	0.00
60	60	0.70	0.00
55	65	0.51	0.00
50	70	0.62	0.00
45	75	0.54	0.00
40	80	0.44	0.00
35	85	0.30	0.00
30	90	0.28	0.00
25	85	0.44	0.00
20	80	0.40	0.01
15	75	0.20	0.10

Table 32. Loop 8 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.57	0.00
105	15	0.56	0.00
100	20	0.56	0.00
95	25	0.54	0.00
90	30	0.58	0.00
85	35	0.59	0.00
80	40	0.55	0.00
75	45	0.51	0.00
70	50	0.62	0.00
65	55	0.56	0.00
60	60	0.52	0.00
55	65	0.37	0.00
50	70	0.48	0.00
45	75	0.50	0.00
40	80	0.42	0.00
35	85	0.47	0.00
30	90	0.18	0.14

Table 33. Loop 9 of sample scheme reducing.

Remaining points	Points deleted	Moran I	p
115	5	0.57	0.00
110	10	0.57	0.00
105	15	0.58	0.00
100	20	0.55	0.00
95	25	0.58	0.00
90	30	0.58	0.00
85	35	0.55	0.00
80	40	0.54	0.00
75	45	0.57	0.00
70	50	0.53	0.00
65	55	0.48	0.00
60	60	0.51	0.00
55	65	0.44	0.00
50	70	0.47	0.00
45	75	0.45	0.00
40	80	0.29	0.00
35	85	0.14	0.14

Table 34. Loop 10 of sample scheme reducing.

The semi-variogram of SOC, concerning the point of sample scheme optimization, is reported in Figure 5. The semi-variogram showed a range of influence of 66 meters and Sill of 0.13. The model of Semi-variogram has been the Spherical model. The predictive map of RK is reported in Figure 6.
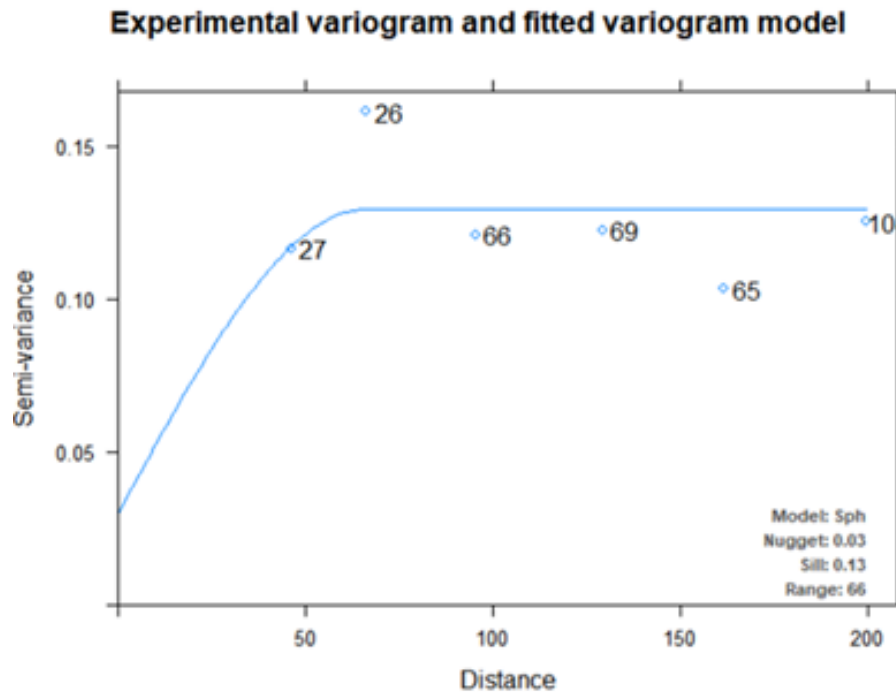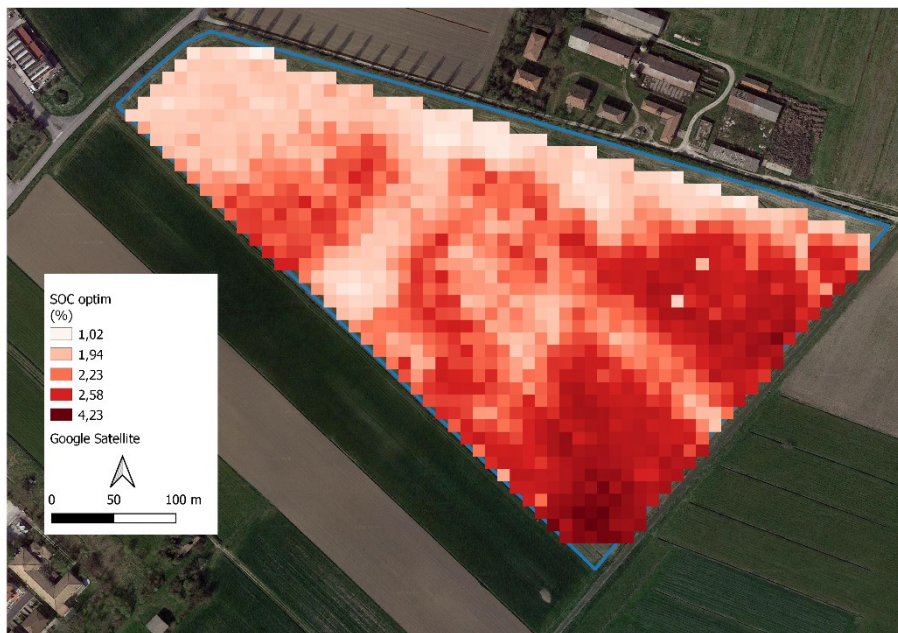
Figure 5. Semi-variogram of SOC.

Figure 6. SOC optim map.

3.6.   Maps comparison

Cross-correlograms were computed and cross-correlation coefficients at specific lags (0 m, 25 m, 75 m, 100 m) between the two maps are extracted (Table 19). SOC optim map was strongly correlated with SOC original map (0.80). The two maps remain correlated until 25 meters.

Meters	Cross-correlation
0	0.80
25	0.59
75 100	0.16

Table 35. Cross-correlation coefficients computed at different lags.

Literature has shown how using SSA and Soil electrical conductivity (ECa) can optimize sampling optimization (Barca et al. 2015; Brunsdon et al. 2010). ECa highlights numerous limitations (Chapter 1); the relation between SYSI and ECa has been fully explained in Chapter 1. This study provides a preliminary analysis of SYSI in sampling optimization. Using SISY allows a reduction of sample points by about 75%.

4.   Conclusions

The study and management of the Soil Organic Carbon variability are essential for a regenerative and sustainable agriculture. The rapid expansion of Precision Agriculture requires rapid and inexpensive methods for SOC mapping. This study confirmed that bare soil indices provided by SYSI can explain SOC variability. Moreover, this study provides a preliminary analysis of SYSI in sampling optimization.

5.   References

Ladoni, M.; Hosein, A.E.; Ae, A.B.; Kazem, S.; Ae, A.; Norouzi, A.A. Estimating Soil Organic Carbon from Soil Reflectance: A Review., doi:10.1007/s11119-009-9123-3.

Oades, J.M. *Soil Organic Matter and Structural Stability: Mechanisms and Implications for Management*;

Iticha, B.; Takele, C. Digital Soil Mapping for Site-Specific Management of Soils. *Geoderma* **2019**, *351*, 85–91, doi:10.1016/J.GEODERMA.2019.05.026.

4.   Lal, R. *Soil Carbon Sequestration Impacts on Global Climate Change and Food Security*; 2005;

Farooq, M.; Rehman, A.; Pisante, M. Sustainable Agriculture and Food Security. In *Innovations in Sustainable Agriculture*; Springer International Publishing: Cham, 2019; pp. 3–24.

Pisante, M.; Stagnari, F.; Grant, C.A. Agricultural Innovations for Sustainable Crop Production Intensification. *Italian Journal of Agronomy* **2012**, *7*, 40, doi:10.4081/ija.2012.e40.

Pisante, M.; Stagnari, F.; Acutis, M.; Bindi, M.; Brilli, L.; di Stefano, V.; Carozzi, M. Conservation Agriculture and Climate Change. In *Conservation Agriculture*; Springer International Publishing: Cham, 2015; pp. 579–620.

Lugato, E.; Berti, A.; Giardini, L. Soil Organic Carbon (SOC) Dynamics with and without Residue Incorporation in Relation to Different Nitrogen Fertilisation Rates. *Geoderma* **2006**, *135*, 315–321, doi:10.1016/J.GEODERMA.2006.01.012.

de Gruijter, J.J.; McBratney, A.B.; Minasny, B.; Wheeler, I.; Malone, B.P.; Stockmann, U. Farm-Scale Soil Carbon Auditing. *Geoderma* **2016**, *265*, 120–130, doi:10.1016/j.geoderma.2015.11.010.

Zhang, H.; Guo, L.; Luo, M.; Zhangyang, C.; Zeng, C.; Wang, S. Spatial Modelling of Soil Organic Carbon Stocks with Combined Principal Component Analysis and Geographically Weighted Regression. **2018**, doi:10.1017/S0021859618000709.

Costa, E.M.; Tassinari, W. de S.; Pinheiro, H.S.K.; Beutler, S.J.; dos Anjos, L.H.C. Mapping Soil Organic Carbon and Organic Matter Fractions by Geographically Weighted Regression. *Journal of Environmental Quality* **2018**, *47*, 718–725, doi:10.2134/jeq2017.04.0178.

Wang, S.; Zhuang, Q.; Jia, S.; Jin, X.; Wang, Q. Spatial Variations of Soil Organic Carbon Stocks in a Coastal Hilly Area of China. *Geoderma* **2018**, *314*, 8–19, doi:10.1016/J.GEODERMA.2017.10.052.

Kumar, S.; Lal, R.; Liu, D.; Rafiq, R. Estimating the Spatial Distribution of Organic Carbon Density for the Soils of Ohio, USA. *J. Geogr. Sci* **2013**, *2013*, 280–296, doi:10.1007/s11442-013-1010-1.

Zhang, C.; Tang, Y.; Xu, X.; Kiely, G. Towards Spatial Geochemical Modelling: Use of Geographically Weighted Regression for Mapping Soil Organic Carbon Contents in Ireland. *Applied Geochemistry* **2011**, *26*, 1239–1248, doi:10.1016/j.apgeochem.2011.04.014.

Mishra, U.; Lal, R.; Liu, D.; van Meirvenne, M. Predicting the Spatial Variation of the Soil Organic Carbon Pool at a Regional Scale. *Soil Science Society of America Journal* **2010**, *74*, 906–914, doi:10.2136/sssaj2009.0158.

Szatmári, G.; Barta, K.; Pásztor, L. An Application of a Spatial Simulated Annealing Sampling Optimization Algorithm to Support Digital Soil Mapping. *Hungarian Geographical Bulletin* **2015**, *64*, 35–48, doi:10.15201/hungeobull.64.1.4.

Brus, D.J. Sampling for Digital Soil Mapping: A Tutorial Supported by R Scripts. *Geoderma* **2019**, *338*, 464–480, doi:10.1016/j.geoderma.2018.07.036.

Brus, D.J.; de Gruijter, J.J. Random Sampling or Geostatistical Modelling? Choosing between Design-Based and Model-Based Sampling Strategies for Soil (with Discussion). *Geoderma* **1997**, *80*, 1–44, doi:10.1016/S0016-7061(97)00072-4.

van Groenigen, J.W. Spatial Simulated Annealing for Optimizing Sampling. In; 1997; pp. 351–361.

Corwin, D.L.; Scudiero, E. Field-Scale Apparent Soil Electrical Conductivity. *Soil Science Society of America Journal* **2020**, *84*, 1405–1441, doi:10.1002/saj2.20153.

Corwin, D.L.; Plant, R.E. Applications of Apparent Soil Electrical Conductivity in Precision Agriculture. *Computers and Electronics in Agriculture* **2005**, *46*, 1–10, doi:10.1016/J.COMPAG.2004.10.004.

Barca, E.; Castrignanò, A.; Buttafuoco, G.; de Benedetto, D.; Passarella, G. Integration of Electromagnetic Induction Sensor Data in Soil Sampling Scheme Optimization Using Simulated Annealing. *Environmental Monitoring and Assessment* **2015**, *187*, 422, doi:10.1007/s10661-015-4570-y.

Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A Powerful Data Mining Procedure to Retrieve Soil Spectral Reflectance from Satellite Images. *Remote Sensing of Environment* **2018**, *212*, 161–175, doi:10.1016/j.rse.2018.04.047.

Silvero, N.E.Q.; Demattê, J.A.M.; Amorim, M.T.A.; Santos, N.V. dos; Rizzo, R.; Safanelli, J.L.; Poppiel, R.R.; Mendes, W. de S.; Bonfatti, B.R. Soil Variability and Quantification Based on Sentinel-2 and Landsat-8 Bare Soil Images: A Comparison. *Remote Sensing of Environment* **2021**, *252*, 112117, doi:10.1016/j.rse.2020.112117.

Safanelli, J.L.; Chabrillat, S.; Ben-Dor, E.; Demattê, J.A.M. Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe. *Remote Sensing* **2020**, *12*, 1369, doi:10.3390/rs12091369.

Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society* **2020**, *146*, 1999–2049, doi:10.1002/qj.3803.

Walkley, A.; Black, A.; An Examination of the Degtjareff Method for Determining Soil Organica Matter, and a Proposed Modification of the Chromic Acid Titration Method. **1934**.

Drusch, M.; del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **2012**, *120*, 25–36, doi:10.1016/J.RSE.2011.11.026.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27, doi:10.1016/J.RSE.2017.06.031.

Mathieu, R.; Pouget, M.; Cervelle, B.; Escadafal, R. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment. *Remote Sensing of Environment* **1998**, *66*, 17–28, doi:10.1016/S0034-4257(98)00030-3.

Weiss, M.; Jacob, F.; Duveiller, G. Remote Sensing for Agricultural Applications: A Meta-Review. *Remote Sensing of Environment* **2020**, *236*, 111402, doi:10.1016/J.RSE.2019.111402.

Thaler, E.A.; Larsen, I.J.; Yu, Q. A New Index for Remote Sensing of Soil Organic Carbon Based Solely on Visible Wavelengths. *Soil Science Society of America Journal* **2019**, *83*, 1443–1450, doi:10.2136/sssaj2018.09.0318.

Xiao, J.; Shen, Y.; Tateishi, R.; Bayaer, W. Development of Topsoil Grain Size Index for Monitoring Desertification in Arid Land Using Remote Sensing. *International Journal of Remote Sensing* **2006**, *27*, 2411–2422, doi:10.1080/01431160600554363.

Segal, D. Theoretical Basis for Differentiation of Ferric-Iron Bearing Minerals, Using Landsat MSS Data. *Proceedings of Symposium for Remote Sensing of Environment, 2nd Thematic Conference on Remote Sensing for Exploratory Geology, Fort Worth, T* **1982**, 949–951.

Rowan, L.C.; Mars, J.C. Lithologic Mapping in the Mountain Pass, California Area Using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Data. *Remote Sensing of Environment* **2003**, *84*, 350–366, doi:10.1016/S0034-4257(02)00127-X.

36.   Douaoui, A.E.K.; Nicolas, H.; Walter, C. Detecting Salinity Hazards within a Semiarid Context by Means of Combining Soil and Remote-Sensing Data. *Geoderma* **2006**, *134*, 217–230, doi:10.1016/J.GEODERMA.2005.10.009.

Lamichhane, S.; Kumar, L.; Wilson, B. Digital Soil Mapping Algorithms and Covariates for Soil Organic Carbon Mapping and Their Implications: A Review. *Geoderma* **2019**, *352*, 395–413, doi:10.1016/J.GEODERMA.2019.05.031.

Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis* **2010**, *28*, 281–298, doi:10.1111/j.1538-4632.1996.tb00936.x.

Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Routledge, 2017; ISBN 9780203753781.

Steichen RJRT, T.J.; Cox, N.J. *A Note on the Concordance Correlation Coefficient*; 2002; Vol. 2;.

van Groenigen, J.W.; Siderius, W.; Stein, A. Constrained Optimisation of Soil Sampling for Minimisation of the Kriging Variance. *Geoderma* **1999**, *87*, 239–259, doi:10.1016/S0016-7061(98)00056-1.

Brus, D.J.; Heuvelink, G.B.M. Optimization of Sample Patterns for Universal Kriging of Environmental Variables. *Geoderma* **2007**, *138*, 86–95, doi:10.1016/J.GEODERMA.2006.10.016.

Keskin, H.; Grunwald, S. Regression Kriging as a Workhorse in the Digital Soil Mapper's Toolbox. *Geoderma* **2018**, *326*, 22–41, doi:10.1016/J.GEODERMA.2018.04.004.

Anselin, L.; Rey, S. Properties of Tests for Spatial Dependence in Linear Regression Models. *Geographical Analysis* **2010**, *23*, 112–131, doi:10.1111/j.1538-4632.1991.tb00228.x.

Castellini, M.; Stellacci, A.M.; Tomaiuolo, M.; Barca, E. Spatial Variability of Soil Physical and Hydraulic Properties in a Durum Wheat Field: An Assessment by the BEST-Procedure. *Water* **2019**, *11*, 1434, doi:10.3390/w11071434.

Adhikari, D.; Yang, Y. Selective Stabilization of Aliphatic Organic Carbon by Iron Oxide OPEN. *Nature Publishing Group* **2015**, doi:10.1038/srep11214.

CHAPTER III: Preliminary suitability analysis of *cv.* "Senatore Cappelli": Effect of soil and topographic properties on yield variability

1.   Introduction

The Durum wheat (*Triticum durum* Desf.) *cv.* "Senatore Cappelli" is a variety selected by Nazareno Strampelli. The cultivar "Senatore Cappelli", a pure line belonging to the Mediterranean type extracted from the North African population Jean Retifah in 1915.

"Senatore Cappelli" is characterized by very tall plants, low production, and high grain protein content (Dinelli et al. 2013; Giunta et al. 2019; Pagnani et al. 2020). Until the advent of the Green Revolution in late 1960, "Senatore Cappelli" was the most important variety in the Mediterranean country (Mefleh et al. 2018). The progressive release of new cultivars, more appreciated by farmers for their higher productivity, has significantly reduced the areas cultivated with "Senatore Cappelli", limiting it to a niche market. In the last years, "Senatore Cappelli" has met with considerable interest from the market for its high quality (Acquistucci et al. 2020; Bianco et al. 2019; Mascia et al. 2014; Mefleh et al. 2018; Rocco et al. 2019) and nutraceutical properties (Mascia et al. 2014). Future European agricultural production will need to be at the same time more productive and more sustainable due to the increasing world population and climate change (Farooq et al. 2019) Furthermore, consumers demand products of national origin grown sustainably (Borsellino et al. n.d.). A sustainable agri-food system is a knowledge-based system requiring new knowledge (Colizzi et al. 2020; Sharma et al. 2018; Vecchio et al. 2020), such as Crop Suitability (Akpoti et al. 2019; el Baroudy 2016).

Suitability analysis is a process of matching crop requirements with field variability to identify production suitability at a given location (Elsheikh et al. 2013) Therefore understanding interactions between yield and topographical features and soil properties is fundamental knowledge for sustainable agriculture (Jiang and Thelen n.d.; Kravchenko and Bullock n.d.). Remote sensing allows collecting this information cheaply and quickly (Iticha and Takele 2019). Extraction of topographical features from Digital Elevation Model (DEM) is widely used (McBratney et al. 2003) while satellite bare soil images in the suitability analysis have not been extensively investigated. Recent studies have been developed new methodologies to obtain bare soil images, such as the Geospatial Soil Sensing System (GEOS3) Demattê et al. (2018) born on Landsat satellite and subsequently developed using Sentinel-2 (Silvero et al. 2021).

The objective of this study is:

a) to determine how selected topographic information and soil properties can explain yield variability of *cv*. "Senatore Cappelli" at open field scale;

b) to assess the ability to predict such variability through remote sensing data.

2.   Materials and Methods

2.1.   Study area

The study area is located in the deltaic plain of Po River, near Massa Fiscaglia (44° 79' N; 11°98' E), in Northern Italy. The study site area is composed of a field (ID 19), as reported in Figure 1, with a surface of approximately 12 ha. A 2-year field experiment was carried out during the 2020 and 2020–2021 growing seasons (referred in the text as 2020 and 2021, respectively). The climate of this area is temperate. According to ERA 5 LAND (Hersbach et al. 2020), in the last 40 years (1981- 2021), the yearly average precipitation is 770 mm; yearly average air temperature is 14.61 °C, yearly average minimum air temperature 0.43 °C, and yearly average maximum air temperature 28.79 °C.

2.2.   Crop and Yield data

In the season 2020, the sowing has been done on January 20th: with a density of 250 seed m^{-2}, and the harvest has been done on July 10th. In the season 2021, the sowing was done on December 15th, 2020, with same density of 250 seed m^{-2}; and the harvest was done on July 12th, 2021. Each year, a 1 m^2 plot of wheat was harvested at each of 122 previously sampled sites.

2.3.   Weather data

Weather data about the growing period of two seasons were collected by ERA 5 LAND, particularly average precipitation, air temperature, average minimum air temperature, and air temperature have been recorded.

2.4.   Soil Sampling

A Soil sampling was carried to determine the predominant soil properties in August 2020. A grid of 122 points, 61 according to a regular grid and 61 chosen randomly (Fig. 1), was used to take the sample from topsoil (0-30 cm).

The coordinates of each sampling site were determined with a GPS model SP20 handheld GNSS (Spectra Geospatial), and a private laboratory carried out the soil analysis. The samples were air-dried and analysed using the Robinson pipette for particle size distribution.

Organic Carbon was determined by dichromate oxidation (Walkley and Black 1934).

2.5. Topographical data

DEM provided by Emilia Romagna Region (https://geoportale.regione.emilia-romagna.it/catalogo/dati-cartografici) with a pixel spatial resolution of 5 meters, was resampled at the exact resolution of VIS-NIR bands of Sentinel-2 Satellite (10 meters) using the algorithm Nearest Neighbour in Qgis. Using the resampled DEM, several topographic indices were calculated (Table 1) using Qgis software.
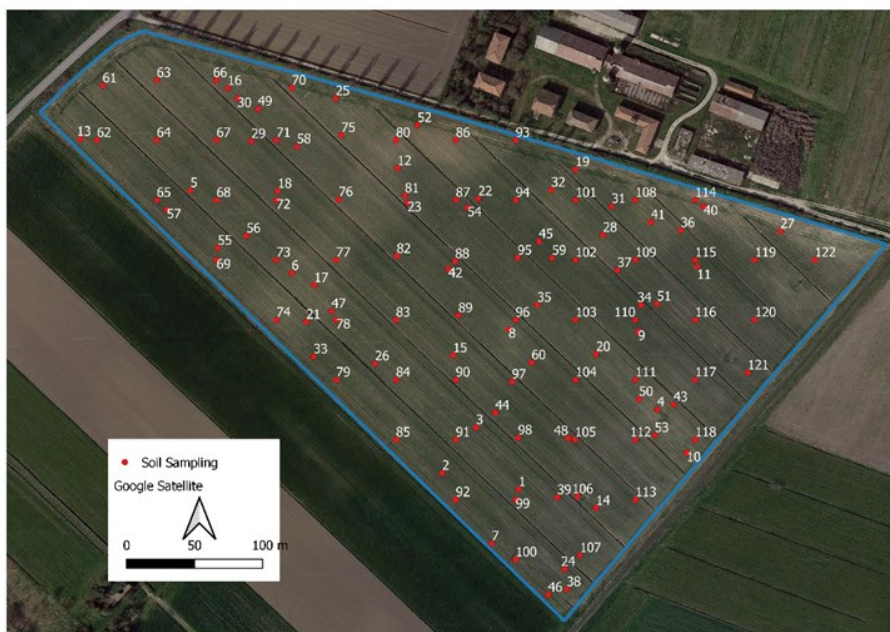

Figure 14. Soil sampling grid.

Topographic Indices
Slope
Tangential Curvature
Longitudinal Curvature
Total Curvature
Profile Curvature
General Curvature
Plan Curvature
Minimal Curvature
Maximal Curvature
Aspect
Hillshade
Elevation

Table 1. Topographic Indices.

2.6. Satellite data and multi-temporal images

Sentinel2-MSI satellite data (Drusch et al. 2012), as surface reflectance products, were used for this work. The Sentinel-2 MultiSpectral Instrument (S2-MSI) mission was launched by the European Space Agency (ESA) in 2015. Two satellites (2A and 2B) were placed in orbit, provide five days revisit time. Sentinel2-MSI has spatial resolutions ranging from 10 to 60 meters and thirteen spectral bands that covet the visible, near-infrared, red-edge, shortwave-infrared and water and cirrus region. A multi-temporal bare soil image named SYSI was obtained by applying the GEOS3 (Demattê et al. 2018) using Sentinel-2 satellite data (Silvero et al. 2021) on the Google Earth Engine platform (GEE) Gorelick et al. (2017).

One hundred fourteen images from 2018-2021 with at least 80% of the area without clouds and shadows were used. The GEOS3 method consists of (1) obtaining the images, (2) masking pixels that were non-bare soil, and (3) calculating the median reflectance of the masked images.

After obtaining Sentinel-2 images, NDVI (Normalized Difference Vegetation Index) and NBR2 (Normalized Burn Ratio 2) were calculated and used to mask the images.

Following the GEOS3 method, the range between 0 and 0.25 for NDVI was considered bare soil, and the values above this threshold were flagged as N.A..

NBR2 values > 0.075 were also flagged as N.A. and corresponded to sites covered with straw or burned. For each image, quality masked was used to exclude areas with clouds and shadows. Finally, the masked images were ordered by date and the median reflectance was calculated.

Form SYSI VIS-NIR bands (Table 2) were extracted with a spatial pixel resolution of 10 meters.

Sentinel -2 Bands	Central wavelength (nm)
Band 2 (B2) - Blue	490
Band 3 (B3) - Green	560
Band 4 (B4) - Red	665
Band 8 (B8) - NIR	842

Table 2. Sentinel-2 Bands characteristics.

2.7. Data Analysis Procedure

2.7.1. Selection of topographical factors

Principal Component Analysis (PCA) was run to investigate topographic variance and identify important topographic variables to be used as inputs for suitability analyses. Principal components (PCs) formula can be written:

$$Y_i = a_{i1}X_1 + a_{i_2}X_2 + \cdots . + a_{ip}X_p$$

Where Y_i is the *i*-th PC; X_1, …,X_p are the original variables; and a_{i1},…, a_{ip} are the coefficients of the *i*-th PC and an index of relative importance of the variable to that PC. The greater the absolute values of a_{i1}, the greater the importance of those soil variables to the PC. A covariance matrix was used in the PC calculation due to differences in order of magnitude between soil variables measured. Topographic variables in each retained PC were empirically analyzed and selected based on their loading coefficients (a_i). The loadings with values more than 0.40 were considered in each PC and were believed to have a more significant effect on yield variability.

2.7.2. Correlation and stepwise regression between topographic variables, soil data, and yield

Correlation coefficients were calculated among the selected topographic variables, soil data, and yield. Finally, a stepwise regression was used to analyze soil properties and topographic data's combined effect on crop yield.

2.7.3. Stepwise regression between topographic variables, SYSI, and yield

Finally stepwise regression was performed again, replacing the soil data with SYSI VIS-NIR bands.

3. Result and discussion

3.1. Weather data

Weather data of two seasons were reported in Table 3. In the 2020 season, the average air temperature was to 14.18 °C, and the total precipitation was about 272 mm. The season 2021 has recorded a minor average temperature (12.78 °C) and higher rainfall level (385 mm) than 2020. These data are in line with historical data; therefore, no climatic anomalies have been found in these two years.

	2020	**2021**
Temperature Mean, °C	14.18	12.78
Temperature Minimum, °C	8.69	7.43
Temperature Maximal, °C	19.91	18.23
Total Precipitation, mm	272.03	385.52

Table 3. Weather data.

3.2. Yield data

The yield data (t/ha) of the two growing seasons are summarized in Table 4.

Yield	**2020**	**2021**
Mean	4.55	3.74
Standard Deviation	0.74	0.97
Minimum	2.43	1.13
Q1	3.94	3.06
Median	4.58	3.93
Q3	5.14	4.39
Maximum	6.30	5.86

Table 4. Yield data (t/ha).

The average yield of the sampling point in the first year was 4.55 t/ha, having a maximum of 6.30 t/ha and a minimum of 2.43 t/ha. In the second year, the average yield was 3.74 t/ha, recording a decrease of about 18% than the first year. The maximum yield value in 2021 was 5.86 t/ha, while the minimum value was 3.74. These results confirmed the low productivity of Senatore Cappelli, in accordance with literature (Dinelli et al. 2013; Giunta et al. 2019; Pagnani et al. 2020).

3.3. Soil Sampling

As reported in Chapter 1, the 122 sample points were classified according to USDA (United States Department of Agriculture) Textural soil classification (Table 5).

N°	Textural soil classification
1	(Cl) Clay
18	(ClLo) Clay Loam
13	(Lo) Loam
25	(SiCl) Silty Clay
56	(SiClLo) Silty Clay Loam
9	(SiLo) Silty loam

Table 5. Textural soil classification.

Table 6 provides a statistical summary of the entire SOC (Soil Organic Carbon) data set of 120 observations after the remove of outliers, that has been explained in Chapter 2.

N	Min.	Max.	Mean	STD	Skew	Kurt
120	1.16	4.38	2.25	0.54	0.78	1.49

Table 6. Summary Statistics of SOC Sample Data.

N = Number of samples; Min. = Minimum; Max. = Maximum; STD = Standard Deviation;

Skew = Skewness; Kurt = Kurtosis

3.4. Data Analysis Procedure

3.4.1. Selection of topographical factors: PCA

The PCA results are presented in Table 7. The first four PCs all had an eigenvalue greater than one and cumulatively explained 90% of the total sample variance. In PC1, variable loadings of absolute dominance were not observed, as in PC4. In contrast, PC2 and PC3 had higher loadings, General curvature (0.403) in PC1. Plan Curvature (-0.516) and Tangential Curvature (-0.538) in PC2 (Table 8). Elevation, Hillshade, Maximal Curvature in PC3 were identified as critical variables.

Importance of components	PC1	PC2	PC3	PC4	PC5
Standard deviation	2.42	1.60	1.15	1.05	0.75
Proportion of Variance	0.49	0.21	0.11	0.09	0.04
Cumulative Proportion	0.49	0.70	0.81	0.90	0.95

Table 7. Individual and cumulative variance and eigen values.

Variables	PC1	PC2	PC3	PC4
Elevation	0.121	-0.002	0.590	0.396
Hillshade	-0.119	-0.150	-0.574	-0.330
General C.	0.403	-0.057	-0.117	0.089
Longitudinal. C	0.381	0.142	-0.180	0.176
Maximal C.	0.111	-0.287	-0.413	0.650
Minimal C.	0.386	0.089	0.090	-0.247
Plan C.	0.158	-0.516	0.151	-0.192
Profile C.	0.352	0.248	-0.205	0.186
Tangential C.	0.171	-0.538	0.125	-0.145
Total C.	-0.338	-0.265	-0.095	0.229
Transverse C.	0.327	-0.363	0.0158	-0.074
Slope	-0.316	-0.203	0.0759	0.241

Table 8. Variable loading.

3.4.2. Correlation and stepwise regression between topographic variables, soil data, and yield

Significant correlation coefficients ($p \leq 0.10$) between soil properties and topographical features selected by PCA, are shown in Table 9. Elevation was correlated with all soil properties, in particular a positive correlation whit Silt and negative correlations with SOC (-0.30), Clay (- 0.24), and Sand (-0.24). Moreover, SOC was correlated with Hillshade (0.37), while Silt was also correlated with Plan Curvature (0.15) and Maximal Curvature (0.24). This means that soil properties follow an elevation-slope gradient, as widely explained in Chapter 1. Field topography can have an indirect effect through its influence on the distribution of specific soil chemical and physical properties, such as particle size distribution and organic matter (Jiang and Thelen n.d.; Kravchenko and Bullock n.d.).

General C	Silt	Sand	Clay	SOC
Plan C	0.15			
Tangential C				
Elevation	0.45	- 0.13	- 0.24	- 0.30
Hillshade				0.37
Maximal C	0.24			
General C				

Table 9. Correlation between topographic variables and soil data.

Correlation coefficients ($p \leq 0.10$) between topographic features and yield and soil properties are shown in Table 10. In the season 2020, the yield had a positive correlation with the topographic feature, in particular, Plan Curvature (0.15) and General Curvature (0.31). While in the second growing season, the yield was negatively correlated with Maximal Curvature (-0.15) and General Curvature (- 0.27). These results highlight that field terrain conformation can positively or negatively affect the yield. This difference can be explained that yield spatial variability is affected by the interaction between the soil characteristics, position in the landscape, and weather (Maestrini and Basso 2018). The difference in total precipitation can explain this difference. In fact, in the second year, total precipitation was approximately 40% greater than 2020 and mainly distributed in the early cultural stages. Therefore, in the concave areas of the field, there is a more significant accumulation of water, which may have negatively affected the development of the plant, generating low yields than other zones of the field. These results are in agreement with the reported literature (Kravchenko and Bullock n.d.) .

General C	Yield 2020	Yield 2021
Plan C	0.15	
Tangential C		
Elevation		
Hillshade		
Maximal C		-0.15
General C	0.31	-0.27
Silt		
Sand		
Clay		
SOC		

Table 10. Correlation between topographic variables and yield.

Finally, stepwise multiple regression was performed to determine how selected topographic information and soil properties can explain yield variability. The covariates in which significant contribution to regression ($p < 0.10$) were retained in the regression equation. The covariates selected for the two sasons, were:

$$Yield\ 2020 \sim Silt + Sand + SOC + General\ Curvature$$
$$Yield\ 2021 \sim Silt + SOC + General\ Curvature + Elevation$$

The coefficient of determination from multiple linear regression between yields and soil properties and topographic features are 0.130 in 2020 and 0.135 in 2021. This means that soil properties and topographic factors can explain only 13% yield variability. These results confirmed that the field's elevation-slope gradient indirectly affects crop behavior. Therefore, topographic features will be considered in site-specific crop management. Among the soil properties studied, SOC and Silt are the most influence on the yield. As reported in the literature (Kravchenko and Bullock n.d.) Organic Matter (OM) content was a more critical yield-affecting factor in soil with low OM content than in soil with high OM content. According to the literature (Kravchenko and Bullock n.d.) in sites with relatively low OM content ($< 3\%$), yields were affected by OM. However, no correlation between yields and OM content was observed for sites with high OM ($> 3\%$). This assumption can explain the results of this study reported in Table 10, where no correlation between soil properties and yield has been found. Therefore, rapid methods of SOC mapping as reported in Chapter 2, can be fundamental in suitability analysis.

3.4.3. Stepwise regression between topographic variables, SYSI, and yield

The stepwise multiple regression was performed to assess the ability to predict yield field variability using remote sensing data, in particular replacing soil properties with SYSI (Synthetic Soil Images). The covariates in which significant contribution to regression ($p < 0.10$) were retained in the regression equation. The covariates selected for the two seasons are:

$$Yield(rs)2020 \sim B8 + General\ Curvature$$
$$Yield(rs)2021 \sim B4 + B8 + General\ Curvature$$

The coefficient of determination from Multiple Linear Regression between yields and VIS-NIR SYSI and topographic features was 0.112 in 2020 and 0.118 in 2021. Between regression with soil data and SYSI images, minimal difference in coefficient of determination was found. The correlation between SYSI and soil properties have been widely explained in Chapter 1 and Chapter 2. These results further confirmed the importance of the topographic factors in yield variability and that SYSI is a valid tool for detecting soil variability.

Therefore, combined use of the topographic features and VIS-NIR SYSI they will be able to provide a useful contribution in the management of the specific site of Durum wheat *cv.* "Senatore Cappelli".

3.   Conclusion

The Durum wheat *cv.* "Senatore Cappelli" is characterized by very tall plants, low production, and high grain protein content. In the last years, the *cv.* "Senatore Cappelli" has met with considerable interest from the market for its grain quality and nutraceutical properties. This study provided a preliminary suitability analysis for the sustainable cultivation of this specific cultivar of Durum wheat. The results highlight that soil and topography properties explain only 13% of with-in yield variability. Additional investigation is needed to study the possible correlation of wheat lodging and yield, given that this aspect is relevant for the sustainability of management. SYSI alone produced accurate predictions and confirm that it is a good way to quickly obtain information on soil variability and suitability for the cultivation of specific crop or cultivar, as in this Ph.D. project.

4.   References

Dinelli, G.; Marotti, I.; di Silvestro, R.; Bosi, S.; Bregola, V.; Accorsi, M.; di Loreto, A.; Benedettelli, S.; Ghiselli, L.; Catizone, P. Agronomic, Nutritional and Nutraceutical Aspects of

Durum Wheat (Triticum Durum Desf.) Cultivars under Low Input Agricultural Management. *Italian Journal of Agronomy* **2013**, *8*, 12, doi:10.4081/ija.2013.e12.

Giunta, F.; Pruneddu, G.; Motzo, R. Grain Yield and Grain Protein of Old and Modern Durum Wheat Cultivars Grown under Different Cropping Systems. *Field Crops Research* **2019**, *230*, 107–120, doi:10.1016/J.FCR.2018.10.012.

Pagnani, G.; Galieni, A.; Stagnari, F.; Pellegrini, M.; del Gallo, M.; Pisante, M. Open Field Inoculation with PGPR as a Strategy to Manage Fertilization of Ancient Triticum Genotypes. *Biology and Fertility of Soils* **2020**, *56*, 111–124, doi:10.1007/s00374-019-01407-1.

Mefleh, M.; Conte, P.; Fadda, C.; Giunta, F.; Piga, A.; Hassoun, G.; Motzo, R. From Ancient to Old and Modern Durum Wheat Varieties: Interaction among Cultivar Traits, Management, and Technological Quality. **2018**, doi:10.1002/jsfa.9388.

Rocco, M.; Tartaglia, M.; Izzo, F.P.; Varricchio, E.; Arena, S.; Scaloni, A.; Marra, M. Comparative Proteomic Analysis of Durum Wheat Shoots from Modern and Ancient Cultivars. *Plant Physiology and Biochemistry* **2019**, *135*, 253–262, doi:10.1016/j.plaphy.2018.12.010.

Acquistucci, R.; Melini, V.; Galli, V. Durum Wheat Grain and Pasta from Locally-Grown Crops: A Case-Study on Saragolla (Triticum Turgidum Ssp. Turanicum) and Senatore Cappelli (Triticum Turgidum Ssp. Durum) Wheats. *Emirates Journal of Food and Agriculture* **2020**, 47, doi:10.9755/ejfa.2020.v32.i1.2058.

Mascia, I.; Fadda, C.; Dostálek, P.; Olšovská, J.; Caro, A. del Preliminary Characterization of an Italian Craft Durum Wheat Beer. **2014**, doi:10.1002/jib.176.

Bianco, A.; Fancello, F.; Balmas, V.; Dettori, M.; Motroni, A.; Zara, G.; Budroni, M. Microbial Communities and Malt Quality of Durum Wheat Used in Brewing. **2019**, doi:10.1002/jib.555.

Farooq, M.; Rehman, A.; Pisante, M. Sustainable Agriculture and Food Security. In *Innovations in Sustainable Agriculture*; Springer International Publishing: Cham, 2019; pp. 3–24.

Borsellino, V.; Schimmenti, E.; Bilali, H. el Agri-Food Markets towards Sustainable Patterns., doi:10.3390/su12062193.

Colizzi, L.; Caivano, D.; Ardito, C.; Desolda, G.; Castrignanò, A.; Matera, M.; Khosla, R.; Moshou, D.; Hou, K.-M.; Pinet, F.; et al. Introduction to Agricultural IoT. In *Agricultural Internet of Things and Decision Support for Precision Smart Farming*; Elsevier, 2020; pp. 1–33.

Sharma, R.; Kamble, S.S.; Gunasekaran, A. Big GIS Analytics Framework for Agriculture Supply Chains: A Literature Review Identifying the Current Trends and Future Perspectives. *Computers and Electronics in Agriculture* **2018**, *155*, 103–120, doi:10.1016/j.compag.2018.10.001.

Vecchio, Y.; Agnusdei, G.P.; Miglietta, P.P.; Capitanio, F. Adoption of Precision Farming Tools: The Case of Italian Farmers. *International Journal of Environmental Research and Public Health* **2020**, *17*, 869, doi:10.3390/ijerph17030869.

el Baroudy, A.A. Mapping and Evaluating Land Suitability Using a GIS-Based Model. *CATENA* **2016**, *140*, 96–104, doi:10.1016/j.catena.2015.12.010.

Akpoti, K.; Kabo-bah, A.T.; Zwart, S.J. Agricultural Land Suitability Analysis: State-of-the-Art and Outlooks for Integration of Climate Change Analysis. *Agricultural Systems* **2019**, *173*, 172–208, doi:10.1016/j.agsy.2019.02.013.

Elsheikh, R.; Mohamed Shariff, A.R.B.; Amiri, F.; Ahmad, N.B.; Balasundram, S.K.; Soom, M.A.M. Agriculture Land Suitability Evaluator (ALSE): A Decision and Planning Support Tool for

Tropical and Subtropical Crops. *Computers and Electronics in Agriculture* **2013**, *93*, 98–110, doi:10.1016/j.compag.2013.02.003.

Kravchenko, A.N.; Bullock, D.G. *SPATIAL VARIABILITY Correlation of Corn and Soybean Grain Yield with Topography and Soil Properties*;

Jiang, P.; Thelen, K.D. *SITE-SPECIFIC ANALYSIS Effect of Soil and Topographic Properties on Crop Yield in a North-Central Corn-Soybean Cropping System*;

Iticha, B.; Takele, C. Digital Soil Mapping for Site-Specific Management of Soils. *Geoderma* **2019**, *351*, 85–91, doi:10.1016/J.GEODERMA.2019.05.026.

McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On Digital Soil Mapping. *Geoderma* **2003**, *117*, 3–52, doi:10.1016/S0016-7061(03)00223-4.

Demattê, J.A.M.; Fongaro, C.T.; Rizzo, R.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A Powerful Data Mining Procedure to Retrieve Soil Spectral Reflectance from Satellite Images. *Remote Sensing of Environment* **2018**, *212*, 161–175, doi:10.1016/j.rse.2018.04.047.

Silvero, N.E.Q.; Demattê, J.A.M.; Amorim, M.T.A.; Santos, N.V. dos; Rizzo, R.; Safanelli, J.L.; Poppiel, R.R.; Mendes, W. de S.; Bonfatti, B.R. Soil Variability and Quantification Based on Sentinel-2 and Landsat-8 Bare Soil Images: A Comparison. *Remote Sensing of Environment* **2021**, *252*, 112117, doi:10.1016/j.rse.2020.112117.

Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society* **2020**, *146*, 1999–2049, doi:10.1002/qj.3803.

Walkley, A.; Black, A.; An Examination of the Degtjareff Method for Determining Soil Organica Matter, and a Proposed Modification of the Chromic Acid Titration Method. **1934**.

Drusch, M.; del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment* **2012**, *120*, 25–36, doi:10.1016/J.RSE.2011.11.026.

Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone. *Remote Sensing of Environment* **2017**, *202*, 18–27, doi:10.1016/J.RSE.2017.06.031.

Maestrini, B.; Basso, B. Predicting Spatial Patterns of Within-Field Crop Yield Variability. *Field Crops Research* **2018**, *219*, 106–112, doi:10.1016/J.FCR.2018.01.028.

CHAPTER IV: Research Conclusions

The Durum wheat *cv*. "Senatore Cappelli" is characterized by very tall plants, low production, and high grain protein content. In the last years, *cv*. "Senatore Cappelli" has met with considerable interest from the market for its high grain quality and nutraceutical properties.

The main goal of this thesis was to develop a Decision Support System (DSS) for the sustainable cultivation of *cv*. "Senatore Cappelli". The nature of the decision itself may depend on the availability of data. Therefore, the basic idea of this Ph.D. project was to create tools capable of intercepting field variability quickly and economically.

This DSS can be divided into two modules: a) Digital Soil Mapping Module (DSM), which uses rapid methods to assess field variability and map soil properties. b) Crop Suitability module where the information obtained from the first module are used to explain yield variability.

In particular, alternative methods at the apparent soil electrical conductivity (ECa) were evaluated in the DSM module. Geospatial ECa mapping is a sensor technology that has a significant role in spatial field characterization. However, this method has several limitations; therefore, the rapid expansion of Precision Agriculture required an alternative, faster way to map spatial field variability.

The results of chapter one can be summarized:

- Multi-temporal images with bare soil pixels (SYSI), can be a valid alternative to ECa, understood as a proxy of soil variability.
- The red band (B4) of SYSI and the bare soil index SOCI (Soil Organic Carbon Index) can be used as a rapidly and economically proxy of soil variability.

Moreover, in DSM module, rapid methods for Soil Organic Carbon mapping were assessed.

The study and management of the Soil Organic Carbon variability are essential for a new sustainable agriculture. The rapid expansion of Precision Agriculture requires rapid and inexpensive methods for SOC mapping. The results of chapter two are:

- Bare soil indices provided by SYSI can explain SOC variability.
- SYSI could be used as auxiliary information in sampling optimization.
- SYSI allows a reduction of sample scheme by about 75% with a consequent significant cost reduction.

In Crop Suitabilty module, some selected topographic information and soil properties have been used to explain yield variability of *cv.* "Senatore Cappelli" at field scale.

In chapter three, preliminary suitability analysis for the sustainable cultivation of *cv.* "Senatore Cappelli" was carried out using soil data and bare soil images by SYSI. The results of chapter three can be summarized:

- Soil and topography properties explain only 13% of with-in grain yield variability.
- SYSI alone produced accurate predictions and confirmed that it is a good way to quickly obtain information on soil variability and suitability for the cultivation of a specific crop or cultivar.

  These results require further validations in other areas of study, and additional investigation is needed to study the possible correlation of wheat lodging and yield of the *cv.* "Senatore Cappelli". This method could be a vital impulse for Precision Agriculture adoption and diffusion. Therefore, this result could open the way to new accurate and cheaper services based on data and informed decisions.

Further Research

Further research is required to confirm and to wide these conclusions. A list of main foreseen tasks follows:

- Geospatial data play a significant role in evaluating the spatial variability of cultivated fields and to support the rapid expansion of Precision Agriculture it is necessary to overcome current limitations by investing in more transdisciplinary research and development on a territorial scale;

- better integration of data and systems for to evaluate the influence of agricultural practices, soil health and soil–water interactions, fostering closer partnerships with farmers, who will be the people responsible for implementing new techniques and approaches;

- the preliminary analysis of the suitability for cultivation to evaluate the real sustainability of the cultivation systems, appears to be a promising model at territorial scale, and also at crop and cultivar level, as demonstrated in this study, which requires systematic and diversified applications.