Contribution des séries temporelles satellitaires à la cartographie du carbone organique des sols cultivés à divers échelons régionaux

Diego Fernando Urbina Salazar

THESE DE DOCTORAT

NNT : 2023UPASB028

université PARIS-SACLAY

Contribution des séries temporelles satellitaires à la cartographie du carbone organique des sols cultivés à divers échelons régionaux

Satellite time series contribution to organic carbon mapping in cultivated soils at various regional scales

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 581 : agriculture, alimentation, biologie, environnement et santé (ABIES)
Spécialité de doctorat : Sciences de l'environnement
Graduate School : Biosphera. Référent : AgroParisTech

Thèse préparée dans l'UMR **ECOSYS** (Université Paris-Saclay, INRAE, AgroParisTech), sous la direction de **Emmanuelle VAUDOUR**, Directrice de Recherche, le co-encadrement de **Dominique ARROUAYS**, Ingénieur de Recherche

Thèse soutenue à Paris-Saclay, le 15 mai 2023, par

Diego Fernando URBINA SALAZAR

Composition du Jury

Membres du jury avec voix délibérative

Christian WALTER Professeur, L'Institut Agro Rennes	Président
Sabine CHABRILLAT Professeure, Leibniz University Hannover (Allemagne)	Rapporteur & Examinateur
François COLIN Professeur, L'Institut Agro Montpellier	Rapporteur & Examinateur
Dominique COURAULT Directrice de Recherche, INRAE (Université d'Avignon)	Examinatrice

Titre : Contribution des séries temporelles satellitaires à la cartographie du carbone organique des sols cultivés à divers échelons régionaux

Mots clés : Carbone organique du sol ; cartographie numérique des sols; séries satellitaires Sentinel; sol nu; terres agricoles.

Résumé : Le carbone organique du sol (COS) dans les zones agricoles joue un rôle clé dans la sécurité alimentaire et l'atténuation du changement climatique. La quantification du COS est nécessaire pour mettre en œuvre des techniques et des pratiques de stockage. Cependant, l'échantillonnage du COS dans un monde qui comporte environ 1,5 milliard d'hectares de sols agricoles est un véritable défi. C'est pourquoi l'utilisation de technologies telles que les capteurs satellitaires pourrait être une alternative prometteuse pour quantifier et cartographier le COS dans différents types d'agroécosystèmes à travers le monde. L'objectif de cette thèse est celui d'évaluer le potentiel des images satellitaires Sentinel-2 (S2) et Sentinel-1 (S1) pour la cartographie du COS dans les agroécosystèmes de la France métropolitaine en utilisant des modèles spectraux et spatio-spectraux. Le chapitre 1 aborde l'état d'avancement de la cartographie du COS en France et présente les principales limitations et méthodes actuellement reposant sur les données d'images satellitaires pour la prédiction du COS. Le chapitre 2 présente les régions d'étude situées en Bretagne, Occitanie et Centre Val de Loire. De plus, les principaux ensembles de données utilisés sont décrits et une analyse préliminaire de l'une des régions d'étude est présentée. Le troisième chapitre évalue le potentiel des images S2 et des produits dérivés de S1 et S2 pour prédire le SOC à l'aide d'images à date unique (mono-date). Dans ce chapitre comme dans le second, des limitations liées principalement aux conditions de surface du sol ont été observées ; et les meilleures dates d'image pour détecter le SOC ont été identifiées. Dans le quatrième chapitre, au lieu d'images à date unique, l'utilisation de mosaïques temporelles S2 de sol nu (S2Bsoil) par périodes est abordée, tout comme l'utilisation de covariables dérivées de l'imagerie satellitaire et du terrain. Ce chapitre traite de l'importance de la sélection des périodes de production de S2Bsol et de l'utilisation de covariables pertinentes pour comprendre la variabilité spatiale du COS à l'échelle régionale. Enfin, le dernier chapitre aborde les principaux constats et perspectives à envisager dans un futur proche.

Title : Satellite time series contribution to organic carbon mapping in cultivated soils at various regional scales

Keywords : Soil organic carbon; digital soil mapping; Sentinel time series; bare soil; croplands.

Abstract : Soil organic carbon (SOC) in agricultural areas plays a key role in food security and climate change mitigation. SOC quantification is necessary in order to implement storage techniques and practices. However, SOC sampling in a world that approximately encompasses 1.5 billion hectares of agricultural soils is quite challenging. Therefore, the use of technologies such as satellite sensors may be a promising alternative to quantify and map SOC in various types of agroecosystems around the world. The objective of this thesis is to evaluate the potential of Sentinel-2 (S2) and Sentinel-1 (S1) satellite imagery for SOC mapping in agro-ecosystems in mainland France using spectral and spatio-spectral models. Chapter 1 addresses the progress of SOC mapping in France and presents the main limitations and methods currently used with satellite image data for SOC prediction. Chapter 2 presents the study areas located in the Brittany, Occitanie and Centre Val de Loire regions. In addition, the main datasets used are described and a preliminary analysis of one of the study areas is shown. The third chapter evaluates the potential of S2 images and derived products of S1 and S2 to predict SOC using single-date images. In this chapter as in the second, limitations related mainly to soil surface conditions were observed; and the best image dates for detecting SOC were identified. In the fourth chapter, instead of single date images, the use of S2 temporal mosaics of bare soil (S2Bsoil) by periods are addressed as the use of covariates derived from satellite imagery and terrain. This chapter deals the importance of selecting periods for the production of S2Bsoil and the use of relevant covariates to understand the spatial variability of SOC on a regional scale. Finally, the last chapter discusses the main findings and perspectives to be considered in the near future.

FINANCEMENTS

Diego Urbina Salazar a reçu une allocation doctorale financée conjointement par l'Agence de la transition écologique (ADEME) et par l'Institut National de Recherche pour l'Agriculture, l'alimentation et l'environnement (INRAe).

Cette thèse a été réalisée dans le cadre de deux projets : au niveau national, le projet PEDO_POLYPHEME (2020-2022) du programme TOSCA financé par le Centre National d'Etudes Spatiales (CNES), et au niveau européen, le projet STEROPES (2021-2024) du programme européen Horizon H2020 de recherche et d'innovation sur la gestion des sols agricoles "European Joint Programme Cofund on Agricultural Soil Management" (EJP-SOIL grant number 862695).

Cette thèse a aussi bénéficié, à son début, d'un soutien du Programme National de Télédétection Spatiale à travers le projet TELEMOS (PNTS, http://programmes.insu.cnrs.fr/pnts/), grant n° PNTS-2020-17.

REMERCIEMENTS

Il y a sept ans, je finissais tout juste ma licence à l'université en Colombie. Ayant toujours été attiré par l'agriculture, l'environnement et la technologie, j'ai décidé de faire un doctorat en 2019, à l'occasion d'un stage dans une ferme et du 5e workshop international sur la détection proximale du sol dans le Missouri. Je dois dire que cela a été un parcours stimulant, au cours duquel j'ai appris, beaucoup ri, pleuré et même subi une quarantaine en raison du confinement imposé par la Covid19. Cela fait 3 ans et demi que je suis arrivé en France et le fait d'être si près de terminer cette étape de ma vie est une source de joie. Je voudrais tout d'abord remercier Dieu de m'avoir soutenu jusqu'à ce jour, de m'avoir donné la force, la sagesse, la santé et chacune de ses bénédictions et sa grâce infinies.

Je tiens à remercier ma directrice de thèse, Dr. Emmanuelle Vaudour, pour sa qualité professionnelle et humaine. Je la remercie d'avoir veillé à mon bien-être tout au long de cette thèse et d'avoir accompagné le bon développement de ce projet. Merci Emmanuelle pour le partage de vos connaissances approfondies en télédétection du sol, pour ces années de travail et d'échanges sur des sujets très variés. Merci également à mon co-directeur de thèse, Dr. Dominique Arrouays pour ses conseils, ses échanges éclairants, son excellent sens de l'humour et pour le partage de sa richesse de connaissances en sciences du sol. Merci Emmanuelle et Dominique pour votre grand soutien, votre grande motivation et votre patience.

Je souhaite remercier les membres de mon jury de thèse, Pr. Sabine Chabrillat, Pr. François Colin, Dr. Dominique Courault et Pr. Christian Walter; pour avoir gentiment accepté d'évaluer mon travail ainsi que d'avoir accepté d'être présents le jour de ma soutenance.

Je souhaite également remercier Anne C. Richer de Forges pour sa grande aide, les échanges à distance avec Emmanuelle et Dominique, qui ont beaucoup aidé à l'élaboration et à la réalisation de ce travail.

Aux membres de mon comité de thèse, Nadia Boukhelifa, Catherine Ottlé, Thomas Eglin et Manuel Martin, merci beaucoup pour les échanges enrichissants qui m'ont beaucoup aidé au début et qui m'ont aidé à structurer mon calendrier de thèse après le confinement.

Je souhaite remercier Eric Ceschia et Didier Michot pour les données de sols et SOC collectées lors des campagnes de terrain.

Un grand merci à Nicolas Baghdadi, Sébastien Lehmann, Songchao Chen et Guillaume Martelet pour leurs précieux retours et leur grande contribution tout au long de ma thèse.

Je tiens à remercier l'ADEME, l'INRAe et l'Université Paris-Saclay pour le soutien et le financement de ma thèse, et plus généralement pour leur engagement à soutenir la recherche de nouveaux talents.

Je remercie également Pierre Larraufie d'avoir toujours été disponible pour toute question relative à mon Ecole Doctorale ABIES, jusqu'au dernier moment avant de terminer ma Thèse et de m'avoir donné des idées pour la rédiger et la finaliser.

Je tiens à remercier Stéphanie et Elisabeth pour leur accueil chaleureux dès le premier instant où je n'ai pu que dire "Bonjour". Sans votre aide, je n'aurais tout simplement pas réussi à apprendre à parler français. Merci beaucoup pour votre grande motivation, votre patience et votre sens de l'humour pendant les cours de français.

Merci à tous les collègues de l'unité de recherche ECOSYS et Agronomie avec qui j'ai travaillé durant ces 3 ans et demi. Remerciements particuliers à Jean-Marc Gilliot et Corentin Barbu pour leur soutien, les échanges scientifiques que nous avons eus à certaines occasions. A Haotian, Florent, Léa, Maxence, Asma, Tristan, Marie-Liesse, Juliette, Israël, Ottone, Charlotte, Tobias pour leur amitié et partenariat au fil des années.

A Pauline, Moza et Patricia, mes collègues de bureau pour leur grand accueil et leur amitié durant ces années. A Madame Marie-Angèle Briand qui m'a accueilli et m'a remis les clés de la maison le premier jour au Château de Grignon pour sa gentillesse et son amitié. A mes colocataires de Grignon, Anaïs, Sixtine et Sonia pour leur amitié, fêtes, apéritifs, raclette et barbecues etc. Une belle opportunité de s'immerger dans la culture française.

Pour la période de confinement, un grand merci à Maria, Pedro, Azariel, Aliia et Luiz pour le temps partagé; c'était vraiment plus facile et plus agréable de passer cette période avec vous.

Je tiens à remercier mes frères en Christ de l'Église évangélique de Plaisir. Surtout Elizabeth, Sophie, Myriam Guilherme, Corentin, Denise, Xavier, Akli, Sabine, Bob et Vero pour leur amitié et leurs prières.

Merci à mes amis et frères en Christ au Brésil. Oswaldo, Fellipe, Pedro, Filipe, Claudia, Nelida, Clecia, Bah, Luh et Nathalia qui, malgré la distance, sont restés en contact pour parler, échanger une idée ou un simple message et même prier.

Merci à Vivian, Edwin, Anaelle et Alice pour leur amitié, balades dans Paris, cinéma et restaurants.

Merci beaucoup à Jorge, Alicia, Gabi d'être la famille qui m'a adoptée comme fils dès le premier jour où j'ai mis les pieds en France à l'aéroport le 27 octobre 2019. Avec vous tout était plus facile, beau et amusant. Merci à Diana pour son amitié du Brésil qui m'a permis de venir dans la maison de cette belle famille.

Enfin, et surtout, ces dernières lignes sont dédiées à ma famille en Colombie et à ma fiancée.

Je tiens à remercier mes parents Jairo et Adriana et ma petite sœur Valentina pour leur amour et leurs prières au cours de ces dernières années depuis que j'ai quitté la maison, malgré la distance nous avons toujours essayé d'être unis et en contact, je vous aime.

Je voudrais également remercier mon oncle Miguel pour ses appels et ses paroles de force et de motivation. Merci à toute ma famille, même si je ne peux pas tous les citer, ils sont dans mon cœur.

Je voudrais remercier ma fiancée et bien-aimée Oleksandra d'avoir toujours été là pour moi, pour son amour, sa foi, son affection et sa douceur. Pour son soutien inconditionnel dans les moments de joie et de difficulté. Tout mon parcours de doctorant a été plus léger et plus gratifiant avec toi à mes côtés.

Enfin, je tiens à remercier chaleureusement toutes les personnes mentionnées et non mentionnées dans ces lignes qui ont contribué directement ou indirectement à rendre ce travail possible.

LIST OF PUBLICATIONS

Posters and oral communications

- Urbina-Salazar, D., Vaudour, E., Baghdadi, N., Ceschia, E., Arrouays, D., 2021. Combined use of Sentinel-2 images and Sentinel-1-derived moisture maps for soil organic carbon content mapping in croplands, South-western France, session SSS10.4, abstract EGU21-8836, vPICO presentation.
- Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Biney, J., Baghdadi, N., Borůvka, L., Wetterlind, J., Arrouays, D., 2022. Mapping topsoil organic-carbon content over cropland regions. The contribution of the Sentinel-2 and Sentinel-1 satellites for Beauce in central France. 22nd World congress of soil science, Glasgow. Poster.
- Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Bialkowski, A., Martelet, G., Baghdadi, N., Wetterlind, J., Van Egmond, F., Arrouays, D., 2023. Potential of using Sentinel-2 temporal mosaics over a 5-year period to map soil organic carbon (SOC) in Beauce, Northern France. 2nd joint Workshop of the IUSS Working Groups Digital Soil Mapping and Global Soil Map "Soil Mapping for a Sustainable Future", Orléans

Scientific articles

- Urbina-Salazar, D., Vaudour, E., Baghdadi, N., Ceschia, E., Richer-de-Forges, A.C., Lehmann, S., Arrouays, D., 2021. Using Sentinel-2 images for soil organic carbon content mapping in croplands of southwestern France. The usefulness of Sentinel-1/2 derived moisture maps and mismatches between sentinel images and sampling dates. Remote Sens. https://doi.org/10.3390/rs13245115.
- Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Martelet, G., Baghdadi, N., Arrouays, D., 2023. Sentinel-2 and Sentinel-1 bare soil temporal mosaics of 6-Year periods for soil organic carbon content mapping in central france. Remote Sensing. https://doi.org/10.3390/rs15092410.

Other co-authored articles

- Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. International Journal

of Applied Earth Observation and Geoinformation 96, 102277. https://doi.org/10.1016/J.JAG.2020.102277

- Vaudour, E., Gholizadeh, A., Castaldi, F., Saberioon, M., Borůvka, L., Urbina-Salazar, D., Fouad, Y., Arrouays, D., Richer-De-Forges, A.C., Biney, J., Wetterlind, J., van Wesemael, B., 2022. Satellite Imagery to Map Topsoil Organic Carbon Content over Cultivated Areas: An Overview. Remote Sensing 2022, Vol. 14, Page 2917 14, 2917. https://doi.org/10.3390/RS14122917

Contents

CHAPTER 1
GENERAL INTRODUCTION

1 GENERAL INTRODUCTION

Soil is a natural resource that is part of the environmental matrices of life. Just as water and air play a fundamental role in the life cycle, soils are a biologically active component that performs multiple functions supporting life on Earth. It uniquely protects the environment and humans by promoting buffering, transformation and filtering capacity between the land surface, the atmosphere and groundwater (Blum, 2005). Soils also provide a wide range of goods and services such as food, fuel, fiber, climate and water regulation, habitat for biodiversity, and recreation that are linked to their different properties (Adhikari and Hartemink, 2016). The United Nations Conference on Sustainable Development in Rio de Janeiro, Brazil in 2012 was the basis for giving rise to the Sustainable Development Goals (SDGs) related to addressing global environmental, political and economic challenges by 2030. Although soil is not explicitly mentioned in these goals, it is included in four targets (2.4; 3.9; 12.4; 15.3) (Bouma et al., 2019) and is intrinsically connected to at least 12 of the 17 SDGs (Table 1) (Lal et al., 2021). The fact that soils are referred to in the targets of the SDGs opens up interdisciplinary challenges for the scientific community to contribute to the development of metrics, in order to propose measurable indicators related to soil quality such as organic carbon stock (Bouma et al., 2019).

The extent of agricultural soils worldwide has doubled the annual expansion rate in the last two decades to 9 million hectares per year (Potapov et al., 2022) and they play fundamental and global roles in food security, climate change and the protection of environmental resources (Lal, 2008; Koch et al., 2013; Paustian et al., 2016). It is therefore essential to obtain rapid and measurable information from field to national or global scales, on soil properties to assist farmers in improving land use management and to contribute to the implementation of current and future public policies concerning soil

(Grunwald et al., 2011; Arrouays et al., 2014a; Montanarella et al., 2016).

Table 1.1 *Sustainable Development Goals in which soil is intrinsically connected and the four targets where soil is mentioned (Adapted from Lal et al., 2021 and Bouma et al., 2019)*

SDG	Focus	Targets	The SDG targets in which soil is specifically mentioned
1	End poverty in all its forms everywhere		
2	End hunger, achieve food security and improved nutrition and promote sustainable agriculture	2.4	By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding and other disasters, and that progressively improve land and **soil quality**
3	Ensure healthy lives and promote well-being for all at all ages	3.9	By 2030, substantially reduce the number of deaths and illnesses from hazardous chemicals and air, water, and **soil pollution**, and contamination
4	Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all		
5	Achieve gender equality and empower all women and girls		
6	Ensure availability and sustainable management of water and sanitation for all		
9	Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation		
11	Make cities and human settlements inclusive, safe, resilient and sustainable		
12	Ensure sustainable consumption and production patterns	12.4	By 2020, achieve the environmentally sound management of chemicals and all wastes throughout their life cycle, in accordance with agreed international frameworks, and significantly reduce their release to air, water and **soil** in order to minimize their adverse impacts on human health and the environment
13	Take urgent action to combat climate change and its impacts		
15	Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	15.3	By 2030, combat desertification, restore degraded land and **soil**, including land affected by desertification, drought and floods, and strive to achieve a land degradation-neutral world
16	Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels		

In this sense, there are three ways to acquire soil data to face current and future challenges: i) field measurements usually through an auger or a soil pit for later chemical analysis or by using proximal sensors such as spectroradiometers, ii) field measurements via proximal sensors and drone-mounted or airborne cameras; and on a larger scale iii) data extraction from remote sensors such as satellites (Figure 1.1). Remote and proximal sensor observations at varied observational scales have been widely used in the last decades to quantify and describe soil attributes (e.g., Stoner and Baumgardner, 1981; Escadafal et al., 1989; Ben-Dor et al., 1997, 2003; Ben-Dor and Banin, 1995; Demattê et al., 2004, 2018; Viscarra Rossel et al., 2006; Gomez et al., 2012; Chabrillat et al., 2014, 2019). Therefore, the increasing use of these sensors covering the visible, near-infrared and mid-infrared regions (VisNIR; 350-2500 nm and MIR; 2500-25000 nm, Figure 1.1) of the electromagnetic spectrum, which are key for the detection of soil properties features, has positioned them as a fast, cost-effective and environmentally friendly alternative for soil monitoring (Nocita et al., 2015; Demattê et al., 2019a). In conjunction with the availability of sensors, computational power and geographic information systems, the importance of soils in a world that demands more and more of this natural resource for food production, decision making, highlights the need to produce rapidly accurate soil maps that provide spatial information for territorial planning through approaches such as digital soil mapping (DSM, McBratney et al., 2003; Minasny and McBratney, 2016). Some examples using DSM (Viscarra Rossel et al., 2015; Mulder et al., 2016a; Padarian et al., 2017; Loiseau et al., 2019) have been developed aiming to map at least one of the properties recommended by the *GlobalSoilMap* initiative (Sanchez et al., 2009; Arrouays et al., 2014a), with soil organic carbon (SOC) and clay being the most frequent over large regions (>10,000 km^2) (Chen et al., 2022).

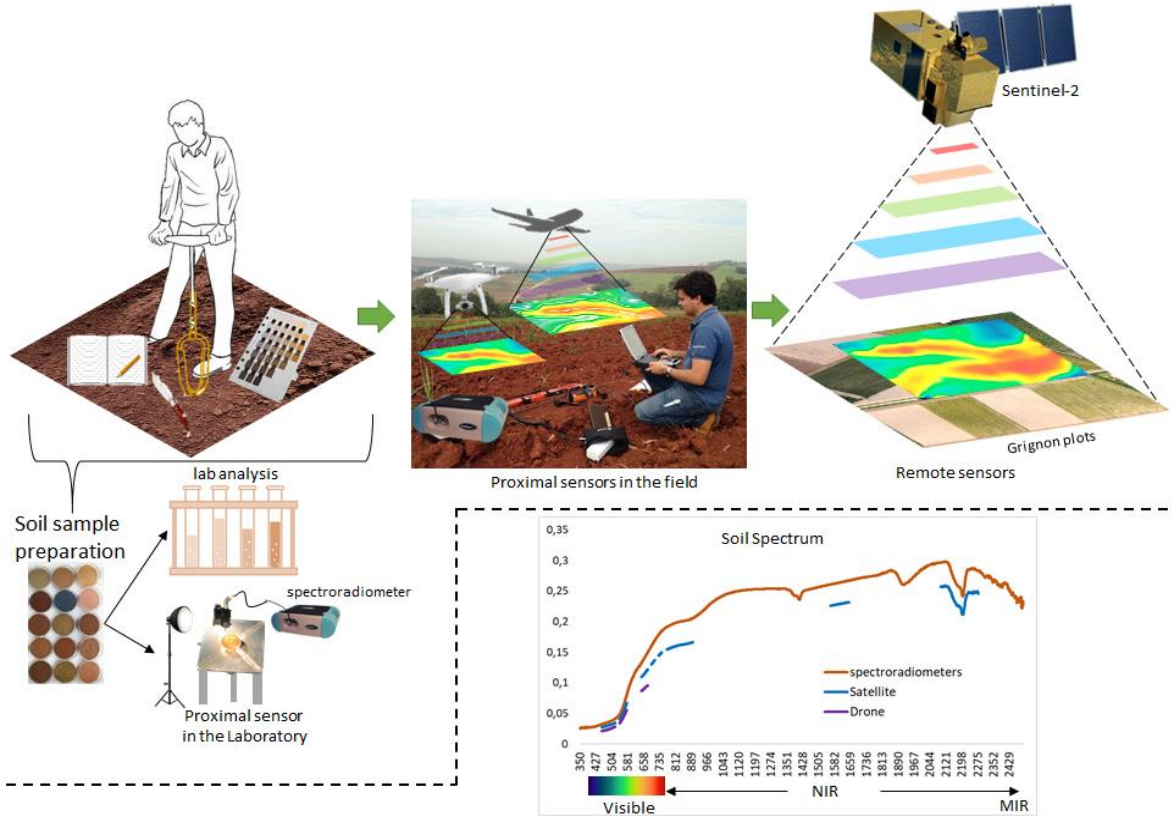


Figure 1.1 *Brief overview of the current ways and equipment used for soil data acquisition and example of the spectral coverage of three different equipments (spectroradiometer, drone and satellite) along the soil electromagnetic spectrum.*

1.1 SOIL ORGANIC CARBON CONTENT PREDICTION AND MAPPING

Two thirds of countries worldwide have mapped soils at a scale of 1:1 million or finer and more of the same amount have not achieved mapping soils even at a scale of 1:1 million (Hartemink et al., 2013). Therefore, there is a need to update the scale of detail of soil attribute and soil class maps as well as generate new ones where they do not exist, to enable decision support under initiatives such as "4 per 1000", which requires determining SOC storage on croplands (Minasny et al., 2017; Arrouays and Horn, 2019). Mapping and knowing the variability of SOC at different spatial scales is fundamental for its monitoring, as it allows a better understanding of the global carbon cycle in a specific

area or region and contributes to the development of diverse environmental applications such as the use of soil carbon sequestration strategies (Mishra et al., 2010; Arrouays et al., 2014a). Some works mapping SOC content and stocks have been carried out from field or farm scales (e.g., Crème et al., 2016; Poeplau et al., 2016), regional scales (Zaouche et al., 2017) up to the national level (e.g., Martin et al., 2011; Poggio et al., 2013; Mulder et al., 2016a) and even at the global level at a resolution of 250 meters (SoilGrids250m, https://soilgrids.org/, v2.0, Poggio et al., 2021) or at 1 km resolution with the global soil organic carbon map (GSOCmap, http://54.229.242.119/GSOCmap/, v1.5.0). In the following sections of this chapter we will give an overview of some topics related to satellite imagery-derived methods for mapping SOC content.

1.1.1 Satellite data overview and main methods for SOC prediction

The first views of the Earth from space obtained from a satellite date back to 1960 with the launch of the first operational meteorological satellite (TIROS-1, Television Infrared Observation Satellite) equipped with two television cameras and two magnetic tape recorders (Sivakumar 2004). In 1972, the first satellite of the Landsat series was launched (Landsat-1 or Earth Resources Technology Satellite "ERTS'') equipped with the first multispectral sensors; Return Beam Vidicon (RBV) and Multispectral Scanner (MSS). Subsequently in 1982, a new generation of Landsat sensor (TM, Thematic Mapper) was put into orbit with the Landsat-4 satellite with a resolution of 30 m, and four years later in 1986 the "*Satellite pour l'observation de la Terre*" (SPOT) equipped with the "*Haute Résolution Visible*" (HRV) sensor with 20 m resolution was launched. Today these two series of satellites continue to operate with the Operational Land Imager (OLI) on Landsat 8 and 9 and with the New AstroSat Optical Modular Instrument (NAOMI) on SPOT 6 and 7. Some pioneering studies on SOC content detection were carried out using the SPOT HRV bands (Agbu et al., 1990; Berthier et al., 2008; Vaudour et al., 2013) and Landsat (Nanni and Demattê 2006;

Huang et al., 2007; Jarmer et al., 2010).

Between 2000 and 2017, aboard the Earth Observing 1 satellite, the Hyperion sensor made available hyperspectral images with 30 m resolution; some SOC modelling approaches via this sensor were used (Gomez et al., 2008; Jaber et al., 2011; Nowkandeh et al., 2013). More recently in 2019 the "*PRecursore IperSpettrale della Missione Applicativa*" (PRISMA) was launched with 239 bands delivering images at 30 m resolution (Loizzo et al, 2018) that have been used to predict SOC at local (Mzid et al., 2022; Angelopoulou et al., 2023) and regional (Gasmi et al., 2022) scales. The Environmental Mapping and Analysis Program (EnMAP) was launched early in 2022 and further works with this satellite mission are expected in the near future, however, studies simulating its bands have already been performed (Guanter et al., 2015; Steinberg et al. 2016; Ward et al., 2020; Chabrillat et al., 2022). In addition, satellite sensors with high spatial resolution mainly used at plot, farm or local level in SOC modelling such as IKONOS with 4 m resolution and PlanetScope with 3 m resolution, are currently operational (Sullivan et al., 2005; Žížala et al., 2019; Silvero et al., 2021a). In the framework of this thesis we focus on the use of multispectral image series acquired from the Sentinel-2A (S2A) and Sentinel-2B (S2B) satellites with a revisit period of approximately 5 days, launched in 2015 and 2017 respectively by the European Space Agency (ESA).

Nowadays satellite SOC prediction is commonly calculated according to three main methods:

I. purely spectral prediction models derived from satellite images (e.g. Landsat, PlanetScope, Sentinel and ASTER) (Žížala et al., 2019; Urbina Salazar et al., 2020).
II. "Bottom up" spectral prediction models, i.e. using spectral information from satellite images in combination with soil spectral libraries (Castaldi et al., 2018), being notably Land Use

and Coverage Area Frame Survey (LUCAS, Orgiazzi et al., 2018) of the European Union, the Mediterranean GEOCRADLE (Tziolas et al., 2019), the Brazilian Soil Spectral Library (BSSL) (Demattê et al., 2019b) or even local spectral libraries (Aïchi et al., 2020).

III. Mixed prediction models, i.e. using satellite imagery or satellite-derived products in combination with non-spectral covariates such as terrain-derived variables (Loiseau et al., 2019; Zhou et al., 2020a; Matinfar et al., 2021).

1.1.2 Main Limitations to detect SOC via satellite

Regardless of the method used, the accuracy that can be obtained from multispectral satellites can be impacted by several factors.

1.1.2.1 Atmospheric effects

The distance between the Earth's surface and the satellites (i.e., more than 700 km for S2 and Landsat 8) poses a challenge in retrieving accurate spectral data. The radiation that is emitted from the Earth to the satellite-borne sensors is subject to significant interaction with the atmosphere which can be a problem when using multispectral satellites for environmental or agricultural purposes such as land use monitoring (Hadjimitsis et al., 2010). As a result, the atmospheric correction error that is expected for the near infrared and shortwave infrared (SWIR) bands, which are key for soil organic matter retrieval (Viscarra-Rossel et al., 2006; Barthès and Chotte 2021), is much higher than for the visible bands (Vaudour et al., 2014). Therefore, the successful retrieval of soil properties may depend significantly on how well the influence of particles or aerosols present in the atmosphere can be accounted for through atmospheric correction methods (Stamnes et al., 2005). Studies on SOC detection have not yet reported effects related to atmospheric correction methods; however, some such as *MAJA* and *Sen2Cor* used for S2 imaging have been studied on

clay prediction performance (Gomez et al., 2022). The directional effects due to soil roughness (Jacquemoud et al., 1992; Cierniewski and Courault, 1993) play a disturbing role in interaction with atmospheric attenuation. Overall, the reflectance signal recorded by the satellite sensor varies according to soil surface condition, defined as the "composition and organization of soil surface at a given time" (Escadafal 1989a), encompassing soil roughness, soil microscopic constituants including water content, the presence of dry vegetation residues and/or chlorophyllian vegetation, the presence of coarse fragments (Girard, 1978; Girard and Girard, 1989,2003; Vaudour et Girard, 2010).

1.1.2.2 Spatial and spectral resolution

The spatial and spectral resolution, i.e. the pixel size and the number of bands capable of discriminating information in the spectrum of a satellite image can impact the retrieval of topsoil attributes (Silvero et al., 2021a). Pixel size is crucial for the retrieval of soil spectral data; satellite images with a coarse spatial resolution (e.g. greater than 30m resolution) may potentially include spectral information from other targets such as vegetation, forests, water bodies, roadsides among others. The spectral mixture of different targets in the pixels impacts the accuracy of the prediction models of a given soil attribute. The number of spectral bands and the width of these bands is also key in the detection of soil attributes since each soil property such as organic matter along the electromagnetic spectrum has a specific spectral response (Ben-Dor and Banin 1995).

1.1.2.3 Soil data used in modelling

Another factor concerning the performance of satellite spectral models may be related to changes in SOC content over time as a result of land management practices (Costa et al., 2021). This involves not only information acquired on historical land use and management in a given area but also the soil samples analysis date used, which if there

is a big difference between the image acquisition date and its respective analysis date may represent confusing information in the modelling. Spatial density i.e. the number of samples per unit area (e.g. km^2 or hectare) questions the spatial and spectral representativeness of the sampling sites location in terms of soil type, crop rotation, tillage practices and amendments among others, which alter the performance of models (Vaudour et al., 2019a,b).

1.1.2.4 Soil surface condition and bare soil availability

Studies using satellite imagery to detect SOC often do not explore the relationship between model performance and soil surface condition. For instance, soil moisture and roughness are two common factors on croplands that impact soil carbon modelling and monitoring. These factors can vary from one plot or region to another over time due to management practices and land use (Vaudour et al., 2019a).

The retrieval of soil attributes as SOC using satellite information leads to another limitation, the availability of bare soil. Bare soil retrieval, in other words the identification of pixels from a satellite image containing pure soil information without any vegetation cover or residue is commonly based on the use of threshold indices such as the Normalized Difference Vegetation Index (NDVI, values lower than 0.35 to 0.20) and the Normalized Burn Ratio (NBR2, values lower than 0.35 to 0.05). The choice of NDVI thresholds does not seem to have a major impact on SOC prediction performance (Vaudour et al., 2021) and thresholds that are too strict for both indices significantly reduce the number of samples available for model calibration and validation (Castaldi et al., 2019a; Vaudour et al., 2021). The use of these indices is key not only for bare soil pixel retrieval on a single image date but also in the production of bare soil composite images from satellite data over a time series that allows the mapping of soil attributes such as SOC over the largest possible percentage of area (Diek et al., 2017; Demattê et al., 2018). However, some difficulties still remain and need

to be studied in depth in relation to the spectral variability of carbon, the effect of the thresholds of the applied indices and the sensors used in different types of local or regional agroecosystems (Vaudour et al., 2021).

1.2 CONTEXT IN FRANCE

Regardless of the fact that France covers the largest exclusive economic zone of the world due to its overseas departments and regions, only the land surface of France, which is much smaller (0.45% of the world's land surface) places her amongst the countries with the largest pedodiversity of the world (Minasny et al., 2010). The spatial variation of soil types in mainland France is quite wide and is determined by the large diversity of parent material across the national land surface. Leptosols, Cambisols, Luvic Cambisols and Haplic Luvisols are some of the dominant soils derived from calcareous rock formations, loamy and clayey materials (Gis Sol. 2011). The used agricultural area (In French, "*superficie agricole utilisée*" (SAU)) in 2019 represented 45 % of the whole country, including mainland France and overseas territories (INSEE, 2021). Production is quite diversified; according to the 2020 agricultural census conducted by municipalities; areas devoted to Multi-crop or poly-livestock and cereals or oil-protein crops represented 32.3 % and 12.4 % respectively, followed by other major field crops (10.4%), livestock farming (8.8%) and viticulture (6.8%) (AGRESTE, 2020) (Figure 1.2). These data represent not only the diversity and extent of agricultural production but also the heterogeneity of soil conditions, climate and indeed of the different soil management practices applied in French agroecosystems.

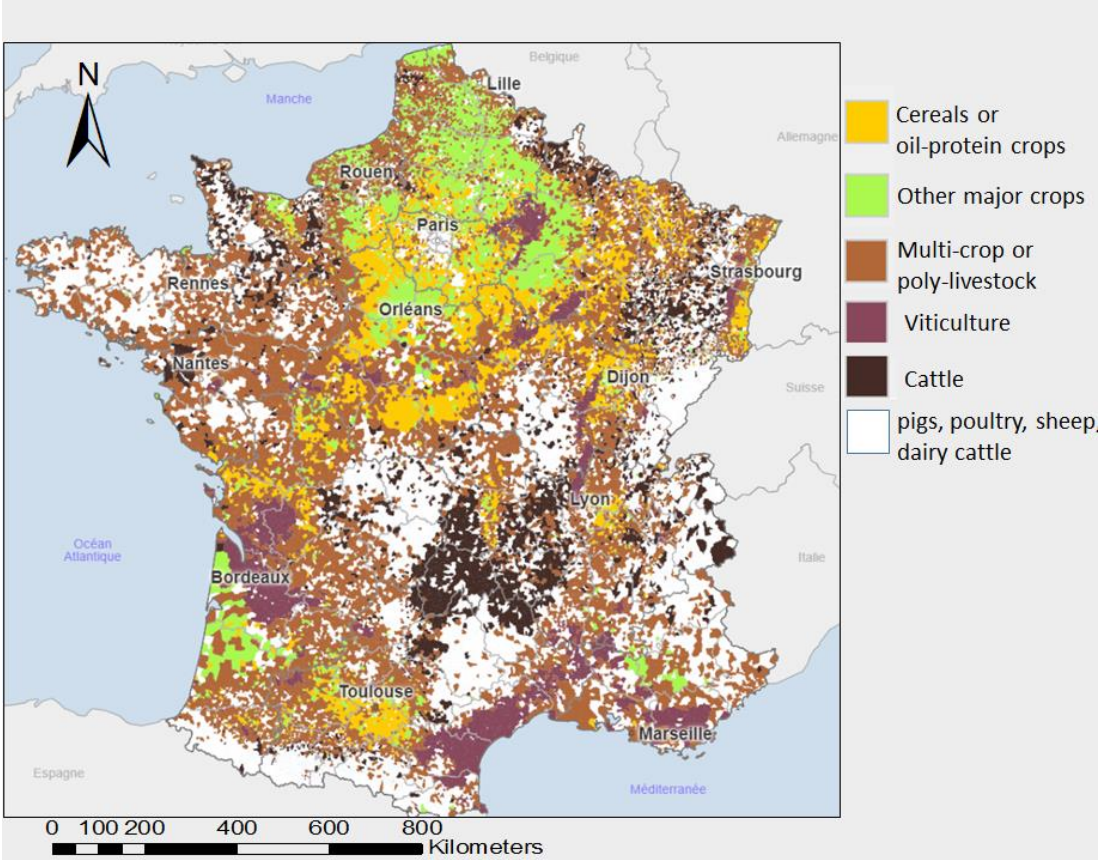


Figure 1.2 *Main agricultural productions in 2020; an agricultural census conducted by municipalities across the mainland French territory (Adapted from AGRESTE, 2020).*

1.2.1 SOC mapping

Since the mid-1990s, digital SOC mapping has been addressed in France (Table 1.3). Arrouays et al. 1995a, published the first top SOC map in a temperate area in southwestern France. These authors were able to map SOC using multiple regression models relating clay, temperature and distance upstream of the sites. This study was then refined using relief attributes (Arrouays et al., 1998) and mapping threshold values of SOC was successfully applied using thematic mapper data on bare soils (Arrouays et al., 1996). Although most other studies have focused on non-spectral methods i.e. they did not consider band data from proximal or remote sensors, in 2008 and 2013 in a small region of Beauce and in the Versailles plain top SOC was

quantified with a field spectroradiometer and SOC maps with a resolution of 20 meters were produced using SPOT 4 satellite data from different dates in spring (Berthier et al., 2008; Vaudour et al., 2013), then from Vis-NIR hyperspectral AISA-Eagle airborne data with a resolution of 1 m (Vaudour et al., 2016). Other studies were performed in small regions using land use and terrain data (Lacoste et al., 2014a); and only S2 satellite spectral data (Vaudour et al., 2019a; 2019b; 2021). Interestingly the national scale SOC maps (i.e. mapping of mainland France) have all been carried out through models that considered soil, climate, terrain data and satellite derived products (e.g. SPOT 4 satellite derived land use map and MODIS satellite derived net primary production map) but did not include direct spectral data of bare soil (Martin et al., 2011; Meersmans et al., 2012; Lacoste et al., 2014b; Mulder et al., 2016a; Chen et al., 2019a). The use of the S2 satellite series for SOC mapping in croplands has a long way to be applied, findings related to the comparison between single date prediction performances (Vaudour et a., 2013; 2019a; 2019b) and to the potential of the use of composite images of bare soil on French territory (Vaudour et al., 2021) leave a clear message to further explore the capability of satellite sensors. Table 1.3 shows that the scale of the maps at the national level has been improved from 12 km (Martin et al., 2011) to 90 m in the most recent studies (Mulder et al., 2016a; Chen et al., 2019a) fulfilling the requirement of the *GlobalSoilMap* project (Sanchez et al., 2009; Arrouays et al., 2014a). Over a few large regions (up to 10,000 km^2) and very large regions (>10,000 km^2), SOC content has also been mapped in Languedoc-Roussillon, Brittany and Alsace with a resolution similar to national scale maps (Vaysse and Lagacherie, 2015,2017; Ellili-Bargaoui et al., 2020; Lemercier et al., 2022). In that sense, exploring the use of satellite sensors such as S2 not only aiming to map at a more detailed scale but also to assess the relationship between the performance of prediction models, natural or human effects, and the spectral and spatial variability of SOC on croplands in France should be further explored.

Table 1.3 *SOC prediction and mapping in France*

Study area	Year	Landuse	Area (km^2)	Sample size	collection depth cm	Data used	Method	Map resolution	Reference
small region									
Montours	2007	Temporary meadow	<1	1162	0-5;10-15;40-45;50-55	SL+TC	non-spectral		Follain et al., 2007
Beauce Chartraine	2008	Croplands	26	27	0-10	SL+PS+RS	purely spectral	20 m	Berthier et al., 2008
Versailles plain	2013	Croplands	221	128	0-8	SL+PS+RS	purely spectral	20 m	Vaudour et al., 2013
Brittany	2014	Several landuses	10	70/49	0-30; 30-105	SL+TC	non-spectral	2 m	Lacoste et al., 2014a
	2016		<1			SL+CL+TC	non-spectral	2 m	Lacoste et al., 2016
Versailles plain	2016	Croplands	221	267	0-8	SL+RS	purely spectral	2 m	Vaudour et al., 2016
Versailles plain	2017	Croplands	221	253	0-8	SL+TC	non-spectral	25 m	Zaouche et al., 2017
Versailles plain	2019	Croplands	221	72	0-8	SL+RS	purely spectral	20 m	Vaudour et al 2019b
La Peyne	2019	Vineyards	48	104	0-5	SL+RS	purely spectral	ng	Vaudour et al 2019b
Pleine-Fougères	2019	Croplands, grasslands, forest	10	64	0-7.5;7.5-15;15-30;30-45	SL+TC	non-spectral	10 m	Ellili-Bargaoui et al., 2019
Versailles plain	2019	Croplands	221	199	0-8	SL+RS	purely spectral	ng	Vaudour et al., 2019a
Lusignan	2020	Croplands and grasslands	<1	400	0-30;30-60;60-90	SL	non-spectral		Crème et al., 2020
Versailles plain	2021	Croplands	221	268	0-8	SL+RS	purely spectral	20 m	Vaudour et al., 2021
Large region									
Pyrenean piedmont	1995	Forests		194	0-30	SL+CL+TC	non-spectral	1 km	Arrouays et al., 1995a
Beauce	2023	Croplands	4838	391	0-10	SL+TC+RS	mixed	25 m	4th chapter of this thesis
Very large region									
Languedoc-Roussillon	2015 2017	all landuses	27236	1986 + RMQS (105)	0-5; 5-15; 15-30; 30-60	SL+CL+TC+RS	non-spectral*	100 m	Vaysse and Lagacherie 2015,2017
Brittany	2020	all landuses	27040	1020 strata	0-5 ;5-15 ;15-30 ;30-60 ;60-100 ;100-200	SL	non-spectral	50 m	Ellili-Bargaoui et al., 2020
Pyrenean piedmont	2021	Croplands	22177	625	0-10	SL+RS	purely spectral	10 m	3rd chapter of this thesis
National									

Entire France	2011	all landuses	RMQS (1.974)	0-30	SL+CL+RS	non-spectral*	12 km	Martin et al., 2011
Entire France	2012	all landuses	RMQS (2.158)	0-30	SL+CL+RS	non-spectral*	250 m	Meersmans et al., 2012
Entire France	2014	Forest	DoneSol (228); RMQS (131); other (115)	O Horizon	SL+CL+TC+RS	non-spectral*	90 m	Lacoste et al., 2014b
Entire France	2016	all landuses	RMQS and DoneSol (34.295)	0-5; 5-10; 15-30; 30-60; 60-100; 100-200	SL+CL+TC+RS	non-spectral*	90 m	Mulder et al., 2016a
Entire France	2019	Croplands	RMQS (1.089)	0-30; 30-50	SL+CL+RS	non-spectral*	90 m	Chen et al., 2019a

*SL: soil legacy data; PS: proximal sensing data; RS: remote sensing data; TC: Terrain covariates; CL: climate data; *: no direct spectral data were used, but satellite-derived products were used.*

1.3 FRAMEWORK AND OBJECTIVES OF THE THESIS

This thesis was carried out in the framework of two main projects: at national level, the PEDO_POLYPHEME project of the TOSCA program funded by the CNES (2020-2022), and at the European level, the *Stimulating Novel Technologies from Earth Remote Observation to Predict European Soil Carbon* Project (STEROPES; https://ejpsoil.eu/soil-research/steropes (accessed on 15 February 2023)) which aims to use satellite time series to predict the SOC content of agricultural soils over different soil-climatic conditions and cropping systems across Europe (2021-2024). STEROPES, in addition to building models from the satellite time series spectra included in this thesis, also aims to analyze the influence of factors such as soil moisture, texture, management practices, crop residues and soil salinity on SOC prediction performance. This thesis was also supported by Programme National de Télédétection Spatiale through TELEMOS project (PNTS, http://programmes.insu.cnrs.fr/pnts/), grant n° PNTS-2020-17.

1.3.1 Objectives and structure of the thesis

We have seen that SOC prediction by remote sensing techniques has been widely accepted and used worldwide in the framework of common efforts and global initiatives (e.g. "*4 per 1000*" and "*GlobalSoilMap*") due to the importance that this soil attribute represents in SDGs such as achieving food security, promoting sustainable agriculture and mitigating climate change (Arrouays and Horn, 2019). However, several questions emerge about the capability to detect SOC via satellite imagery in diverse agroecosystems where SOC availability may be strongly influenced by pedoclimatic conditions and human practices that in turn may affect the predictive potential of satellite-borne sensors. Therefore, an initial question is, ***which is the optimal period to acquire satellite images for SOC estimation?*** This is a frequent inquiry in soil attributes detection and is mainly

associated with climatic or seasonal factors specific to each study area. Rainfall frequency, cloud cover, temperature and even management practices such as tillage are factors that change soil surface condition and may hinder good performance in prediction models. As it is well known SOC is characterized by having a labile pool that can be readily decomposed by organisms over time. In most cases the satellite image dates do not necessarily coincide with the soil sampling periods, therefore, this poses the following question: ***can differences between the SOC sampling periods and the date of satellite imagery influence the model performances?*** This, as well as the first question, depends on several factors that may vary from one area to another.

The acquisition of satellite image series in periods with ideal conditions to obtain as much exposed soil as possible has been addressed with bare soil composite imaging techniques. However, few such methods have been explored to predict SOC in agroecosystems at different scales. In that context it is worth asking: ***What strategy should we use to retrieve bare soil over a satellite time series in order to improve SOC prediction in a given area?*** Finally, following this same scenario and considering a mixed method (see section 1.1.1), ***what would be the most relevant covariates to include in the models?*** Although the list of questions may actually be longer; due to the complexity that SOC remote sensing can represent, these are the main questions that we address in this thesis. It should be noted that during the development of this document and in the last chapter we discussed these questions and exposed some remaining questions.

The objective of this thesis project is to evaluate the potential of Sentinel image time series for SOC monitoring in French mainland agroecosystems. Therefore, we address SOC prediction and mapping through reflectance-derived models, particularly based on spectral data from Sentinel-2 (S2) and Sentinel-1 (S1) satellites and from spatio-spectral models based on S2 and S1 data and terrain-derived covariates at various regional scales in France.

To achieve this objective, we will aim to:

- collect a large amount of Sentinel imagery from 2016 to 2021.
- determine the optimal acquisition periods of the images for SOC prediction.
- develop a composite image series method to extend the top bare soil coverage in order to increase the prediction area.
- implement models that combine multi-date images and other relevant covariates to produce exhaustive spatial maps together with maps of uncertainty.

This manuscript is organized into five chapters. The first one presents an introduction on the general overview and importance of SOC prediction and mapping. It focuses mainly on SOC detection via satellite as well as on the progress of SOC mapping at national and regional level in mainland France. The second presents the study areas, the main data sets used and shows a preliminary analysis of one of the study areas. The third chapter deals with the evaluation of the potential of S2 and derived S1/2 products to predict SOC, using single date imagery in spring during 2017 and 2018 in a large agricultural region in southwestern France (Urbina-Salazar et al., 2021). The fourth one addresses the use of bare soil composite S2 image series for periods between 2016 and 2021 and a set of covariates over a region of intense agriculture in the southwest of Paris (Urbina-Salazar et al., 2023). Finally, in the last chapter we present the discussion and general conclusions of this thesis. It is important to note that chapters 3 and 4 are presented in article format.

CHAPTER 2
STUDY AREAS, DATA AND A PRELIMINARY ANALYSIS

2 STUDY AREAS, DATA AND A PRELIMINARY ANALYSIS

This chapter aims to introduce the different study areas that were used for the development, validation and application of SOC detection models in this thesis. It will also address the description of the soil and satellite data and field observations. At the end, a preliminary analysis of one of the study areas is presented, which served as the basis for the development of the following chapters.

2.1 STUDY AREAS

The development of this work is focused on three different regions in terms of soil types, climatic conditions and landforms located in three regions in France (Brittany, Occitanie and Centre Val de Loire) (Figure 2.1).

2.1.1 Pleine-Fougères

The first is the Zone Atelier Armorique (Pleine-Fougères), which is part of the European Long-Term Ecosystem Research Network. The area covers 10 km^2 and is located in the Brittany region in northwestern France (48°31′N-48°29′N; 1°35′W-1°32′W), which includes, continuous grassland plots, natural areas, annual crop rotation (mainly maize, wheat and barley), annual crop to grassland rotation and vice versa. The main soils in the region are Cambisols, Luvisols, Leptosols and Fluvisols developed from alluvial and colluvial deposits (IUSS Working Group WRB., 2015). These soils are developed on rather diversified geological formations across the following landscape units: a plain on soft schist at the northern end of the area; a plateau on granite (southern extreme); and a hillside on hard schist that presents a transition between granite and soft schist (Lacoste et al., 2014a). This area was chosen to perform the first spectral model tests of this project.

2.1.2 Pyrenean piedmont

This is a large area (22,177 km^2) located in the Occitanie region of southwestern France (43°57′N-42°48′N; 0°15′W-1°51′E). More than 50 % of the total area are croplands, where mainly wheat, sunflower and maize are sown. For the elaboration of the spectral models, only crop parcels were considered through the Land Parcel Identification System (LPIS) declared by farmers at the national level in 2017. The soils of this region mainly originate from molassic deposits that through time and the flow of rivers during the Quaternary formed alluvial terraces from North to South. The main soils found on these terraces and along the area are Calcaric Cambisols, Coluvic Regosols, Fluvisols, Haplic Luvisols, Umbric Leptosols and Haplic Cambisols (Redon et al., 2013).

2.1.3 Beauce

Beauce is located in the Centre Val de Loire region southwest of Paris (48°35′N-47°31′N; 1°0′E-2°28′E), covering 4,838 km^2. According to the LPIS in 2019 approximately 84 % of the entire area are croplands. Winter cereals such as wheat with rotations of rapeseed, sugar beet and maize are the predominant crops grown in the area (Bouarfa et al., 2011). Cambisols and Luvisols are the main soil classes observed (IUSS Working Group WRB., 2015).

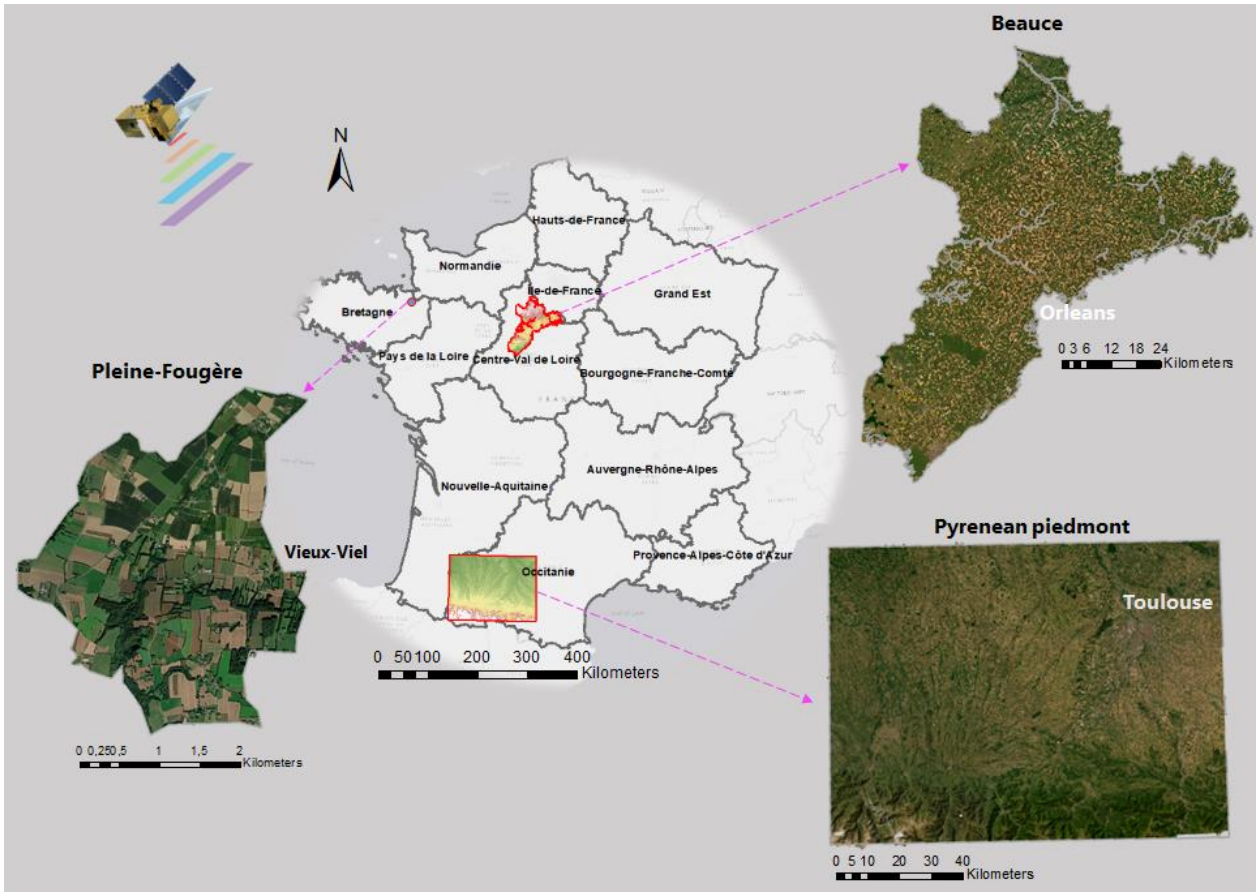


Figure 2.1. *Study areas location in mainland France*

2.2 LEGACY SOIL DATA

From this section we present the data sets used in this thesis. Table 2.1 summarizes the data sets available for each of the study areas. More specific information regarding the datasets used are given in the following chapters.

Table 2.1 *Data available for the study areas*

		Pleine-Fougères	Pyrenean piedmont	Beauce
Soil Legacy	DoneSol		420 sites	341 sites
	LUCAS-Soil		114 sites	50 sites
	RMQS			12 sites
	SSC	111 sites	91 sites	
	Soil map	x*		
Parent material	PMM			x**
	AGRI			x***
Remote sensing Data	S2 images	x	x	x
	SMPs		x	x
	S2****			x
Terrain-derived variables	Landform map		x	
	DEM		x	x
	Others		x	x

*SSC : Soil sampling campaigns; PMM : parent material map; AGRI: airborne gamma ray images; SMPs: soil moisture products; *: Lacoste et al., 2011; ** : King et al., 1994; *** : Martelet et al., 2014; **** : Indices derived from S2 images.*

2.2.1 DoneSol

The French soil profile database (DoneSol) is the result of a collective work that structures and groups soil profiles and auger surveys (approximately 160°000 point data) from conventional mapping programs at all scales at the national level (Inventaire Gestion et Conservation des Sols, IGCS) (Arrouays et al., 2020a). The points are distributed irregularly across mainland France. For each point it provides information such as the coordinates, the sampling date, the description of the profile and horizon, soil types, the results and methods used for granulometric and chemical analysis among others. (to see more details, visit https://www.gissol.fr/outils (accessed on 14 February 2023)).

2.2.2 RMQS

The Soil Quality Measurement Network (in French "*réseau de mesure de la qualité des sols*", RMQS) is a network dedicated to the long-term monitoring of soils in France; it aims to assess status and changes in soil under natural and human factors such as atmospheric deposition, climate change, agricultural practices, and land use and management (Gis Sol. 2011). Since 2000, the RMQS has been based on the monitoring of 2200 georeferenced sites every 15 years, distributed on a grid (16km x 16km) covering all France. The first campaign (RMQS1) was carried out between 2000 and 2009 reaching the sampling of 2170 sites, a second campaign (RMQS2) currently in operation started in 2016 and will be carried out until 2027. Some analyses such as particle size distribution and SOC content were determined for the 0-30cm and 30-50cm layers at these sites. In addition with the results obtained from RMQS1 some trace elements such as cadmium, cobalt, chromium and zinc were mapped and are available (see, https://www.gissol.fr/le-gis/programmes/rmqs-34 (accessed on 15 February 2023)).

2.2.3 LUCAS

The Land Use and Land Cover Survey (LUCAS) also has as its main objective to monitor changes in land use and land cover over time, with the difference that it is carried out across the European Union (EU). This program is managed by the Statistical Office of the European Union (EUROSTAT), which since 2006 has been conducting surveys every three years through a 2km x 2km grid (approximately 1,000,000 georeferenced points in EU) to cover all EU member states (Jones et al., 2020). A subset of these points (LUCA-Soil) is also used to measure some topsoil properties such as SOC in approximately 20,000 points in the EU.

2.2.4 Soil sampling campaigns

Two field campaigns were conducted in 2009 and 2016 in Pleine-

Fougères. The first one was based on 111 points covering different land cover types such as crops, bare soil, forest and grassland. In 2016, 64 points were re-sampled from the 111 points collected in 2009 as part of a previous project aiming to determine SOC changes over 7 years (Ellili-Bargaoui et al., 2019). In the Pyrenean piedmont a field campaign was carried out between May and June 2018 in partnership with the Center for Spatial Studies of the Biosphere of Toulouse (CESBIO) and the unit of functional ecology and ecotoxicology of agroecosystems (EcoSys). A total of 91 topsoil samples were collected from croplands.

Note that DoneSol, LUCAS and samples collected in the field campaigns were used for spectral modelling and cross-validation. RMQS was only considered as an independent validation set of the predicted values in Beauce.

2.3 SATELLITE DATA

In this thesis we address the use of multispectral image series S2A and S2B acquired between 2016 and 2021 with a resolution of 10 to 20 meters using the Multispectral instrument (MSI, 13 spectral bands) (Table 2.2). Soil moisture products (SMPs) derived from S1 and S2 were also used. The S2 images and SMPs were acquired from the Muscate platform of the French Ground Data Center (Theia) (https://www.theia-land.fr/ (accessed on 15 February 2023)).

Table 2.2 *Sentinel-2A and Sentinel-2B bands*

Band Number	Sentinel-2A		Sentinel-2B		Spatial resolution (m)
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	
Band 1	442.7	20	442.3	20	60
Band 2	**492.7**	**65**	**492.3**	**65**	**10**
Band 3	**559.8**	**35**	**558.9**	**35**	**10**
Band 4	**664.6**	**30**	**664.9**	**31**	**10**
Band 5	**704.1**	**14**	**703.8**	**15**	**20**
Band 6	**740.5**	**14**	**739.1**	**13**	**20**
Band 7	**782.8**	**19**	**779.7**	**19**	**20**
Band 8	**832.8**	**105**	**832.9**	**104**	**10**
Band 8a	**864.7**	**21**	**864.0**	**21**	**20**
Band 9	945.1	19	943.2	20	60
Band 10	1373.5	29	1376.9	29	60
Band 11	**1613.7**	**90**	**1610.4**	**94**	**20**
Band 12	**2202.4**	**174**	**2185.7**	**184**	**20**

Bands used in this thesis are shown in bold characters (Adapted from: https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument)

2.4 APPROACHES USED FOR SOC DETECTION

In the first chapter we mentioned the main methods used for SOC estimation using satellite imagery. Following these methods, we first explored the use of purely spectral models in Pleine-Fougères (Table 2.3). Obviously, the number of available samples on croplands in bare soil areas per date could limit this approach based on S2 single date images.

Table 2.3. *Methods, approaches, selection of bare soil pixels and variables used in the models for the study areas*

			Bare soil pixels selection					Variables used				
Study area	Method	Approach	no clouds	Croplands	NDVI	NBR2	SM	S2 bands	S2*	SMPs	terrain derived	AGRI
Pleine-Fougère	purely spectral	Single date	x	x	x			x				
Pyrenean piedmont	purely spectral and mixed	single date	x	x	x			x		x	x**	
Beauce	mixed	Bare soil compo-site	x	x	x	x	x	x	x	x	x	x

*SMPs : soil moisture products; AGRI: airborne gamma ray images; *: Indices derived from S2 images; ** : not included as a covariate but used for other analyses such as the relationship between spatial predicted values and residuals*

Bare soil availability in satellite imagery for SOC detection has currently been addressed using single date imagery in other regions of Europe (Gholizadeh et al., 2018; Castaldi et al., 2019a; Vaudour et al., 2019a-b; Dvorakova et al., 2020; Biney et al., 2021); Asia (Zhou et al., 2020a; Wang et al; 2021); America (Sayão et al., 2018; de Castro-Padilha et al., 2020); in the Middle East (Mohamed et al., 2020; Mirzaee et al., 2016) and Oceania (Gomez et al., 2008). Recently it has been explored using techniques that aim to extend the bare soil surface from satellite imagery time series (e.g., the use of bare soil composite images) (Dvorakova et al., 2021; Silvero et al., 2021b; Vaudour et al., 2021; Zepp et al., 2021). Both approaches have been successfully used. However, prediction performances are directly related to satellite image acquisition dates, soil surface conditions such as soil roughness and soil moisture; which are easily altered by management practices or natural events (Vaudour et al., 2019a; Gomez et al., 2022; Rienzi et al., 2014). The SOC variability and its correlation with other soil attributes in a given area also influence the spectral modelling (Vaudour et al., 2019b). In that sense as well as in Pleine-Fougère, in Pyrenean piedmont a purely spectral approach based on single date images was addressed with the difference that soil moisture data was included as

a covariate in the models (Table 2.3). Finally in Beauce, mixed prediction models were used including different covariates derived from terrain, SMPs and S2-derived indices; based on composite images of bare soil over a 6-year period (Tabla 2.3). These last two areas are addressed in detail in the following chapters of this thesis.

2.5 PRELIMINARY ANALYSIS

Pleine-Fougères was selected to perform the first tests of spectral models elaborated from S2 images. Before starting the modelling, it was noticed in this study area that the soil sampling campaigns were performed considering different land cover types and not only croplands targeted by this project. Some of the main land covers observed were croplands, continuous and rotational pastures, bare soil, bare soil with residues and wooded area (Figure 2.2).

The spectral model building required choosing an algorithm that would fit our collected data set. The choice of algorithms depended on different aspects such as the number of soil samples available for model calibration; the number of predictive covariates; linear or non-linear relationships between SOC and the available covariates, among others. Considering that this thesis focused on the use of satellite data and that the availability of soil samples would depend directly on the land cover type at the sampling points (i.e. on bare soils in croplands) and on favorable conditions of the acquired S2 images (e.g. cloud-free images), we decided to use at first partial least squares regression (PLSR, Geladi et al., 1986) for both Pleine-Fougères and Pyrenean piedmont. This algorithm has the advantage of being less prone to overfitting in modeling with reduced sample sets and has been one of the most widely used algorithms in the last decade to detect SOC via satellite (Vaudour et al., 2022).

In Beauce, due to the larger number of data, i.e. not only the S2 image bands but also the number of legacy data and terrain-derived covariates calculated, we use quantile regression forest models (QRF). QRF is an algorithm that is based on random forests (RF) and offers the advantage of estimating the full distribution of predicted values (Meinshausen, 2006).

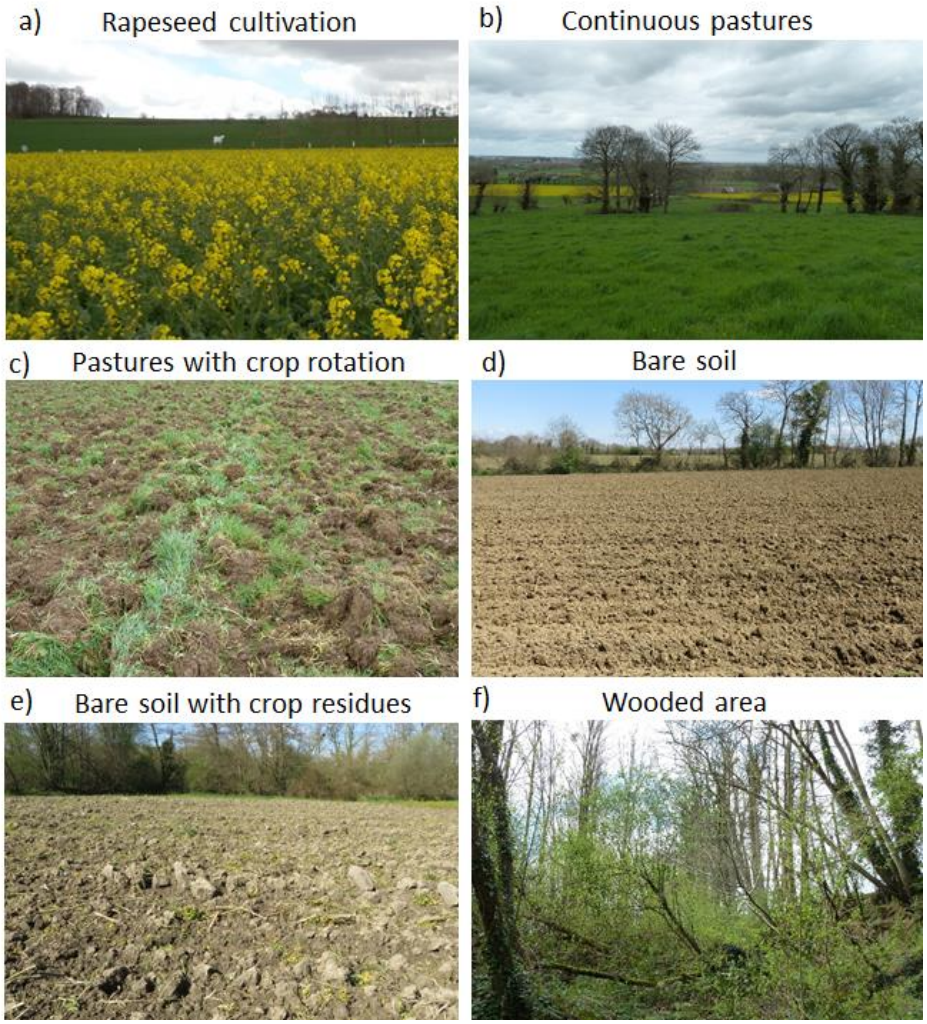


Figure 2.2. *Land use coverage in Pleine-Fougère during the field campaign conducted in April 2016. Photo Credits: Didier MICHOT*

In other words, unlike RF which only provides the mean of the predicted values; QRF calculates quantiles (e.g., 0.05 and 0.95) that can be used as prediction intervals, adding complementary information to

the maps produced and to the understanding of SOC variability and model performance in the study area. However, in this section we only show preliminary results in Pleine-Fougères that helped us to establish the methods and approaches in this Thesis project.

In order to obtain spectral data from the S2 bands as close as possible to the soil sampling performed in spring 2016, S2 images were acquired from the springs of 2016 and 2017. However due to the limited availability of images on the THEIA website and the presence of clouds in Pleine-Fougères it was only possible to obtain them on four different dates between March and May 2017. To extract bare soil pixel spectra, the NDVI index was used, masking all areas with a pixel value less than 0.3. The use of other indices such as NBR2 was not considered in this area due to the limited number of samples available; the use of thresholds of this or any other index would have significantly reduced the set of samples available for the regression models.

Figure 2.3 displays the results of the PLSR models obtained on four different dates. As expected the number of samples in bare soils captured by the S2 images (i.e. soil samples located in the pixels that are bare soil on the acquisition date) available per date was quite low, particularly for the S2 image of March 13, 2017 (3 soil samples). On the date with the highest percentage of bare soil in the whole area (11 %) the model was elaborated with only 25 soil samples.

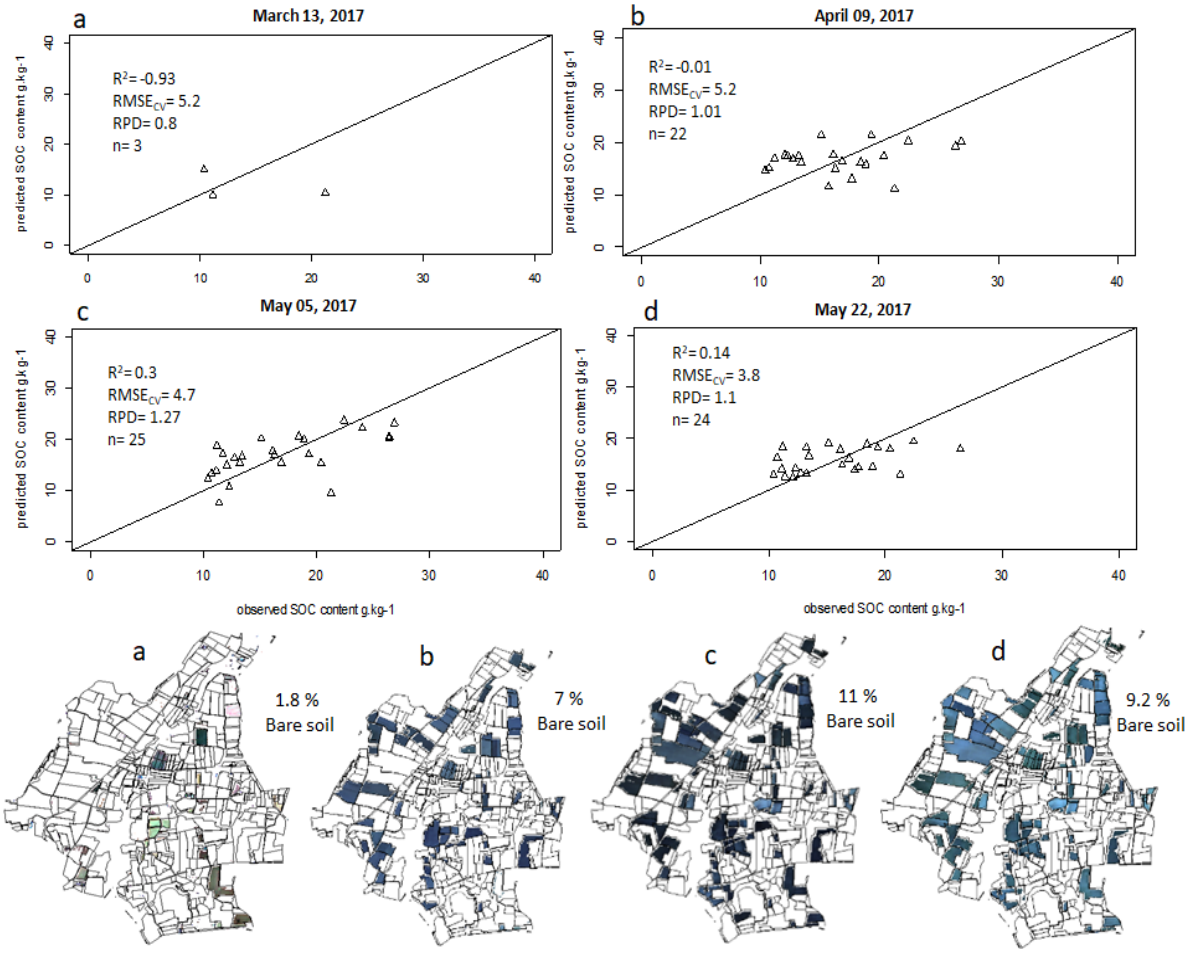
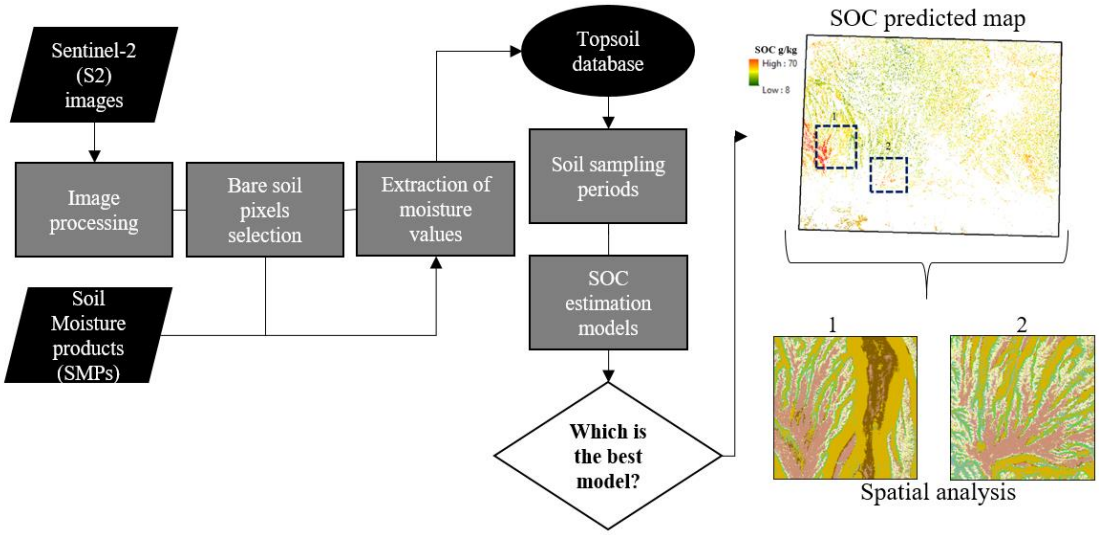


Figure 2.3. *Performance of PLSR models and percentage of bare soils for four dates in Pleine-Fougère*

CHAPTER 3



Graphical Abstract

3 USING SENTINEL-2 IMAGES FOR SOIL ORGANIC CARBON CONTENT MAPPING IN CROPLANDS OF SOUTHWESTERN FRANCE. THE USEFULNESS OF SENTINEL-1/2 DERIVED MOISTURE MAPS AND MISMATCHES BETWEEN SENTINEL IMAGES AND SAMPLING DATES

Diego Urbina-Salazar [1,4], Emmanuelle Vaudour [1], Nicolas Baghdadi [2], Eric Ceschia [3], Anne C. Richer-de-Forges [4], Sébastien Lehmann [4] and Dominique Arrouays [4]

[1] *UMR EcoSys, AgroParisTech, INRAE, Université Paris-Saclay, 78850 Thiverval-Grignon, France*

[2] *UMR TETIS, INRAE, University of Montpellier, UMR TETIS, 500 rue François Breton, 34093 Montpellier CEDEX 5, France*

[3] *CESBIO, Université de Toulouse, CNES/CNRS/INRAE/IRD/UPS, 31400 Toulouse, France*

[4] *Info&Sols Unit, INRAE, CEDEX 2, 45075 Orléans, France*

Urbina-Salazar, D., Vaudour, E., Baghdadi, N., Ceschia, E., Richer-De-forges, A.C., Lehmann, S., Arrouays, D., 2021. Using sentinel-2 images for soil organic carbon content mapping in croplands of southwestern france. The usefulness of sentinel-1/2 derived moisture maps and mismatches between sentinel images and sampling dates. Remote Sens. 13. https://doi.org/10.3390/rs13245115

Keywords: soil organic carbon; sentinel-2; soil moisture; croplands; digital soil mapping; southwestern france; topographic wetness index; slaking crust sensitivity index

3.1 ABSTRACT

In agronomy, soil organic carbon (SOC) content is important for the development and growth of crops. From an environmental monitoring viewpoint, SOC sequestration is essential for mitigating the emission of greenhouse gases into the atmosphere. SOC dynamics in cropland soils should be further studied through various approaches including remote sensing. In order to predict SOC content over croplands in southwestern France (area of 22,177 km^2), this study addresses (i) the influence of the dates on which Sentinel-2 (S2) images were acquired in the springs of 2017-2018 as well as the influence of the soil sampling period of a set of samples collected between 2005 and 2018, (ii) the use of soil moisture products (SMPs) derived from Sentinel-1/2 satellites to analyze the influence of surface soil moisture on model performance when included as a covariate, and (iii) whether the spatial distribution of SOC as mapped using S2 is related to terrain-derived attributes. The influences of S2 image dates and soil sampling periods were analyzed for bare topsoil. The dates of the S2 images with the best performance (RPD ≥ 1.7) were 6 April and 26 May 2017, using soil samples collected between 2016 and 2018. The soil sampling dates were also analyzed using SMP values. Soil moisture values were extracted for each sample and integrated into partial least squares regression (PLSR) models. The use of soil moisture as a covariate had no effect on the prediction performance of the models; however, SMP values were used to select the driest dates, effectively mapping topsoil organic carbon. S2 was able to predict high SOC contents in the specific soil types located on the old terraces (mesas) shaped by rivers flowing from the southwestern Pyrénées.

3.2 INTRODUCTION

Soil organic carbon (SOC) is fundamental to the global carbon cycle. Croplands represent approximately 143.4 Pg of SOC stocks worldwide at a depth of 30 cm, i.e., 20.7% of all land cover types followed by forests and grasslands with 43.1% and 25.9%, respectively (FAO and ITPS, 2020). Agricultural production and land use change account for 24% of global greenhouse gas emissions (Smith et al., 2014). When converted from forest or native vegetation, cultivated soils lose between 20 and 50% of the carbon content in their top layers (0–30 cm), with the highest rates of loss in the first years of disturbance (Schlesinger, 1984; Arrouays and Péelissier, 1994; Jolivet et al., 1997). Therefore, SOC sequestration-related practices on croplands can potentially help ensure soil health and have, thus, been proposed as a temporary solution for mitigating climate change (Minasny et al., 2017; Chenu et al., 2019).

Monitoring changes in SOC is essential, and the use of field or remote sensors to monitor carbon over croplands is highly relevant to this process (Smith et al., 2020; Costa et al., 2021). At global, continental or national scales, mapping the spatial variability of SOC content (Hengl et al., 2015; Mulder et al., 2016b; Poggio et al., 2021) and SOC stocks has been possible using spatial models that include environmental covariates such as those derived from climate and terrain using MODIS satellite imagery and SRTM radar systems (Guevara et al., 2020; Hengl et al., 2014). Such technology is necessary because of increases in arable land due to food demand. Therefore, remote sensing techniques are essential for SOC monitoring in croplands because information regarding soil properties is important for decision-making. The use of spectral data on croplands to predict SOC has been studied by using laboratory/field sensors (Gomez et al., 2012; Stevens et al., 2008), airborne imagery (Stevens et al., 2010; Ou et al., 2020) and satellite imagery (e.g. Castaldi et al., 2019a; Vaudour et al., 2019a,b; Urbina-Salazar et al., 2020; Vaudour et al., 2021); by comparing satellite

and airborne imagery (Castaldi et al., 2019b) or using them together in a common approach (Vaudour et al., 2016); by comparing satellite, airborne and unmanned aerial system-based imagery (Žížala et al., 2019; Biney et al., 2021); and by combining different satellite imagery with satellite-derived covariates and indices over wide areas with different land uses: cropland, pasture and forest (Zhou et al., 2020b; Silvero et al., 2021b). The performance of these approaches for European croplands has demonstrated the ability of the new, free and open-access optical sensor Sentinel-2 (S2) to map and/or to detect changes in SOC (Castaldi et al., 2019b; Vaudour et al., 2019a; Žížala et al., 2019).

The use of the S2 satellite for the quantification of soil attributes has increased due to its characteristics, which include high spectral and spatial resolution. However, factors such as the date of satellite image acquisition, the solar elevation angle on the date of satellite image acquisition, the soil moisture (SM) and the surface soil roughness should be considered and studied in detail in order to predict SOC as well as other attributes. Vaudour et al. (2019a) observed that the performance of S2 images when calculating SOC varied with image acquisition date. Furthermore, these authors found that the best performances (RPD values ≥ 1.4) were mainly related to the solar elevation angle, the soil roughness and the SM. The development of algorithms to obtain soil information using remote sensing could support the study and mapping of soil attributes, e.g., the use of SM products (SMPs) at a sub-plot scale. The SMPs used in our study were obtained through a synergistic combination of Sentinel-1 and -2 (S1/S2) images; these products were generated for croplands and grasslands with or without vegetation cover with an estimated accuracy of 5% vol. (El Hajj et al., 2017; Bazzi et al., 2019). SMPs errors mainly depend on soil roughness (root mean surface height or Hrms) and NDVI values. The best results were obtained for soils with moderate surface roughness (for Hrms between 1 and 3 cm).

Underestimation and overestimation of estimated moisture, respectively, were observed for a Hrms lower than 1 cm and higher than 3 cm. In addition, the developed approach could be applied to cropland plots mainly with NDVI lower than 0.75. In a study of the Versailles plain, Vaudour et al. (2021) compared different approaches for Sentinel 2 images temporal mosaicking, mosaicking either per date or per pixel and either considering or not considering SM maps in order to produce a composite, multi-date, bare topsoil image for predicting SOC content over croplands. The best results were achieved using the per date approach driven by S1-derived moisture content (R^2~0.5, RPD~1.4, RMSE~3.7g.kg−1), which enabled the predicted area to more than double. However, the direct use of SM as a covariate in spectral models for SOC content prediction has not yet been explored.

Moreover, as superficial SOC is a dynamic characteristic, the time-lag between sampling and remote sensing data acquisition may greatly influence prediction performance, especially if changes in land-use or management practices influencing superficial SOC (e.g., organic fertilization) are rather recent. This study relied on the use of S2 time series in addition to time series from S2 in conjunction with SMP time series obtained using S1/S2. Its objectives were i) to evaluate the influence of image date and soil sampling period, ii) to assess the effect of SM as a covariate in the prediction models, and iii) to identify whether the spatial distribution of SOC as mapped in this study is related to specific soilscapes in southwestern France.

3.3 MATERIALS AND METHODS

3.3.1 Study Area

The area studied is located in the administrative region of Occitanie, a large agricultural region in the southwest of France (43°57′N–42°48′N; 0°15′W–1°51′E) covering 22,177 km^2 (Figure 3.1). It is characterized by an oceanic climate (Cfb) according to its Köppen classification (cold

winters and cool summers). A shapefile of the Land Parcel Identification System (LPIS) of 2017 (French National Institute of Geographic and Forest Information (IGN)) determined that about 12,415 km^2 of the total area consists of croplands. Different types of crops are cultivated in the area, mainly using conventional agriculture (e.g., ploughing). The area's main crops include winter wheat, maize, sunflowers and winter rapeseed.

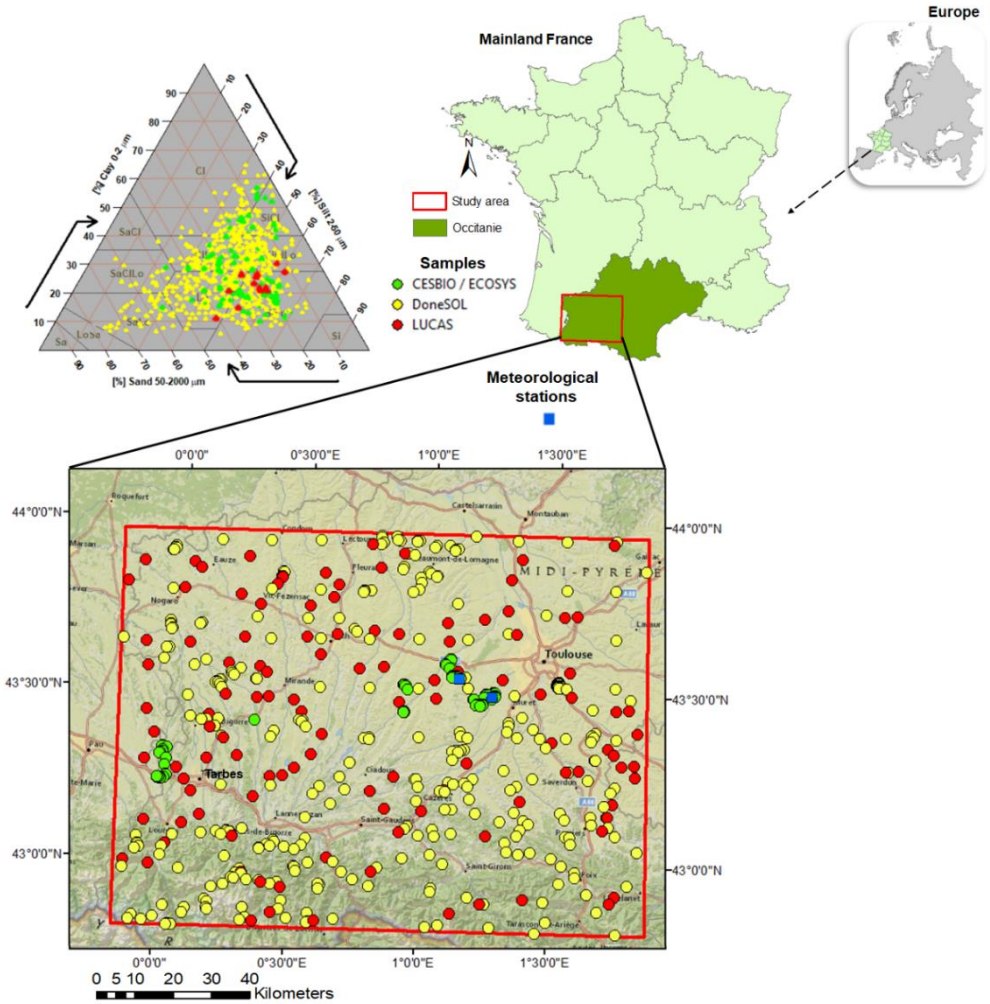


Figure 3.1 *Locations of soil samples and weather stations in the western part of the Occitanie region (France).*

The northeastern part of the study region consists of an old mountain chain, the Massif Central, which was formed approximately 380 million years ago. Calcareous formations shape the base of large plateaus neighboring volcanic and acidic rocks (Guiresse et al., 2014). The soils in the central area located on the left bank of the Garonne River are derived from molassic deposits originating from the area during the Eocene orogenesis of the Pyrenean mountains in combination with an important lacustrine system: these deposits are mainly dominated by marls, giving rise to low permeability. An increase in the flow of rivers during the Quaternary period contributed to incisions of these molassic deposits, which formed north–south-oriented patterns as gravel was deposited during interglacial phases and, thus, created alluvial terraces. The soils developed in these terraces vary from rather-undeveloped alluvial to very illuviated and differentiated soils. Acidic and alkaline soils predominate in the region according to the World Reference Base for Soil Resources with the calcareous soils mainly being Calcaric and Hypereutric Cambisols, Colluvic Regosols, Rendzic Leptosols and Fluvisols and the non-calcareous soils being Haplic Luvisols, Umbric Leptosols and Haplic and Hyperdystric Cambisols (Redon et al., 2013). Most of the naturally acidic soils are limed under cultivation in order to correct their acidity.

3.3.2 Soil Samples

We used 625 topsoil samples collected from 2005 to 2018 that are part of a set that includes three databases used for digital soil mapping (DSM) in France (Figure 3.1). The first set corresponds to the French soil profile database (DoneSol) (to see more details, visit http://www.gissol.fr/ (accessed on 12 December 2021)), in which soil information (soil profiles and augering sampling and analyses) mainly came from data gathered for conventional soil mapping using points that were spread irregularly across the French mainland territory (Arrouays et al., 2020a). The second set was collected in May 2018 by the Center for Spatial Studies of the Biosphere in Toulouse (CESBIO) in

conjunction with the functional ecology and ecotoxicology of agroecosystems unit (ECOSYS), and the third data set used in this study is the 2015 Land Use and Land Cover Survey (LUCAS) from the European Union Statistical Office (EUROSTAT) (Jones et al., 2020). R software's "soiltexture package" (Moeys, 2018) was used to classify the soil samples into eight textural classes according to the US Department of Agriculture (USDA) classification system (Figure 3.1).

In addition, in order to evaluate the possible effect of slaking on the performance of the models, slaking crust sensitivity index (SCSI) values were determined using Equation (1) (for pH values >7) and Equation (2) (for pH ≤ 7) as established by Boiffin (1984).

$$SCSI_{pH>7} = \frac{1.5*FS+0.75*CS}{Cl+10*OM} - 0.2 * (pH - 7) \quad (1)$$

$$SCSI_{pH\leq 7} = \frac{1.5-FS+0.75*CS}{Cl+10*OM} \quad (2)$$

where FS is the fine silt content, CS is coarse silt, Cl is clay and OM is organic matter content as a percentage (%).

3.3.3 Dataset Acquisition

3.3.3.1 Sentinel-2 Time Series

The Sentinel-2 (S2) images were acquired during the springs of 2017 and 2018 in periods corresponding to a maximum bare soil coverage. During these periods, the bare soils mainly comprised plots in maize and sunflower seedbed condition. Four tiles were required to cover the entire study area (T30TYN, T30TYP, T31TCH and T31TCJ); therefore, a total of 24 images were downloaded from the Muscate platform from the French Land Data Centre (Theia, https://www.theia-land.fr/, (accessed on 12 December 2021)) (Figure 3.2).

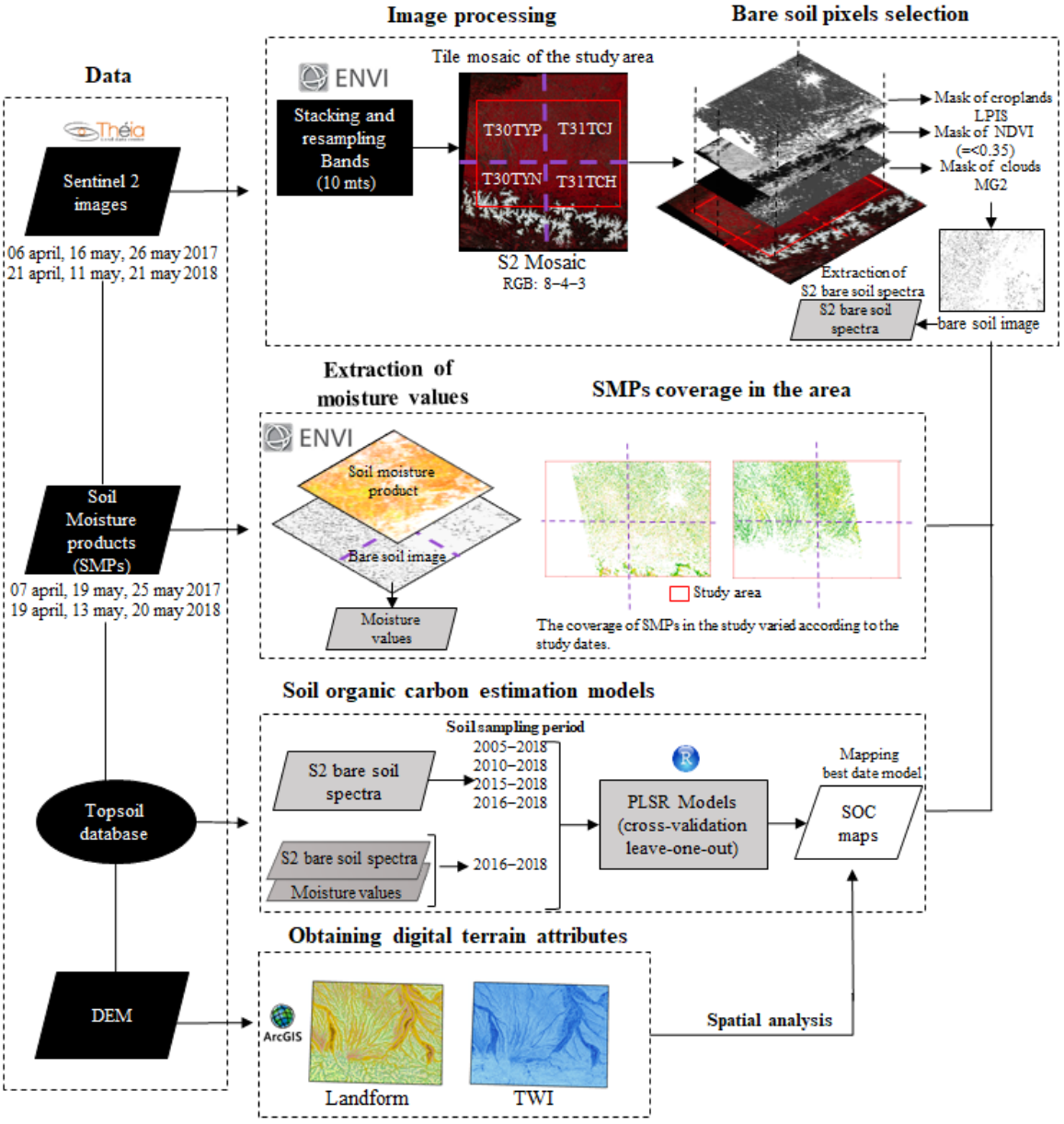


Figure 3.2 *Methodology flowchart*

The images were from six different dates (6 April, 16 May and 26 May 2017, and 21 April, 11 May and 21 May 2018) (Table 3.1). These dates were selected according to the availability of S2 images on Theia's website from between 2016 and 2019 as well as the presence of clouds; we selected those images with lower percentages of clouds (≤ 30%)

covering their surfaces. Bands B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12 were used with slope effects atmospherically corrected ("flat reflectance" or FRE).

Table 3.1. *Main characteristics of the S2 images studied*

Image Date	S2 Tile	Time of Acquisition (u.t gmt)	Viewing Incidence Zenith Angle (°)	Sun Azimuth (°)	Sun Elevation (°)	Cloud/ Shadow Cover by Tuile (%)	Cloud/ Shadow Cover of Study Area (%)
6 April 2017	T30TYN	10:53:17	<4,5	154.6	51.2	4	13.12
	T30TYP	10:53:17	<5.9	155.1	50.5	6	
	T31TCH	10:53:17	<4.3	156.3	51.6	3	
	T31TCJ	10:53:17	<3.3	156.6	50.7	11	
16 May 2017	T30TYN	10:53:22	<4.5	148.8	63.5	23	5.51
	T30TYP	10:53:22	<5.7	149.7	62.8	0	
	T31TCH	10:53:22	≤4.3	151.0	63.9	18	
	T31TCJ	10:53:22	≤3.3	151.7	63.1	2	
26 May 2017	T30TYN	10:55:18	<4.5	146.5	65.3	9	7.15
	T30TYP	10:55:18	≤5.8	147.5	64.5	0	
	T31TCH	10:55:18	≤4.3	148.8	65.7	21	
	T31TCJ	10:55:18	≤3.3	149.6	64.9	1	
21 April 2018	T30TYN	10:56:29	<4.4	153.1	56.6	14	4.79
	T30TYP	10:56:29	<5.7	153.7	55.8	1	
	T31TCH	10:56:29	<4.2	155.0	59.9	3	
	T31TCJ	10:56:29	<3.2	155.4	56.0	4	
11 May 2018	T30TYN	10:58:04	<4.4	149.9	62.4	3	2.33
	T30TYP	10:58:04	<5.7	150.7	61.6	0	
	T31TCH	10:58:04	<4.2	152.0	62.7	4	
	T31TCJ	10:58:04	<3.2	152.7	61.9	0	
21 May 2018	T30TYN	10:57:02	<4.4	147.7	64.5	71	30.09
	T30TYP	10:57:02	<5.7	148.7	63.7	11	
	T31TCH	10:57:02	<4.2	150.0	64.9	72	
	T31TCJ	10:57:02	<3.2	150.7	64.0	20	

In order to extract spectral values from the points sampled in each image, all bands were stacked and resampled at 10 m resolution and mosaics were created from the four tiles covering the area using ENVI 5.5 software (Exelis Visual Information Solutions, Boulder, Colorado) (Figure 3.2). A geophysical mask was used to remove clouds and/or

topographical shadows in each scene ("*masque géophysique*" or MG2) (Baetens et al., 2019); this mask is available for all S2 images that can be downloaded from Theia's web site. The percentages of clouds and shadows differed among tiles with the maximum values of approximately 70–75% observed on the T30TYN and T31TCH tiles (close to the Pyrénées Mountains). When only the study area was considered, on the six dates studied, the minimum coverage value was 2.3%, and the maximum coverage was 30% (Table 3.1). Spectral indices were used to differentiate vegetation and straw residues from bare soil. The normalized difference vegetation index (NDVI, Equation (3)) was used with a threshold ≤ 0.35 to retrieve bare soil pixels. The normalized burn ratio 2 (NBR2, Equation (4)) was used to determine whether a correlation with the residuals of the created models might indicate the presence of crop residues such as straw in the selected samples (Castaldi et al., 2019a).

$$\mathrm{NDVI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{Red}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{Red}}} \quad (3)$$

$$\mathrm{NBR2} = \frac{(\rho_{\mathrm{SWIR1}} - \rho_{\mathrm{SWIR2}})}{(\rho_{\mathrm{SWIR1}} + \rho_{\mathrm{SWIR2}})} \quad (4)$$

where ρ is the surface reflectance (%) of the shortwave infrared (SWIR) (i.e., SWIR1 = B11 band and SWIR2 = B12 band for Sentinel-2), the near-infrared (NIR = B8) and the red (Red = B4) spectral regions.

3.3.3.2 Soil Moisture Products and Climate Data

The soil moisture products (SMPs) derived from the Sentinel-1/Sentinel-2 satellites that were used in this work were provided by the Theia platform (https://www.theia-land.fr/product/humidite-du-sol-a-tres-haute-resolution-spatiale/ (accessed on 12 December 2021)) (Figure 3.2). SM image dates that were as close as possible to the acquisition dates of the S2 images were selected (Table 3.2). The SM images were obtained over croplands and grasslands at plot scale and provide SM estimates with an approximate accuracy of 5 vol.% (El Hajj

et al., 2017) with a six-day temporal resolution (El Hajj et al., 2017; Bazzi et al., 2019). To estimate SM values (0–10 cm depth), El Hajj et al. (2017) inverted the water cloud model parameterized by Baghdadi et al. (2017) for the C-band combined with the integral equation model as modified by Baghdadi et al. (2006). This algorithm inverts Sentinel-1 radar data to SM values and uses the normalized differential vegetation index (NDVI) derived from S2 optical data from agricultural plots as an input.

Table 3.2. *Soil moisture products (SMPs) derived from S1/S2*

S2 Acquisition Date (d_{S2})	SM Date (D_{SM})	D_{SM}-D_{S2} (Days)	SMP Cover in the Study Area (%)	Total SMP Cover * (%)	Rainfall S2 (mm) **	Previous Rain Events (mm) ***	Rainfall SMP (mm) ****
6 April 2017	7 April 2017	1	32.62	2.40	0	2.9	0
16 May 2017	19 May 2017	3	22.64	2.13	0	6.3	59
26 May 2017	25 May 2017	1	32.61	2.48	0	0	0
21 April 2018	19 April 2018	2	12.05	1.98	0	0	0
11 May 2018	13 May 2018	2	18.14	2.72	0	0	58
21 May 2018	20 May 2018	1	23.53	2.90	1.2	2	1

*: Total SMP coverage in the area considering the LPIS, masking clouds and NDVI values >0.35. **: Rainfall event on the same day as S2 images. ***: Rainfall event three days before S2 images and SMPs. ****: Rainfall event on the same day as SMP.

As in Section 3.3.3.1, bare soil spectra and SM values were obtained using ENVI software's "Spectral Library Builder" function (Figure 3.2). NDVI values lower than 0.35, MG2 and the LPIS were also used for extracting spectra from the mosaics created from the S2 and SM images. However, because the SMPs covered a smaller area than the study area (Figure 3.2), only soil samples from plots with SM information were considered (Table 3.2). Rainfall data were acquired from the southwestern France regional space observatory's Auradé and Lamasquère stations ("Observatoire spatial regional", OSR SW) (https://osr-cesbio.ups-tlse.fr/index.php (accessed on 12 December

2021)). OSR is a national observation service dedicated to monitoring the long-term effects of climate change at multiple scales. Rainfall events were recorded that corresponded to the dates of the S2 and SM images as well as the days before the images were taken.

3.3.3.3 Digital Terrain Attributes

The digital elevation model (DEM) source was the BD ALTI® version 2.0 at 25 m resolution in XY and 1 m in Z that is distributed by the IGN (French National Institute of Geographic and Forest Information). Landform classifications and the topographic wetness index (TWI) were derived from the DEM using the Topography Tools for ArcGIS 10.3 and earlier (Dilts, 2015) (Figure 3.2).

3.3.4 SOC Content Prediction Models

The partial least squares regression (PLSR) method was chosen to construct SOC prediction models based on the bare soil reflectance spectra extracted on each date from the sampling locations. The PLSR relates a matrix X consisting of explanatory variables (spectral reflectance bands and covariates) to a dependent variable Y (SOC content) using a linear multivariate model; this method is also characterized by modeling the structures of X and Y (Geladi and Kowalski, 1986; Wold et al., 2001).

The number of samples per sampling year with bare soil pixels varied among the S2 image dates. Therefore, due to the reduced number, or even the absence, of soil samples in some years, the models' performances were not evaluated according to year but instead were evaluated using periods consisting of consecutive years. The soil samples were divided into four sampling periods: (i) 2005–2018, (ii) 2010–2018, (iii) 2015–2018, and (iv) (2016–2018), i.e., the most recent data available. Using these groups, it was possible to include soil samples collected over longer periods (within which the SOC content could have changed) as well as over shorter periods (within which SOC

changes might be considered negligible) to compare the performances of the S2 models. Finally, PLSR SOC prediction models were first built using only the reflectance spectra of the S2 images with 10 selected spectral bands and were then built including S2 bands with moisture values extracted from SMPs as covariates (Figure 3.2).

When SM was included in the matrix of explanatory variables X as a band (covariate), the spectral and SM values were scaled and centered on the mean. The optimal number of latent variables was determined using the prediction residual error sum of squares (PRESS). A leave-one-out cross-validation procedure was applied (Wold, 1978). The quality of model fit was evaluated using the root mean squared error of cross-validation ($RMSE_{cv}$), the coefficient of determination of cross-validation (Rcv^2), the residual prediction deviation (RPD_{cv}) and the ratio of performance to interquartile distance ($RPIQ_{cv}$). The models were constructed in RStudio Software version 1.1.453 using the "*pls*" package (Mevik and Wehrens, 2007). The relationships between the model residuals and the textures of the samples and selected digital terrain attributes (landform and TWI) were explored.

3.4 RESULTS

The SOC content prediction performance was analyzed for the six S2 image dates and soil sampling periods using either S2 data only or S2 data together with surface SM.

3.4.1 Sentinel-2 Prediction Performance Variability and Relationships with Soil Attributes

The performance of the SOC content predictions varied according to the soil sampling periods and the dates of the S2 images, reaching RPD and RPIQ values ≥ 1.7 on two different dates (6 April and 26 May 2017) determined using the soil samples collected between 2016 and 2018. The measured SOC range of these models was ≤ 48.1 $g.kg^{-1}$ (5.03–53.1 and 5–53.1 $g.kg^{-1}$, respectively) (Table 3.3).

Table 3.3. *S2 prediction performance by date of image acquisition and soil sampling period. SOC content statistics for each set of soil sampling period used in the models (better performing models are in bold characters)*

								SOC ($g.kg^{-1}$)						
S-2 Date	**Soil Sampling Periods**	**NS**	**R^2_{CV}**	**$RMSE_{CV}$ ($g.kg^{-1}$)**	**RPD_{CV}**	**$RPIQ_{CV}$**	**NC**	**Min**	**Me**	**x̄**	**Max**	**SD**	**Skw**	**Kr**
6 April 2017	2005–2018	187	0.48	6.74	1.4	1.07	6	2.4	10.1	13.7	53.1	9.45	2	6.95
	2010–2018	165	0.52	6.82	1.46	1.08	6	2.4	10	14	53.1	9.94	1.92	6.2
	2015–2018	132	0.64	5.42	1.7	1.25	7	2.4	9.4	13.1	53.1	9.13	2	6.9
	2016–2018	**98**	**0.7**	**5.58**	**1.83**	**1.68**	**7**	5.03	9.1	14.14	53.1	10.2	1.73	5.26
16 May 2017	2005–2018	195	0.36	6.48	1.26	0.96	4	0.94	9.6	12.6	53.1	8.15	2.28	8.7
	2010–2018	163	0.45	6.32	1.36	0.9	6	2.4	9.5	12.84	53.1	8.6	2.25	8.15
	2015–2018	130	0.58	5.93	1.55	0.95	6	2.4	9.2	12.89	53.1	9.2	2.2	7.6
	2016–2018	**95**	**0.68**	**5.82**	**1.78**	**1.43**	**6**	5.03	9.16	14	53.1	10.3	1.84	5.6
26 May 2017	2005–2018	199	0.4	6.51	1.3	1.06	4	0.94	9.62	13.09	53.1	8.5	1.95	6.8
	2010–2018	169	0.48	6.39	1.4	1.17	4	2.4	9.5	13.45	53.1	8.9	1.87	6.21
	2015–2018	134	0.59	6.12	1.56	1.24	4	2.4	9.32	13.57	53.1	9.5	1.81	5.71
	2016–2018	**100**	**0.65**	**6.17**	**1.7**	**1.75**	**5**	5	9.33	14.7	53.1	10.5	1.51	4.4
21 April 2018	2005–2018	204	0.35	9.16	1.25	0.72	6	0.8	10	13.9	89.8	11.4	3.24	16.8
	2010–2018	182	0.37	8.72	1.27	0.75	6	0.8	10	13.8	89.8	11	3.15	16.7
	2015–2018	148	0.56	5.78	1.51	1.12	5	2.4	9.8	13	53.1	8.71	2.15	7.66
	2016–2018	**122**	**0.6**	**5.7**	**1.6**	**1.14**	**5**	5.01	9.9	13.5	53.1	9.22	2	6.82
11 May 2018	2005–2018	236	0.37	6.09	1.27	1.07	4	0.8	10.42	12.84	53.1	7.73	2.09	8.35
	2010–2018	208	0.42	6	1.32	1.07	4	0.8	10.3	12.96	53.1	7.98	2.1	8.13
	2015–2018	171	0.58	5.42	1.56	1.26	5	2.4	10.3	13.2	53.1	8.43	2.07	7.56
	2016–2018	**135**	**0.66**	**5.22**	**1.73**	**1.53**	**6**	4.21	10.6	13.98	53.1	9	1.89	6.46
21 May 2018	2005–2018	202	0.36	6.47	1.26	1	4	2.4	10.7	13.45	53.1	8.13	2.11	7.93
	2010–2018	183	0.36	6.6	1.26	1	6	2.4	10.6	13.5	53.1	8.36	2.11	7.72
	2015–2018	152	0.47	6.44	1.38	1.15	6	2.4	10.6	13.85	53.1	8.9	1.9	6.87
	2016–2018	**123**	**0.55**	**6.35**	**1.5**	**1.36**	**6**	5.25	10.8	14.63	53.1	9.5	1.8	5.85

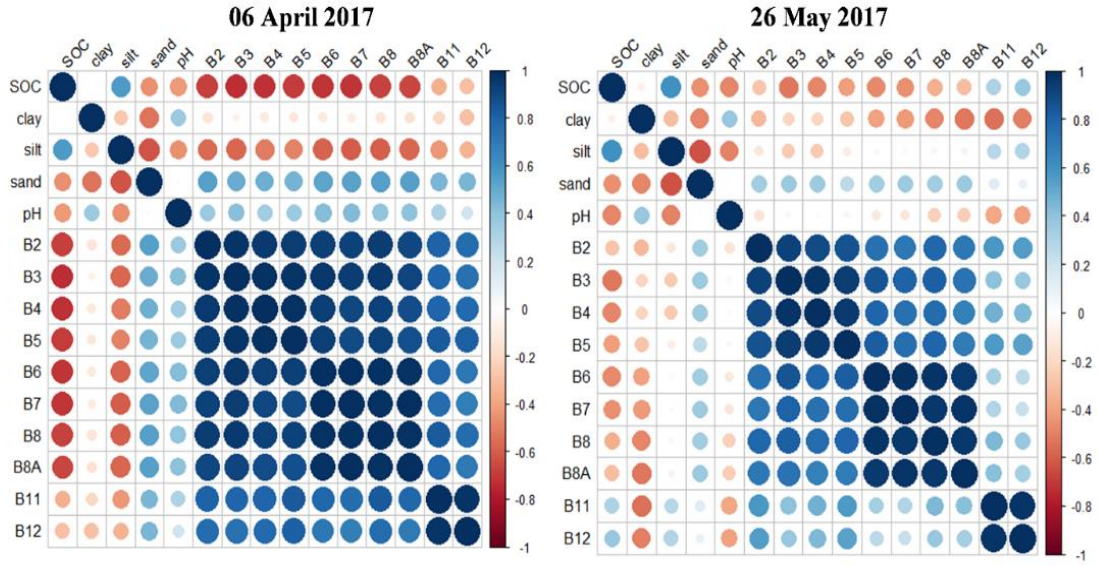
NS: number of samples; R^2_{CV}: coefficient of determination of cross-validation; RMSECV : root mean squared error of cross-validation; RPDCV : residual prediction deviation; RPIQCV : ratio of performance to interquartile distance; NC: number of components; Min: minimum; Me: median; x̄: mean; Max: maximum; SD: standard deviation; Skw: skewness; Kr: kurtosis.

The best performances were obtained for most models using the set of samples from all S2 image dates collected between 2016 and 2018 (Table 3.3). The performances of the models obtained on 21 April and 21 May 2018 are rather poor. The poor performances of these models might be related to changes in soil roughness, e.g., those changes that were due to the soil plowing operations that were common during this

period in plots where corn or sunflower were sown. Another factor contributing to model performance might be the formation of crusts on soil surfaces due to rainfall; for example, in the 2018 soil sample collection, crusting on the soil surfaces of some plots was visually observed. Table 3.2 shows that there was rainfall on the same day that image S2 was taken as well as three days earlier on May 21, 2018, which could explain, at least in some plots, the formation of surface crust that could affect model performance.

S2 images from all dates prior to use of the MG2 had clouds and shadows, with some having less than others. This may be due to the large size of the area and its proximity to the Pyrénées mountains. The best performances (6 April and 26 May 2017) presented low cloud and shadow coverage (13.12 and 7.15%, respectively) (Table 3.1). Conversely, 21 May 2018 had the highest percentage of cloud and shadow coverage (30.09%), and its RMSEcv value (> 6.3 $g.kg^{-1}$) was the highest among the models (2016–2018). In almost all of the scenes, most clouds and shadows were located on the tiles (T30TYN and T31TCH) closest to the Pyrénées mountains (Figures 3.1 and 3.2 and Table 3.1). On 6 April and 26 May 2017, which had the images with the best predictions, a negative correlation trend was observed between SOC content levels and the spectral information from the S2 bands. In both models corresponding to the S2 dates of 6 April and 26 May 2017, there was no correlation between SOC and clay content in the soil samples (Figure 3.3), contrary to what was previously observed at a national scale (Arrouays et al., 2006) or in small regions (Vaudour et al., 2019b). The Pearson's correlation matrix figure is important in our study because as described above the relationships between soil physical attributes such as clay and SOC can be different and lead to different conclusions when studies are done at different scales. Furthermore, it would be useful in further studies for SOC mapping to observe the correlation coefficients between the SOC values and the S2 image bands at single dates. Because, as observed in this work, the

sensitivity of the satellite bands to different soil properties may vary not only considering the study area but also according to dates of acquisition of the S2 images. It is worth noting that in our study we found suitable correlations of spectral bands with soil properties (e.g., for SOC, r = −0.75 for the bands B2 to B8A of 6 April) (Figure 3.3). This approach might support the understanding of SOC variability for digital mapping as a baseline to be applied at national scales in order to obtain maps with better accuracy.



SOC: measured soil organic carbon

Figure 3.3. *Pearson's correlation matrix among Sentinel-2 bands and soil properties (SOC, clay, silt, sand and pH) from the best-performing models*

3.4.2 S2 and SMP Prediction Performance

In this section, SM was included in the models that used soil samples collected between 2016 and 2018. Due to the limited number of samples with SM information, the number of samples used for the models was lower (between 13% and 46% less, according to the model) than in the first approach described in Section 3.4.1.

Nevertheless, new models with lower sample density were created in order to assess whether the use of SM as a covariate influenced performance. The performance of the models including SM was not better than those considering the S2 bands, remaining almost the same (Table 3.4).

Table 3.4. *S2 prediction performance using surface soil moisture from soil moisture products (SMPs) as a covariate. SOC and SM statistics of the sample set used in the models (the models using SM as a covariate are in bold characters)*

								SOC (g.kg^{-1})							Soil Moisture (Vol.%)			
SM Date	**NS**	**MD**	**R^2_{CV}**	**$RMSE_{CV}$ (g.kg−1)**	**RPD CV**	**$RPIQ_{CV}$**	**NC**	**Min**	**Me**	**x̄**	**Max**	**SD**	**Skw**	**Kr**	**Min**	**Me**	**x̄**	**Max**
7 April 2017	84	No	0.66	5.67	1.75	1.52	7	5.48	9.1	14	53.1	9.9	1.82	5.8	6.8	17.8	17.2	24.6
		Yes	**0.66**	**5.66**	**1.75**	**1.53**	**8**											
19 May 2017	69	No	0.67	6.07	1.77	1.77	5	5.03	10.5	15	53.1	10.7	1.63	4.9	9	22.2	21.6	28
		Yes	**0.67**	**6.1**	**1.76**	**1.75**	**6**											
25 May 2017	87	No	0.64	6.45	1.67	1.73	5	5.48	10.2	15	53.1	10.7	1.44	4.1	5.4	13	12.6	19.2
		Yes	**0.63**	**6.54**	**1.65**	**1.7**	**7**											
19 April 2018	66	No	0.62	6.72	1.63	1.46	4	5.01	13.6	17	53.1	10.9	1.26	3.8	8.8	19.3	19.3	26.4
		Yes	**0.6**	**6.82**	**1.6**	**1.44**	**5**											
13 May 2018	79	No	0.76	5	2.06	1.9	3	4.21	13.8	17	53.1	10.4	1.27	4.1	17.4	28	27.6	35.2
		Yes	**0.76**	**5.1**	**2.03**	**1.87**	**4**											
20 May 2018	89	No	0.55	6.37	1.5	1.58	7	5.25	12.3	15	53.1	9.5	1.7	5.9	9.8	22.4	22.3	29
		Yes	**0.58**	**6.12**	**1.56**	**1.65**	**7**											

SM: soil moisture; NS: number of samples; MD: model using moisture data; R^2_{CV}: coefficient of determination of cross-validation; RMSECV: root mean squared error of cross-validation; RPDCV: residual prediction deviation; RPIQCV: ratio of performance to interquartile distance; NC: number of components; Min: minimum; Me: median; x̄: mean; Max: maximum; SD: standard deviation; Skw: skewness; Kr: kurtosis.

Figure 3.4 displays each SMP date and their respective histograms. The dates in 2017 (7 April and 25 May 2017) were characterized by having lower moisture content (15.76 ≥ mean ≥ 14.72) than the other dates, particularly those from 2018 (29.8 ≥ mean ≥19.8). No performance trend was shown for the soil moisture values from the set of samples used in each model (Table 3.4). The SMPs dates were not exactly the

same as those of the S2 images; in some cases, when the date difference was greater than 2 days, a rainfall event occurred between the 2 dates (Table 3.2). The roughness conditions, among other disturbing factors, were unknown for the dates studied herein, so it was not possible to assess their influence. This study does not intend to elucidate disturbing factors such as soil roughness but rather seeks to study the influence of SM on prediction performance.

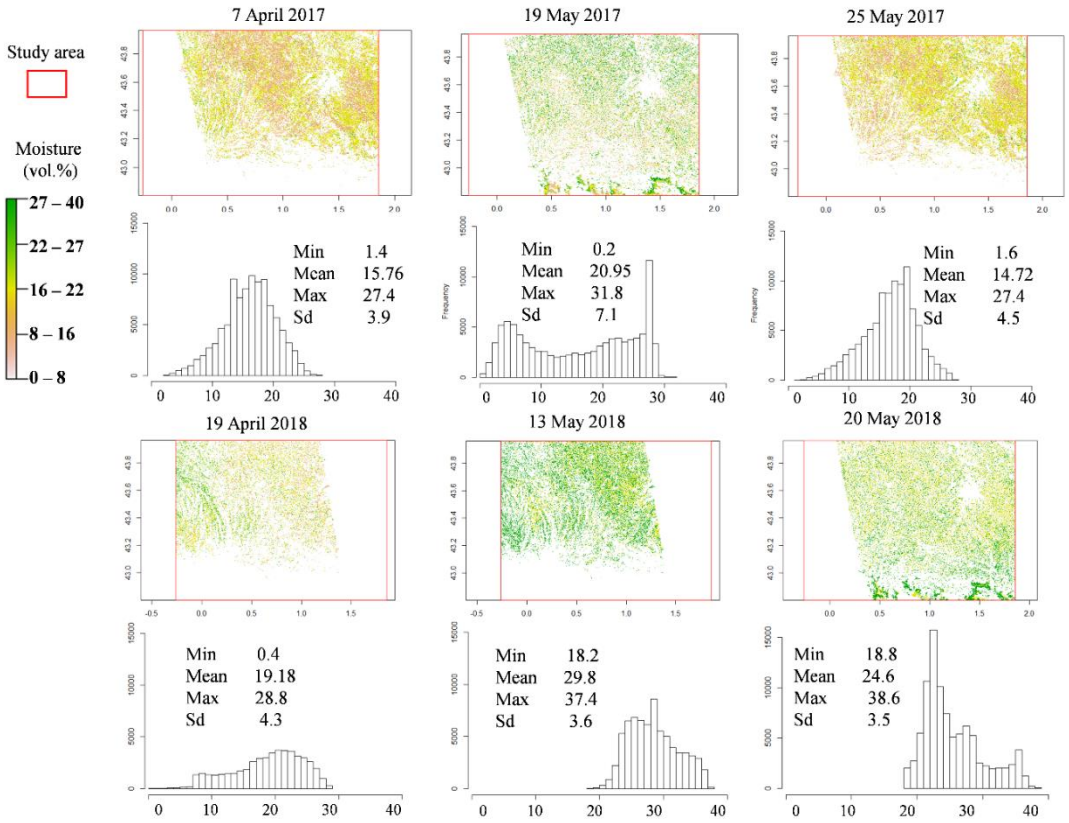


Figure 3.4. *Soil moisture products (SMPs) and their respective histograms*

3.4.3 Spatial Prediction and Characteristics of SOC Maps

Based on the best models in Section 3.4.1, two maps of predicted SOC

contents were produced (Figure 3.5). The mean and SD of the predicted SOC content values were 21.4 and 6.9 g/kg for 6 April 2017 and 27.5 and 7.15 g/kg for 26 May 2017; the models exhibited moderate SD values (<7.2 g/kg). It is noteworthy that the range of predictions varied among the models, while both dates had similar ranges for the observed values for calibration samples (Table 3.3). However, the SOC map predicted for 26 May 2017 showed a wider prediction range than the previous model (6 April 2017). This variation as well as the moderate SD values of the model indicate rather stable predictive capability in an area with diverse soil types and landforms (Figure 3.5 and see, further Section 3.5.3).

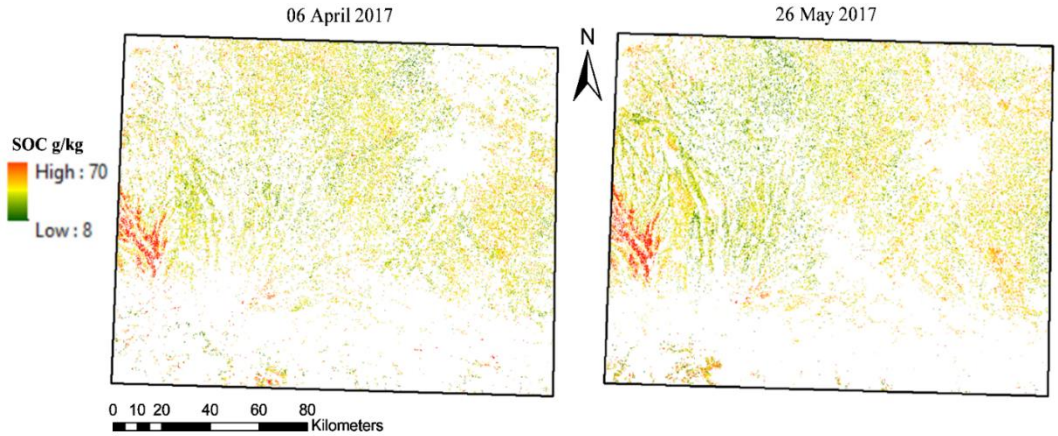


***Figure 3.5**. SOC maps predicted*

3.5 DISCUSSION

Laboratory and satellite remote sensing studies have shown that soil water content impairs SOC prediction accuracy, so this factor is generally either excluded or controlled for in order to ensure better performance predictions. However, in satellite imagery of croplands, factors related to the conditions of the imagery (clouds, shadows, sun elevation angle), the study area (SM, growing season, roughness, landform) and the agricultural practices (crop rotation, weed management) cannot be controlled. Indices are used to avoid some of

these factors. However, if the seasonal relationship between SOC prediction performance and SM is known for a specific region, the best periods from a time series of S2 images can be inferred by considering SM information, improving the performance of SOC predictions.

3.5.1 Optimal Dates and Characteristics of S2 Images and Sampling Periods for SOC Prediction

S2's capability for determining different soil attributes such as SOC for croplands has been studied recently due to its characteristics (spatial and temporal resolution), which allow soil monitoring at both the regional and national scales (Loiseau et al., 2019; Vaudour et al., 2019b). In this study, comparative analysis revealed that both soil sampling period and image date selection affected the accuracy of the SOC prediction models (Table 3.3). It was found that the prediction models achieved better accuracy when recent samples (2016–2018) were selected; poor performances were observed when sequentially old soil samples were incorporated into the prediction models. This could be due to the fact that SOC contents have decreased a lot over time; this is dependent on the duration of continuous maize cropping in particular (Arrouays and Pélissier, 1994; Arrouays et al., 1995b). The dates of the S2 images also influenced the accuracy of the models. This is consistent with the results of Vaudour et al. (2019a), who compared the performance of PLSR models using S2 imagery to predict SOC during different periods in the plain of Versailles, reporting variations in predictive capability for the imagery acquired between 2016 and 2017.

S2 has been widely used to map SOC on croplands in several locations using different approaches. For example, these approaches have included the use of single date images (Castaldi et al., 2019b; Vaudour et al., 2019a;b) and time series creating composite images of bare soil (Silvero et al., 2021b; Vaudour et al., 2021) including S2 data, other satellite sensors and environmental variables as inputs (Zhou et al.,

2020b; 2021). The SOC prediction performances in these studies were similar to those in this work. Castaldi et al. (2019b) achieved RPD values ≥ 2; however, RMSE values grew higher as the RPD increased. Although their performances were not as high (RPD ≤ 1.5), Vaudour et al. (2019a;b), obtained RMSE values lower than those found in this work (RMSE ≤ 5 $g.kg^{-1}$). These results could be related to the different conditions of the study areas, e.g., soil type, relief, management practices, SOC ranges and even sensor accuracy. Moreover, some of the main limitations when determining soil attributes using satellite sensors are the percentages of bare soil available, clouds, vegetation, crop residues and soil moisture in the images, and, thus, algorithms and indices have been developed in order to reduce the effects of these factors (e.g., Demattê et al., 2018; Diek et al., 2017). However, satellite image characteristics such as sun elevation, percentages of clouds and shadows, roughness and SM are often neither related to the performances of the models nor studied in detail (Vaudour et al., 2019a). The acquisition conditions of the scenes, e.g., sun elevation angle, can change from one place to another; this is generally determined by the time and season of acquisition. Table 3.1 shows that all scenes had a sun elevation angle > 50°. Vaudour et al. (2019a) used images from the Versailles plain from different seasons with sun elevation angles between 16 and 52° . These authors observed that the accuracy of the models improved when the sun elevation angle was higher, achieving values of RMSE = 3.02 $g.kg^{-1}$ and RPD = 1.5. In the present study, here the range of the angles was not as wide because all images were acquired in spring; therefore, no trend was found. When observing cloud cover and cloud shadows, the percentage of cloud cover was low on the best performing dates of 6 April and 26 May 2017 (Table 3.1). Further works could examine datasets from different sites in order to compare how image characteristics influence the model performance.

3.5.2 Impact of Soil Moisture

SM content influences spectral response and the quantification of soil attributes (Silvero et al., 2020). SOC detection performance is particularly sensitive to soil moisture (Minasny et al., 2011; Rienzi et al., 2014). Therefore, methods using spectral transformations have been proposed for lab measurements (Minasny et al., 2011; Tan et al., 2021) and via satellites using spectral indices that generally combine infrared bands near the water response to reduce the effect of moisture (Castaldi et al., 2019a; Demattê et al., 2018; Vaudour et al., 2021). Table 3.4 displays RPD and RPIQ values > 1.5 from 2017 and 2018. These performances are higher than those in Table 3.3. The performance of the models did not improve when including SM as a covariate (Table 3.4); however, as mentioned in the results section above, the number of samples considered for these models was smaller, so the lower number of calibration points may have impeded the performance of the models including SM.

Figure 3.6 displays the importance of spectral bands and SM when it was considered in the models. The most important bands were detected in the SWIR region (B11 and B12). Studies have found similar results for SOC prediction using S2 bands (Vaudour et al., 2019b; Žížala et al., 2019). Castaldi et al. (2019b) reported that the importance of bands in PLSR regression models can be different depending the study area and soil physical attributes; bands 11 and 12 in all areas of their study were the most important. Regarding SM its importance was the lowest, this confirms the results (see Table 3.4 and Figure 3.6). Moreover, the best performances considering only the S2 bands were obtained for the driest periods (Table 3.3; Figure 3.4). Rienzi et al. (2014) estimated SOC using a spectrometer with different percentages of water content and found that prediction accuracy was the best (RMSE ≤ 6.38 $g.kg^{-1}$) and there was a larger range in predictions when water content was between 0 and 15%, whereas when the water content was between 20 and 25%, the accuracy was lower (RMSE ≥

7.06 g.kg^{-1}) and the range of predictions was very narrow. Vaudour et al. (2021) obtained similar results with S2 images, although they did not use SM as an input in the models. In addition, their best performances were associated with lower mean SMP values (21.1 ≥ vol.% ≥ 9.8). In our study, the SMP dates were not exactly the same as those of the S2 images (Table 3.2). In the models (6 April and 26 May 2017), there was exactly one day's difference between the S2 images and the SMP, and rainfall events were null on and between these dates (Table 3.2), and therefore, we can deduce that changes in SM were negligible between these two dates. 19 May 2017, and 13 May 2018, presented rainfall values of 59 and 58 mm, respectively, which is consistent with the moisture values of the SMP (Figure 3.4). On 20 May 2018, no rainfall events were observed; however, three days earlier, rainfall occurred in the area (Table 3.2), which could explain the high values of the moisture image (Figure 3.4). The ideal situation could be when the last rain event would be long before both image dates (S2 and SM) with no rain event between the two dates.

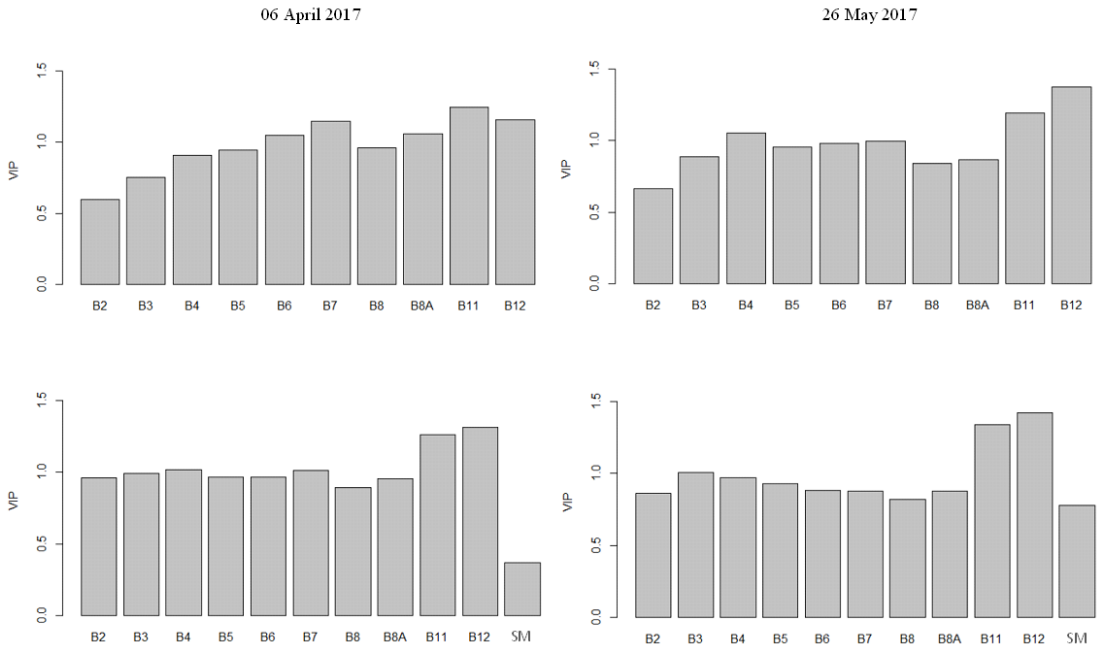


Figure 3.6. *Variable importance (VIP) of PLSR models considering S2 spectral bands (on the top) and S2 spectral bands plus SM (on the bottom).*

3.5.3 Influence of Digital Terrain Attributes on the Predicted SOC Map

The French Pyrenean piedmont (southwest France) is characterized by the diversity of landforms in the area (Figure 3.7). The sample set used by the SOC prediction models in this study was mainly divided into two classes: "upper slopes/mesas" and "U-shaped valleys". An analysis of their SOC content determined the predictive capability of the models, revealing that the highest values were on "upper slopes/mesas" (Figure 3.8). However, other studies have shown that often, soils located on upper slopes are shallow with low SOC content, while soils on lower slopes are deeper and moister with high SOC content (Meliyo et al., 2016; Patton et al., 2019).

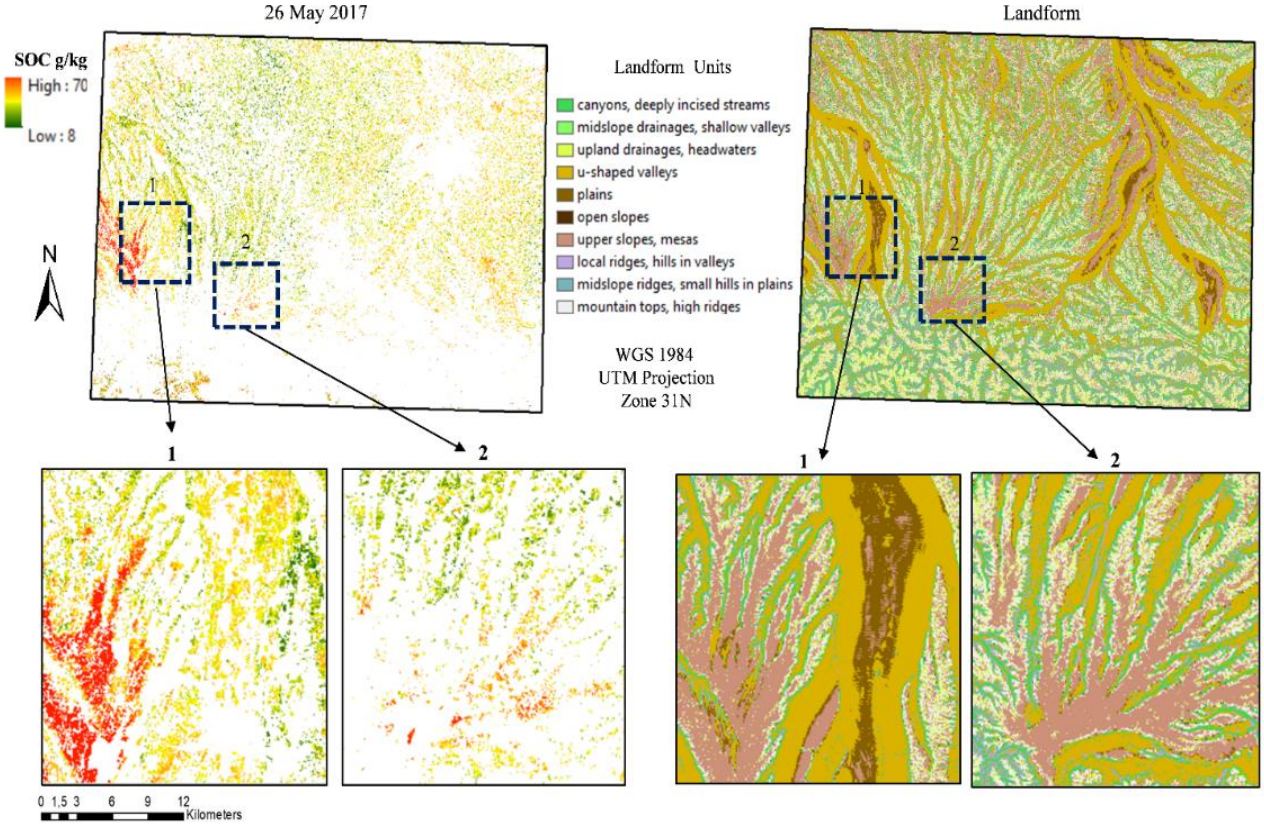


Figure 3.7. *Zooming into the SOC map and the landforms of the study area.*

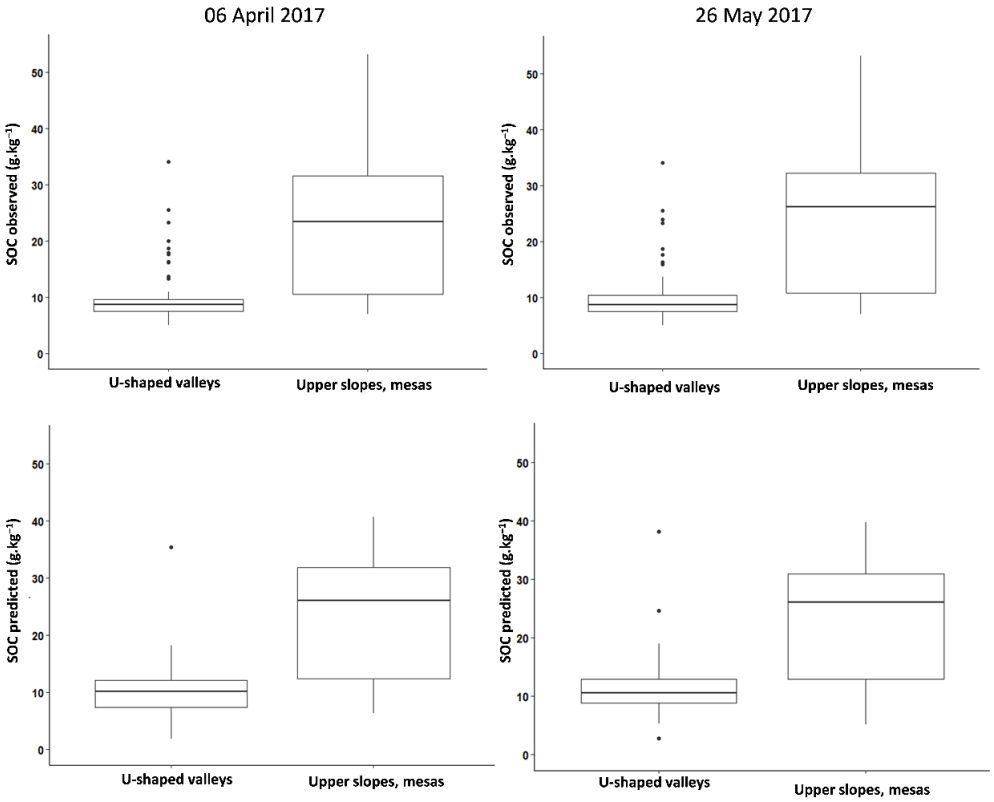


Figure 3.8. *Relationships between landforms and the SOC observed and predicted using the models from 6 April (left column) and 26 May 2017 (right column).*

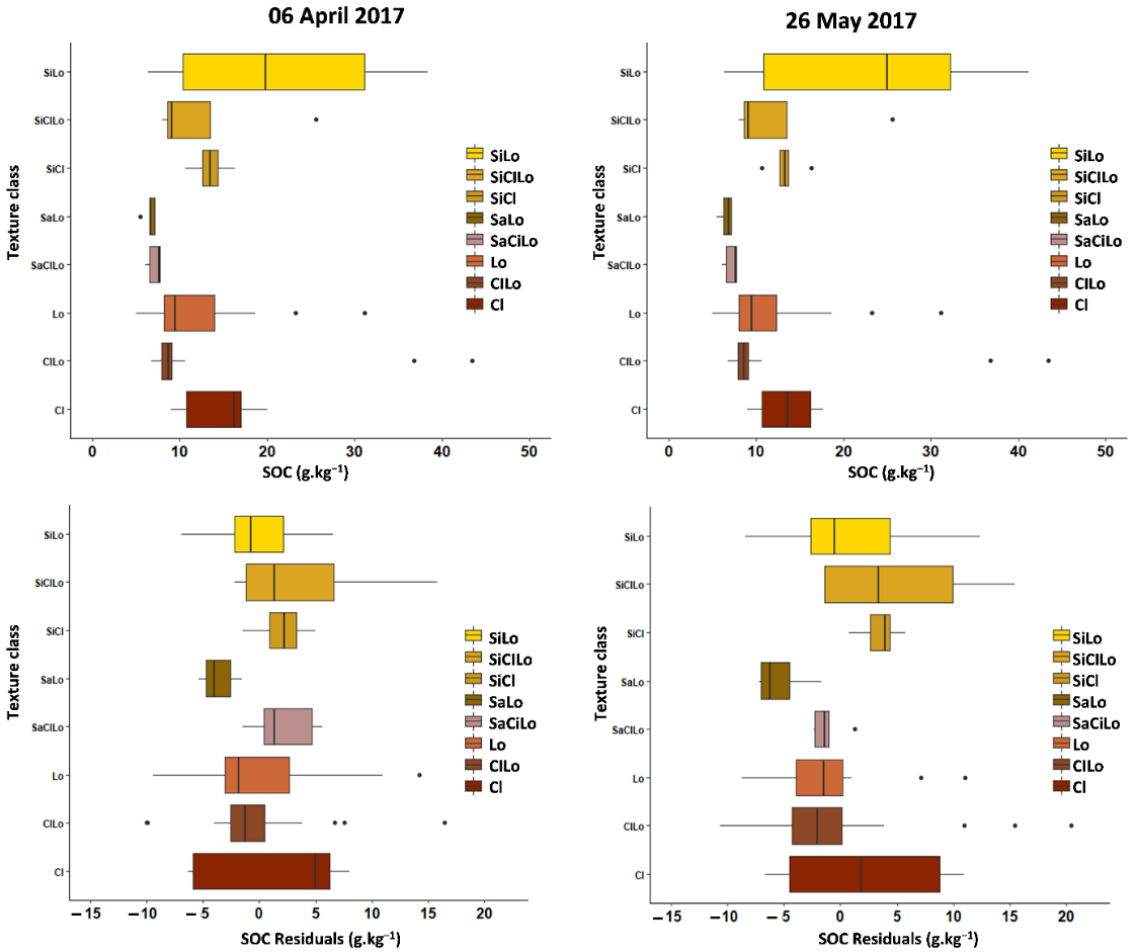
When considering texture, silty loam (SiLo) soil samples had higher SOC values; moreover, careful examination of the residues indicated underestimation of SOC values in silty textures and overestimation in clay loam (ClLo) and clay (Cl) textures (Figure 3.9). It might be possible to infer that the overestimates were associated with calcareous soils, but the number of samples with $CaCO_3$ values > 100 $g.kg^{-1}$ was low (approximately 11 samples). Briefly, the results indicated high SOC contents on "upper slopes/mesas" with SiLo textures. Previous works covering three areas within the southwestern Pyrenean piedmont conducted by Arrouays et al. (1995a;b), Arrouays and Pelissier (1994), Besnard et al. (1996) reported the acidic, humic silty loam soils classified as "Veracrisols" in the French pedological reference base and

as "Vermic Haplumbrepts" in soil taxonomy. These soils developed in Quaternary alluvial deposits and are rich in organic matter; in the past, forests existed on these deposits on ancient terraces, but from the early 1960s to present day, there has been progressive deforestation and conversion into croplands for continuous maize cropping (Arrouays et al., 1992; 1995b). Figure 3.7 shows that the maps predicted high SOC for two zones in the southwest of the study area (see zooms 1 and 2 on the map) with median values of 34 and 29 $g.kg^{-1}$, respectively, in each zone. According to the landform map, these two highlighted areas are located on "upper slopes/ mesas", and most of the soil samples collected in these areas had SiLo textures; this confirms Arrouays et al. (1992; 1995a)'s description of ancient terraces in the study area as well as in more western parts of the Pyrenean piedmont. In other words, our predictions detected very high SOC values on this specific soilscape under the condition that enough bare soils were available on the S2 acquisition date.

As nearly all of the soils located on these ancient terrace soilscapes are under intensive continuous maize cropping (especially in the extreme southwestern part of the region), the soils are almost all bare during the same period, and if this period is dry, the prediction is fairly accurate. Moreover, this prediction has been performed for soils that still have high SOC (Arrouays et al., 1995a), which should, therefore, be protected in order not to release CO_2 in the atmosphere (Martin et al., 2021) through enhanced mineralization under intensive cultivation (Arrouays et al., 1995b). Further studies on these soils could help assess whether using different sets of S2 data could enable the monitoring of SOC decreases in this specific soilscape. Interestingly, these soils also have rather low clay content, showing that statistical relationships established at different scales (Arrouays et al., 2006) can lead to different conclusions. It is very likely that SOC mapping potential also differs depending upon scale and the region of interest. Thus, other case studies should use S1/S2 products to further explore the potential

of SOC prediction as well as the effects of various disturbing factors.

Landforms as well as variables derived from relief and drainage networks influence digital soil mapping and carbon stock estimates (Patton et al., 2019; Mello et al., 2021). Slope processes including erosion and runoff deposition are related to terrain attributes such as the topographic wetness index (TWI). Studies have reported TWI to be a strong predictor of SOC stocks, and it has been used in digital SOC mapping at various scales (Thompson et al., 2006; Zhou et al., 2020b; 2021). In this study, TWI was not used in the prediction models because it was not our main focus, but it was surprising to observe a slight trend from overestimations at lower TWI values (Q1 < 8.9) to underestimations at higher values (Q4 > 10.8) with the predicted SOC residuals in "U-shaped valleys" (Figure 3.10). This might be explained by the relationship between texture and prediction errors (Figure 3.9); this is in accordance with the Australian results reported by Minasny et al. (2013), who observed that loamy texture soils, which were generally located in depressions, retained more moisture. However, rather than soil moisture for the dates shown in Figure 3.10, and as soil moisture was similar and lower than 18% vol. in median for both "U-shaped valley"and "upper slopes/ mesas" for such dates, TWI variation could be consistent with the percentage of coarse fragments of alluvial or colluvial origin, which was previously shown influent on spectral prediction errors (Vaudour et al., 2019b). Of course, it cannot be inferred from Figure 3.10 that TWI alone causes "U-shaped valleys" prediction errors.



SiLo: silty loam; SiClLo: silty clay loam; SiCl: silty clay; SaLo: sandy loam; SaClLo: sandy clay loam; Lo: loam; ClLo: clay loam; Cl: clay.

Figure 3.9. *Relationships between texture classes and observed SOC (up) and SOC prediction residuals (down) from models from 6 April (left column) and 26 May 2017 (right column).*

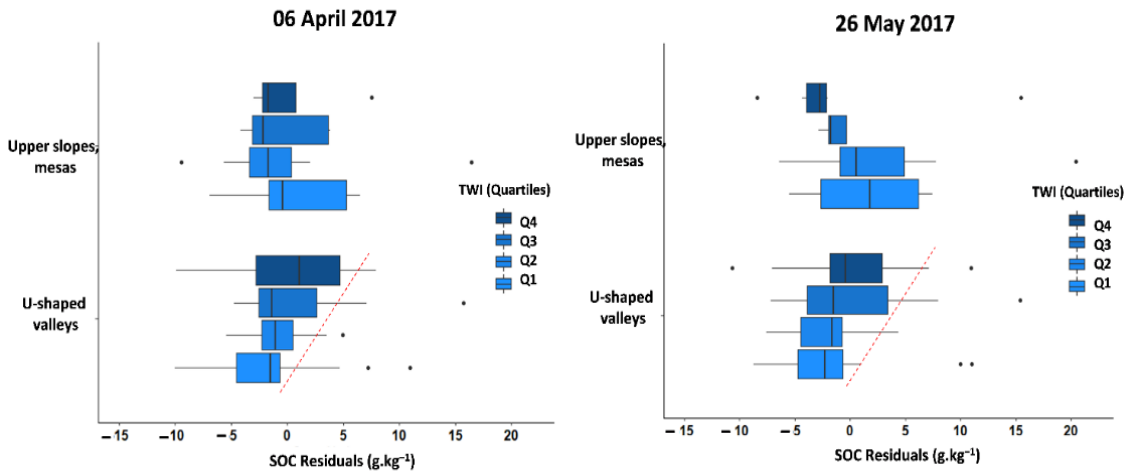


Figure 3.10. *Relationships between quartiles of the topographic wetness index (TWI) of landforms and SOC prediction residuals.*

Finally, the formation of "soil surface slaking crusts" in some plots due to relief and soil attributes was analyzed. According to Aubert et al. (2011) and de Jong et al. (2011), in regions with silty and loamy textures, soil crusting is commonly observed; this effect was seen in some locations during the field campaign conducted in June 2018, which indicates its possible occurrence in the study area. The results obtained using Boiffin's method (Boiffin. 1984) showed higher values for sensitivity to slaking in the prediction errors for "upper slopes /mesas". However, the values for the indices of sensitivity to slaking were either low or very low according to Boiffin's classification, and therefore, it is unlikely that they affected model performances. Moreover, sensitivity to slaking does not necessarily imply crusting, which also depends on rainfall intensity and on the structure of the soil surface before a rainfall event. Thus, although it has been experimentally shown that very silty soils in this region may become sensitive to slaking when they have low SOC contents (Le Bissonnais and Arrouays, 1997), this sensitivity does not necessarily result in crusting during all periods. This research did not focus on all of the

different factors that can limit the DSM of SOC via satellite imagery but rather on how differences among soil sampling dates and using S2 images and SM as inputs could influence SOC prediction performance. However, it is worth mentioning that although this work cannot provide a detailed explanation for all of the disturbing factors, some of them were evidenced. Therefore, it would be interesting for further studies to analyze data sets from different sites in order to evaluate effects related to satellite imagery characteristics by date, effects related to the SM obtained from SMPs and effects associated with digital terrain attributes, soil type and/or land-use history. In addition, the use of composite images might extend the areas of bare soil to be researched during dry and wet periods. Information on roughness and management practices and fieldwork surveys observing soil crust could be useful for SOC mapping.

3.6 CONCLUSIONS

This study confirms the previous statement according to which dates of acquisition of S2 images are crucial (Vaudour et al., 2019a). It is original in extending such statement to the dates of soil sampling periods, which affect SOC prediction. Depending on dynamics of soil organic carbon storage, it is not advisable to use too old samples (> 3–5 years in this study), but some samples older than the acquisition years may be used, thus enabling to gather datasets of sufficient size, better capturing the spatial and spectral diversity of the soilscape units.

The performance of a single date may be related to factors disturbing satellite prediction sensitivity such as the circumstantial conditions present in the study area (SM and soil roughness generated by agricultural operations or natural events). Although the combined use of S2 images and SMPs derived from S1 and S2 sensors had no impact on the performance of the models, this work highlights that SMPs can be used to select the best periods for digital SOC mapping. Indeed, all results converge to demonstrate that the driest periods are the best

suited for mapping SOC on bare soils.

Landform analysis as well as previous digital soil mapping studies conducted in this region have clearly demonstrated that S2 is able to distinguish high SOC content in "upper slopes/mesas" on rather large Quaternary silty alluvial deposits. In other words, S2 was able to map a specific soilscape characterized by high SOC content. This is an important finding, as mapping soils with high SOC content may be a very useful tool for the implementation of practices protecting these soils (for example, favoring cover-crops in winter). Moreover, these results suggest that S2 products can be used as tools to monitor large decreases in SOC content in this specific soilscape. However, more detailed analyses of the use of SMPs in single-date or composite images should be conducted in order to determine whether these products are useful for choosing the best prediction dates and periods and/or whether including SMP values as covariates in synergy with S2 imagery and terrain-derived covariates could improve digital SOC mapping and monitoring performance in areas with different conditions (climate, topography, soil and soil genesis).

CHAPTER 4

4 SENTINEL-2 AND SENTINEL-1 BARE SOIL TEMPORAL MOSAICS OF 6-YEAR PERIODS FOR SOIL ORGANIC CARBON CONTENT MAPPING IN CENTRAL FRANCE

Diego Urbina-Salazar[1], Emmanuelle Vaudour[1], Anne C. Richer-de-Forges[2], Songchao Chen[3,4], Guillaume Martelet[5], Nicolas Baghdadi[6] and Dominique Arrouays[2]

[1] *UMR 1 University of Paris-Saclay, INRAE, AgroParisTech, UMR ECOSYS, 91120, Palaiseau, France*

[2] *INRAE, Info&Sols, 45075 Orléans, France*

[3] *ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou 311200, China*

[4] *Institute of Agriculture Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China*

[5] *BRGM, 3 av. Claude Guillemin, BP36009, 45060, Orléans cedex 2, France*

[6] *UMR TETIS, INRAE, University of Montpellier, UMR TETIS, 500 rue François Breton, 34093 Montpellier Cedex 5, France*

Urbina-Salazar, D.; Vaudour, E.; Richer-de-Forges, A.C.; Chen, S.; Martelet, G.; Baghdadi, N.; Arrouays, D. Sentinel-2 and Sentinel-1 Bare Soil Temporal Mosaics of 6-Year Periods for Soil Organic Carbon Content Mapping in Central France. Remote Sens. 2023, 15, 2410. https://doi.org/10.3390/rs15092410

Keywords: Soil organic carbon, Sentinel-1/2, bare soil Mosaics, Quantile random forest, airborne gamma-ray, digital soil mapping, uncertainty

4.1 ABSTRACT

Satellite-based soil organic carbon content (SOC) mapping over wide regions is generally hampered by the low soil sampling density and the diversity of soil sampling periods. Some unfavorable topsoil conditions, such as high moisture, rugosity, the presence of crop residues, the limited amplitude of SOC values and the limited area of bare soil when a single image is used, are also among the influencing factors. To generate a reliable SOC map, this study addresses the use of Sentinel-2 (S2) temporal mosaics of bare soil (S2Bsoil) over 6 years jointly with soil moisture products (SMPs) derived from Sentinel 1 and 2 images, SOC measurement data and other environmental covariates derived from digital elevation models, lithology maps and airborne gamma-ray data. In this study, we explore (i) the dates and periods that are preferable to construct temporal mosaics of bare soils while accounting for soil moisture and soil management; (ii) which set of co-variates is more relevant to explain the SOC variability. From four sets of covariates, the best contributing set was selected, and the median SOC content along with uncertainty at 90% prediction intervals were mapped at a 25m resolution from quantile regression forest models. The accuracy of predictions was assessed by 10-fold cross-validation, repeated five times. The models using all the covariates had the best model performance. Airborne gamma-ray thorium, slope and S2 bands (e.g., bands 6, 7, 8, 8a) and indices (e.g., calcareous sedimentary rocks, "calcl") from the "late winter–spring" time series were the most important covariates in this model. Our results also indicated the important role of neighboring topographic distances and oblique geographic coordinates between remote sensing data and parent material. These data contributed not only to optimizing SOC mapping performance but also provided information related to long-range gradients of SOC spatial variability, which makes sense from a pedological point of view.

4.2 INTRODUCTION

Cropland soils cover a total of ~1500 million hectares worldwide distributed across different agroecosystems and climatic conditions, as well as diverse cropping systems and management practices (Ramankutty et al., 2008). Soil organic carbon (SOC) monitoring in croplands is crucial for climate change mitigation and food security (Lal, 2008; Paustian et al., 2016). The 4per1000 global initiative has increased the interest in more sustainable soil management and conservation in agricultural regions (Minasny et al., 2017). For about 20 years, the high demand for detailed soil information for policy implementation has driven digital soil mapping (DSM, McBratney et al., 2003). For instance, projects such as SoilGrids (Hengl et al., 2017; Poggio et al., 2021) and GlobalSoilMap (Arrouays et al., 2014b; Mulder et al., 2016a; Sanchez et al., 2009) aim through different approaches and techniques to produce soil property maps over the entire globe. Chen et al. (2022) recently published a review of GlobalSoilMap products on a broad scale, and SOC was the target soil property with the highest number of studies.

The use of satellite imagery to estimate SOC has been widely used including bands from the *Satellite Pour l'Observation de la Terre* (SPOT) (Vaudour et al., 2013), the Thematic Mapper (TM) imaging system of the Landsat satellite (Nanni and Demattê, 2006) and more recently studies of spectra obtained from bare soil imagery by Sentinel-2 (S2) (Castaldi et al., 2019a; Gholizadeh et al., 2018; Vaudour et al., 2019a). The S2 mission of the EU Copernicus program is a constellation of two satellites (Sentinel-2A and Sentinel-2B) launched in 2015 and 2017, respectively. Both satellites are equipped with the Multi-Spectral Instrument (MSI) with a revisit period of every 5 days and spatial resolutions of 10 m (bands 2, 3, 4 and 8) and 20 m (bands 5, 6, 7, 8A, 11 and 12). These high resolutions open up a wide range of applications such as SOC mapping in crop soils. Some previous studies at local and regional scales using S2 images in Brazil (Silvero et al.,

2021b), Belgium (Dvorakova et al., 2021), France (Urbina-Salazar et al., 2021; Vaudour et al., 2019a, 2019b), Czech Republic (Gholizadeh et al., 2018; Žížala et al., 2022) and Germany (Castaldi et al., 2019a) reported that the capability of S2 images to predict SOC performed better in some regions with different environmental and pedological characteristics than in others. For example, Vaudour et al. (2019b) reported different performances in two different regions in France, Versailles plain (R^2 of 0.56, RMSE of 1.23 $g.kg^{-1}$) and la Peyne Valley (R^2 of 0.02, RMSE of 3.71 $g.kg^{-1}$).

SOC mapping is generally hampered by several reasons: (i) low sample density for model calibration due to time and high analysis costs; (ii) soil sampling period of legacy data, when intensive agriculture and soil management practices vary over a period (Urbina-Salazar et al., 2021); (iii) impact of topsoil conditions and acquisition date of satellite imagery (Vaudour et al., 2019a; Urbina-Salazar et al., 2021); (iv) area of bare soil limited or covered by crop vegetation from single date images (Vaudour et al., 2019a, 2021); and (v) heterogeneous sampling in space and the presence of sampling clusters. Alternatives to overcome some of these issues have been proposed, for example choosing the best period of the cropping calendar to use images between sowing and emergence (Dvorakova et al., 2021) or extending the bare soil area through composite imaging using spectral indices alone (Demattê et al., 2018; Diek et al., 2017; Loiseau et al., 2019; Vaudour et al., 2021) or in combination with values of soil moisture products (SMPs) derived from Sentinel 1 and 2 images (Vaudour et al., 2021). However, there is still a gap related to the study of the ability to detect SOC via satellite images over large regional areas influenced by different soil-landscape formations, altitude, parent material, land-use history and management practices; which is crucial to support national or global mapping projects. Therefore, some frequent questions are, for example (i) is it feasible or not to use S2 imagery to map SOC in a particular area, where there might be a large or narrow spectral

variation (e.g. an area X with different landforms or a flat area Y with a large or slight variation of soil classes)? Or (ii) which environmental covariates are relevant and how do they help to map SOC spatial variability? Addressing these questions is important not only to optimize model prediction performances but also to understand in which cases (X or Y area) it is more or less useful to use satellite spectral data to support SOC digital mapping.

The Beauce region, located southwest of Paris is a limestone plateau known as "the granary of France" for its large production of cereal crops for more than 50 years at least (Viel, 2016). Cambisols and Luvisols are the main soil classes in this region. In addition, it is common to find silt, silt loam or silt clay loam layers derived from sediments formed by peri-glacial winds that redistributed Loess deposits across the area (Macaire, 1971). A higher SOC content favors aggregate stability, i.e. there is a greater physical stability that promotes soil structure, which is important in agriculture (e.g. Le Bissonais and Arrouays, 1997; Dexter at al., 2008). SOC improves water retention and drainage capacity, soil aeration and reduces the risk of erosion and nutrient loss by leaching (Lal, 2007). Knowing SOC is essential for agricultural soil management practices such as the application of exogenous organic matter (EOM, e.g. manure and green waste compost) to increase organic carbon and supply crop nutrient demand (Moinard et al., 2021). Therefore, there is a need to provide accurate information on SOC content variability through remote sensing data to contribute to the elaboration of maps of soil carbon status in agricultural regions, all the more than organic amendments spreadings are increasingly common in agricultural regions (Dodin et al., 2021).

The overall objective of this study is to assess the capability of Sentinel-2 imagery to map top SOC content over croplands in the Beauce region of France. In order to achieve this, we addressed (i) the use of S2 temporal mosaics of bare soil (S2Bsoil) over a 6-year period; (ii) the

use of soil moisture maps (SMPs) estimated at plot scale from Sentinel-1/2 images; and (iii) the use of legacy data and a number of environmental covariates derived from digital elevation model, remote sensing data, lithology (soil parent material) map and airborne gamma-ray data. Intending to generate a reliable SOC map at a 25 m-resolution, this paper also aims to answer two questions: (i) is it worthwhile to construct temporal mosaics of bare soils considering as many dates as possible or selecting dates using criteria, such as soil moisture and periods with different soil management practices instead, to obtain a better modelling performance? And (ii) in such an intensively managed agricultural region, is it possible, through DSM, to determine if the use of legacy data and environmental covariates explains the SOC variability and which set of covariates is more relevant?

4.3 MATERIALS AND METHODS

4.3.1 Study area and soil data

This study was conducted in the administrative region of Centre Val de Loire southwest of Paris in an agricultural region known as "Beauce", covering 4,838 km^2 (Figure 4.1). The Land Parcel Identification System (LPIS, in French "registre parcellaire graphique") of 2019 (French National Institute of Geographic and Forest Information (IGN)) determined that about 84% of the total area consists of croplands. Different crops are cultivated, mainly winter cereals and malting barley with crop rotations: rapeseed, sugar beet, maize or vegetables (Bouarfa et al., 2011). This study area covers three departments (Eure-et-Loir, Loir-et-Cher and Loiret) and is located over the Beauce aquifer, one of the main groundwater reservoirs in France (Verley, 2020). The climate is continental-oceanic with a mean annual temperature of 11.5 °C and a mean annual rainfall of 700 mm (Paroissien et al., 2014). Cambisols and Luvisols are the main soil classes observed in the region (IUSS Working Group WRB., 2015). These soils are developed from loess

deposits which covered a bedrock of tertiary continental lacustrine limestone, karstified and slightly weathered (Martelet et al., 2014). However, some Podzols, Gleysols, Fluvisols, Arenosols and Vertisols are more locally present, mainly at the southern border of the region.

We used the LPIS system to identify and used 391 cropland topsoil samples data from two databases used for DSM in France (Figure 4.1). The first set (341 topsoil SOC content data) corresponds to the French soil profile database (DoneSol) (to see more details, visit http://www.gissol.fr/, accessed on 15 September 2022), in which soil information (soil profiles and analyses) mainly came from data gathered for conventional soil mapping using points that were spread irregularly across the French mainland territory (Arrouays et al., 2020a). The second set (50 topsoil SOC content data) is the 2015 Land Use and Land Cover Survey (LUCAS) from the European Union Statistical Office (EUROSTAT) (Jones et al., 2020). As soils were all cultivated and regularly ploughed, the depth of sampling for DoneSol samples was the depth of ploughing which varied between 20 and 25 cm. The depth of sampling for LUCAS-Soil was 20 cm. As tillage homogenized SOC content in these layers, both samplings were comparable. Both LUCAS and French samples were ana-lyzed through dry combustion. Before starting modelling we checked the data set and removed SOC values equal to 0 and NA. Then we used the "duplicated" function of the R software to eliminate possible duplicate records.

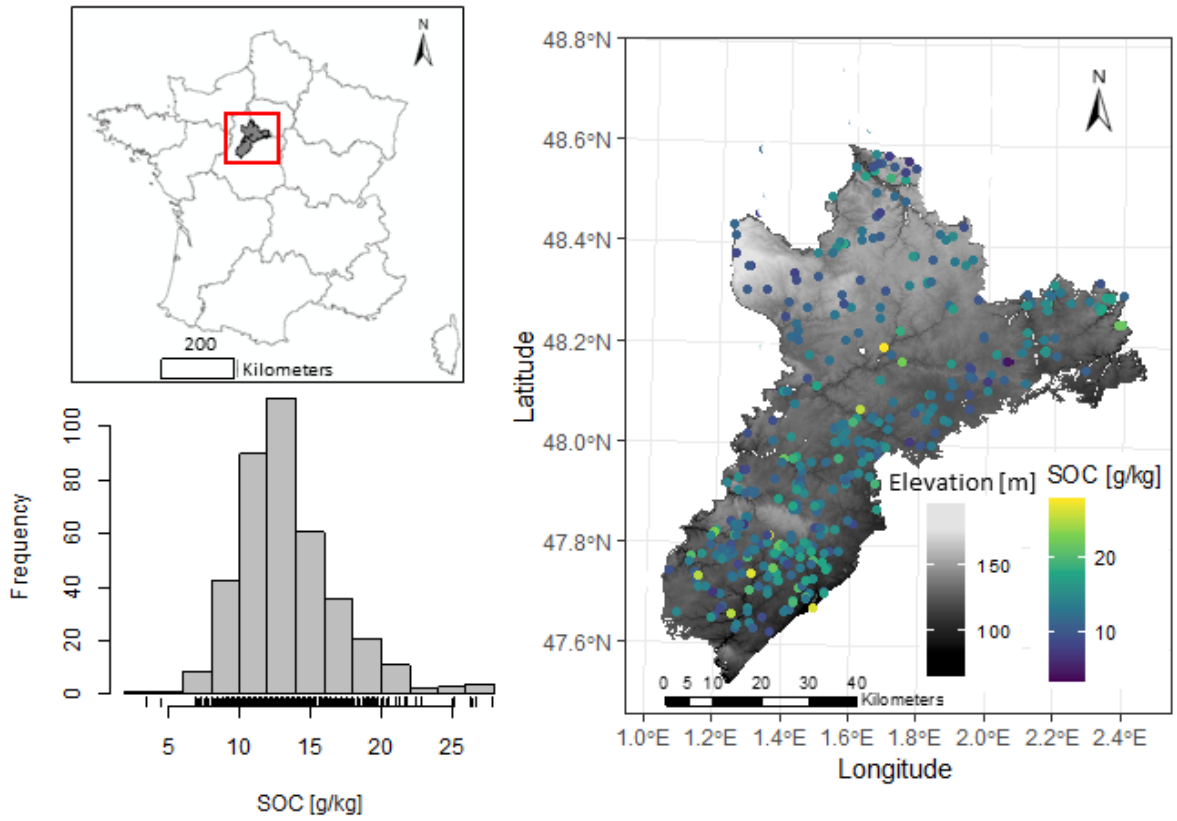


Figure 4.1. *Study area, histogram and density of SOC samples and sampling locations*

4.3.2 Sentinel-2 Time Series

The Sentinel-2 (S2) images were acquired between 2016 and 2021; three tiles were required to cover the entire study area (T31TCN, T31UCP and T31UDP). Therefore, a total of 213 images (from 71 different single dates) were downloaded from the Muscate platform from the French Land Data Centre (Theia, https://www.theia-land.fr/, accessed on 15 September 2022) (Figure 4.2a). These dates were selected according to the availability of S2 images on Theia's website as well as the presence of clouds. We selected those images with lower percentages of clouds (≤ 30%) covering their surfaces. Bands B2, B3, B4, B5, B6, B7, B8, B8A, B11 and B12 were used with slope effects atmospherically corrected ("flat reflectance" or FRE). In order to

prepare the S2 images, the tiles of each band were mosaicked (into 71 single-date images) and subsequently stacked and resampled at 25 meters using the "resample" function and the "nearest neighbor" method in R software, in order to ensure the same spatial extent of all spectral bands.

4.3.3 Soil Moisture Products

The soil moisture products (SMPs) derived from the Sentinel-1/Sentinel-2 satellites were provided by the Theia platform (https://www.theia-land.fr/product/humidite-du-sol-a-tres-haute-resolution-spatiale/ (accessed on 23 September 2022)). SM image dates that were as close as possible to the acquisition dates of the S2 images were selected (Table S1a,b. supplementary material). The SMPs were obtained over croplands and grasslands at plot scale and provided SM estimates with an approximate accuracy of 5 vol.% (El Hajj et al., 2017) with a six-day temporal resolution (Bazzi et al., 2019; El Hajj et al., 2017). To estimate SM values (0–10 cm depth), El Hajj et al. (2017) inverted the water cloud model parameterized by Baghdadi et al. (2017) for the C-band combined with the integral equation model as modified by Baghdadi et al. (2006). This algorithm inverts Sentinel-1 radar data to SM values and uses the normalized differential vegetation index (NDVI) derived from S2 optical data from agricultural plots as an input. SMPs were clipped to the size of the study area and the main statistics were computed. For some dates (27 March 2020, 23 August 2016 and 25 July 2019), full coverage of the area was not achieved because no data was available for the same date in some parts of the study area on Theia's web platform (Figure S1a,b. supplementary material).

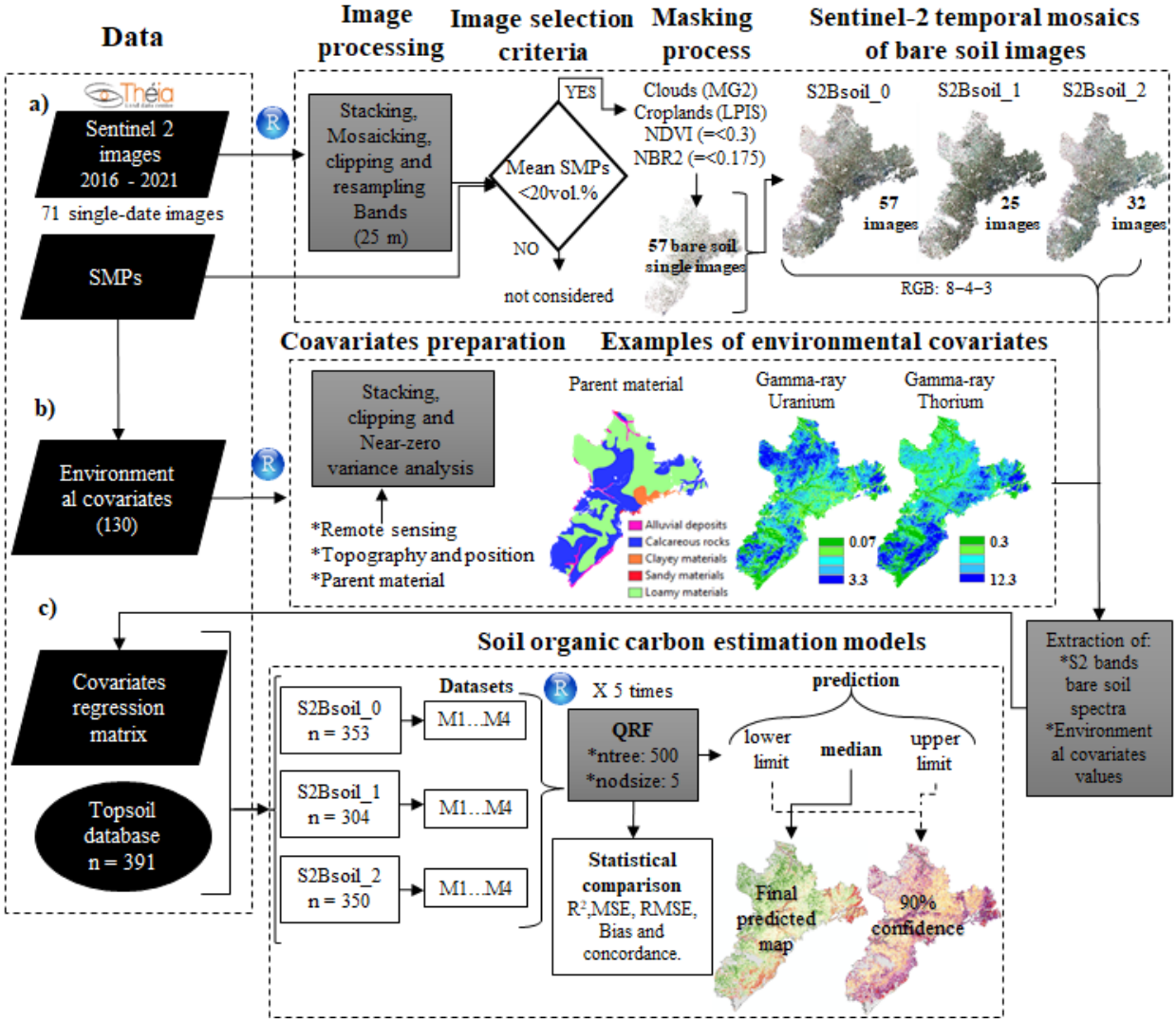


Figure 4.2. *Flowchart methodology. Sentinel-2 temporal mosaics of bare soil (S2Bsoil) over 6 years: S2Bsoil_0, 57 images used; S2Bsoil_1, 25 images between February and May; S2Bsoil_2, 32 images between July and November. Sets of covariates: M1, the 10 S2Bsoil bands; M2, the 10 S2Bsoil bands plus spectral indices were considered (24 covariates); M3, the same covariates used in (M2) plus soil moisture were used (25 covariates); M4, all covariates used in (M1 to M3) plus covariates of topography, and position and parent material (85 covariates*

4.3.4 Temporal mosaics for extending bare topsoil

One of the main challenges to obtain information on soil properties such as SOC through spectral reflectance acquired by satellites is to

achieve a sufficient percentage of bare soil coverage. A large extent of bare topsoil makes it possible to use a larger number of soil samples over the whole area to detect SOC spatial variations. Bare topsoil extension techniques with satellite data have been developed around the world, reporting good accuracy in obtaining bare soil and as input for soil attribute modelling (Demattê et al., 2018; Vaudour et al., 2021; Zepp et al., 2021; Žížala et al., 2022). Therefore, we focused our study on bare soil using S2 data. To create the Sentinel-2 temporal mosaic of bare soil (S2Bsoil) (Figure 4.2a), the processing involved 4 main steps:

1. Vaudour et al. (2021) reported better performances predicting SOC when soils with mean volumetric water content values higher 25vol.% were excluded. In order to eliminate highly wet soils, we chose S2 images using as criteria the mean pixel values of SMPs (< 20vol.%) for each date. In total 57 S2 images were considered for the next step and the LPIS was used to mask all non-agricultural sites (Figure 4.2a).
2. A geophysical mask was used to remove clouds and/or topographic shadows and/or snow cover in each single-date S2 image ("masque géophysique" or MG2) (Baetens et al., 2019); this mask is available for all S2 images that can be downloaded from the Theia website.
3. NDVI (Normalized Difference Vegetation Index) and NBR2 (Normalized Burn Ratio 2) were calculated to mask the vegetated (or covered by crop residues) pixels. Following some previous studies (Urbina-Salazar et al., 2021; Vaudour et al., 2021), we considered as bare soil, NDVI values between 0 and 0.30 and values above were flagged as NA. NBR2 values > 0.175 were marked as NA to exclude sites covered by straw or crop residues. Castaldi et al. (2019a) reported that the most suitable threshold for good performance in SOC prediction models was considering NBR2 index values up to 0.175. At this point, 57

single-date images of bare soil were obtained.

4. Finally, we calculated the median reflectance of all the images to obtain S2Bsoil.

According to previous studies (Urbina-Salazar et al., 2021; Vaudour et al., 2019a, 2021), climatic conditions, soil physical conditions and maximum bare soil coverage (conditioning the sample size) influence the digital mapping performance of soil organic carbon. In this sense, we selected all 57 images and separated them by subsemester to evaluate two periods where soil conditions vary by crop rotation and management practices as well as seasonal changes "late winter-spring" and "summer-autumn". Therefore, we obtained an S2Bsoil image based on all the images (S2Bsoil_0), one using the images between February and May (25 bare soil images) (S2Bsoil_1) and another from the images between July and November (32 bare soil images) (S2Bsoil_2) over a 6-year period (Figure 4.2a).

4.3.5 Environmental covariates

The covariates used in this study are listed in Table 4.1 separated into remote sensing data, topography, position and parent material. We used these covariates based on the main environmental factors of the scorpan model that may control the variability of the SOC proposed by McBratney et al. (2003). The resolution and scales of the environmental covariates were different. Therefore, we harmonized them to 25 m resolution using the nearest neighbor method for spatial predictive modelling and mapping at non-visited locations. In addition, the nearZeroVar function of the caret package (Kuhn, 2022) in R (R Core Team, 2020) was used to remove covariates with zero or near-zero variance (Figure 4.2b).

Table 4.1. *Environmental covariates used as predictors for digital soil mapping.*

Variable	Number	Scale/ resolution	Expression[a]	Reference
Remote sensing				
S2Bsoil_0[b]	10	25 m		This study
S2Bsoil_1[b]	10	25 m		This study
S2Bsoil_2[b]	10	25 m		This study
SM_0[c]	1	25 m		This study
SM_1[c]	1	25 m		This study
SM_2[c]	1	25 m		This study
Hue index (HI)	3	25 m	(2*R-G-B)/(G-B)	Mathieu et al., 1998
Grain size index (GSI)	3	25 m	(R-B)/(R+G+B)	Xiao et al., 2006
Calcareous Sedimentary rocks(CalcI)	3	25 m	(SWIR 1-G)/(SWIR 1+G)	Boettinger et al., 2008
Saturation Index (SI)	3	25 m	(R-B)/(R+B)	Mathieu et al., 1998
Brightness Index (BI)	3	25 m	$\sqrt{(R^2+G^2+B^2)}/3$	Escadafal. 1989b
Coloration Index (CI)	3	25 m	(R-G)/(R+G)	Pouget et al., 1991
Carbonate index (CaI)	3	25 m	R/G	Boettinger et al., 2008
Geological response (Geol)	3	25 m	(SWIR1-SWIR 2)/(SWIR 1+SWIR 2)	Nield et al., 2007
First three PCs of NDVI S2Bsoil[d]	9	25 m		This study
First three PCs of monthly MODIS NDVI[d]	3	500,300 m		Loiseau et al., 2019
Topography and position				
Elevation	4	25 m		IGN (2011); Chen et al., 2021
Slope	4	25 m		IGN (2011); Chen et al., 2021
Slope position (PS)	4	25 m		IGN (2011); Chen et al., 2021
Slope length (LS)	4	25 m		IGN (2011); Chen et al., 2021
Terrain wetness index (TWI)	4	25 m		IGN (2011); Chen et al., 2021
Valley depth (VD)	4	25 m		IGN (2011); Chen et al., 2021
Vertical distance to channel network (VDCN)	4	25 m		IGN (2011); Chen et al., 2021
Multiresolution index of valley bottom flatness (MrVBF)	4	25 m		IGN (2011); Chen et al., 2021
Channel network base level (CNBL)	4	25 m		IGN (2011); Chen et al., 2021
Plan curvature	4	25 m		IGN (2011); Chen et al., 2021
Profile curvature	4	25 m		IGN (2011); Chen et al., 2021
Coordinates (Latitude, Longitude)	2	25 m		Chen et al., 2021; Møller et al., 2019
Oblique coordinates (OC)[e]	10	25 m		
Parent material				
Parent material	1	1:1M		King et al., 1994
Gamma ray (K, U, Th, TC)	4	250 m		Martelet et al., 2014

[a] R: Red; G: Green; B: Blue; SWIR1: short-wave infrared1; SWIR2: short-wave infrared2. [b] S2Bsoil (0,1,2): Sentinel-2 temporal mosaics of bare soil images. [c] SM (0,1,2): soil moisture product means calculated for same dates as S2Bsoil (0,1,2). [d] PCs, principal components; NDVI, normalized difference vegetation index. [e] Oblique coordinates at angles of 15°, 30°, 45°, 60°, 75°, 105°, 120°, 135°, 150° and 165°.

4.3.5.1 Remote sensing

The 10 bands of the three S2Bsoil were used and SMPs pixelwise means were calculated considering the same bare soil mosaic periods; using all images (SM_0), between February and May (SM_1) and between July and November (SM_2). Several spectral indices related to parent material and soil were derived from the bands of the three S2Bsoil. Table 4.1 provides more details of the 8 indices calculated from the S2 spectral bands for each mosaic.

NDVI by single date was used in this study; the 57 NDVI dates were collected and reduced to the first three principal components (PCs) (keeping the same periods of the S2Bsoil) by principal component analysis (PCA) to eliminate their multicollinearity. We chose to use PCA under the assumption that the sensitivity of spectral indices such as NDVI to SOC and soil texture are linked to different soil types and parent material (Demattê et al., 2017; Gholizadeh et al., 2018). In addition, the use of PCA applied to NDVI time series has shown the potential to identify seasonal changes in land cover (Bellón et al., 2017). We also included the three PCs of 24 monthly NDVI data in 2003 (extremely warm and dry year) and 2016 (normal year) acquired from MODIS (500 m resolution) and the PROBA-V 10-day product level 2B TOC (300 m resolution). For more details, we refer to Loiseau et al. (2019).

4.3.5.2 Topography and position

The Digital Elevation Model (DEM) of mainland France (25 m resolution) derived from the BD TOPO 3 of the French National Geographic Institute (IGN, 2011) was used. The covariates derived from the DEM were computed in the SAGA GIS Channels Library (Conrad et al., 2015) (Table 4.1). According to McBratney et al. (2003), neighboring locations on the topography can provide useful information in soil modelling. Grinand et al. (2008) and Loiseau et al. (2019) investigated the potential of incorporating local neighborhood information into

training pixels using convolution filtering operations to predict soil class distribution and clay content respectively. In a recent study, Chen et al. (2021) used mean convolution circular windows to calculate the focal means to map the soil thickness of loess deposits in our study area. We therefore included the focal means of topography covariates with radius at 200, 500 and 1000 m, which were produced earlier. For more details, we refer to Chen et al. (2021). Finally, the position covariates (Latitude and Longitude) extracted for each recorded pixel were used for modelling. In addition, 10 oblique geographic coordinates (OGC) at angles of 15°, 30°, 45°, 60°, 75°, 105°, 120°, 135°, 150° and 165° computed by Chen et al. (2021) were included as covariates.

4.3.5.3 Parent material

The parent material raster was obtained from the French national parent material map (King et al., 1994). Alluvial deposits, clayey materials and sandy materials are present in the study area. However, calcareous rocks and loamy materials are the most common parent materials (42 and 49% of the area respectively) (Figure 4.2b). Parent material was also characterized by means of airborne gamma-ray spectrome-try, which measures the emission of three naturally occurring elements that have radi-oisotopes with sufficient energy and intensity to be measured from the air: uranium (U), thorium (Th), and potassium (K) (Minty, 1997). K is measured directly through the isotope 40K, whereas 232Th and 238U are measured through their radioactive daughters (208Tl and 214Bi, respectively). These emissions capture information from the upper 50 cm of the soil/rocks that is related to their geochemical composition, mineralogy, and weathering. The gamma-ray data (K, Th, U) and total count (TC)) used in this study were acquired by an airborne high-resolution magnetic and radiometric survey over the Région Centre by Terraquest Ltd Canada®, under the supervision of the BRGM (Martelet et al., 2014). The survey was conducted during the autumn and winter of 2008-2009. In order to

minimize the effects of soil moisture from rainfall on gamma ray emissions during rainy days the measurements were interrupted for several hours after rainfall. The footprint of each measurement is an ellipse of approximately 150 × 300 m^2 and the data were interpolated using a standard minimum curvature interpolation on 250 m grids. Gamma-ray emissions being presumably stable over the time-span since the airborne acquisition, there was no need to match Sentinel-2 data.

4.3.6 Datasets for modelling

The flowchart in Figure 4.2c summarizes that before SOC modelling (see, section 4.3.7), we focused on the bare soil mosaics elaborated in section section 4.3.4. First, we extracted the band values of all S2Bsoil and other covariates (Table 4.1) and obtained our regression matrix. Then, the set of covariates was divided into 4 sets for each S2Bsoil:

- ✓ M1, the bare soil reflectance of the 10 S2Bsoil bands was used (10 covariates);
- ✓ M2, The 10 S2Bsoil bands plus spectral indices were considered (24 covariates);
- ✓ M3, The same covariates used in (M2) plus soil moisture were used (25 covariates);
- ✓ M4, All covariates used in the previous models plus covariates of topography, position and parent material were integrated (85 covariates).

Note that for each S2Bsoil we obtained four different models, i.e. sixteen models in total were run.

4.3.7 Quantile regression forest and model performance evaluation for SOC prediction

The challenge of producing more accurate and robust digital soil property maps has driven the use of several machine learning models. Quantile regression forest (QRF, Meinshausen, 2006), based on the

random forest model (RF, Breiman, 2001) has been largely used at regional and national scales (Chen et al., 2021; Lagacherie et al., 2019; Loiseau et al., 2019; Lombardo et al., 2018; Vaysse and Lagacherie, 2017). Therefore, QRF was used to model SOC in this study (Figure 4.2c).

QRF algorithm estimates a large set of bootstrap trees by randomly sampling of n independent observations. For each node of a bootstrap tree, several dependent variables (covariates) are randomly sampled to split the observations. For all decision nodes, QRF, in contrast to RF, keeps the value of all observations, not just their mean as RF. In addition, from an analysis of the conditional distribution based on the stored information, the prediction intervals are calculated.

We used the package quantregForest (Meinshausen, 2017) in R version 4.2.0 for implementing QRF to derive the median prediction (50% quantile) and 90% prediction intervals (PIs, 5% and 95% quantiles). In order to obtain comparable and robust statistics, we optimized the selection of observations from the dataset in each model (see section 4.3.6.) by 10-fold cross-validation with the caret package (Kuhn, 2022) in R (R Core Team, 2020), and through the for function 10 models were generated by selecting a different calibration and validation set each time. This process was repeated 5 times, i.e., 50 repeated models were obtained. QRF is often described as a method that is not very sensitive to overfitting. However, this is not completely true. The repeated k-fold cross-validation avoids putting all the samples in model fitting and allows to control a posteriori if over-fitting occurs (i.e. if learning performances are much better than validation performances). A basic rule to avoid overfitting consists of putting in the learning covariates set, covariates that have a real soil process meaning. More sophisticated selections of co-variates like recursive feature elimination (Guyon et al., 2002) or Boruta algorithm (Kursa et al., 2010) exist, but they are often used when the set of covariates is very large and/or redundant. We did not use these selections in this study.

The default number of tree (ntree = 500) and the minimum node size (nodsize = 5) were considered. The number of variables randomly sampled as candidates in each split (mtry) was set as default (total number / 3) for each model. The QRF includes the option to calculate the variable importance determined by the increased mean squared error (IncMSE, in %) between the model excluding and including a given variable (Chen et al., 2021). Variable importance was calculated by averaging the 50 repeated models mentioned above. Performance was evaluated on the validation sets by the coefficient of determination (R^2, Eq.(1)); root mean square error (RMSE, Eq.(2)); bias (Eq.(3)) and the concordance (Eq.(4)) (Lin, 1989):

$$R^2 = 1 - \frac{\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n}} \quad (2)$$

$$bias = \sum_{i=1}^{n}\frac{(\hat{y}_i - y_i)}{n} \quad (3)$$

$$Concordance = \frac{2*\sum_{i=1}^{n}\frac{(y_i - \bar{y})*(\hat{y}_i - \ddot{Y})}{n}}{\sigma_{y_i}^2 + \sigma_{\hat{y}_i}^2 + (\bar{y} - \ddot{Y})^2} \quad (4)$$

where *n* is the number of samples, y_i and $\hat{y}_i$ are the the observed value and the predicted value for sample i, $\bar{y}$ and $\ddot{Y}$ is the mean of the observed and predicted values; $\sigma_{y_i}^2$ and $\sigma_{\hat{y}_i}^2$ are the observation and prediction variances.

After predictions and performance assessments, we selected, mapped, and discussed the model approach that performed the best. The 90% PIs were used to display the local uncertainty of the SOC map. In addition, we used 12 independent soil samples from the French soil monitoring network (RMQS; Jolivet et al., 2006; Mulder et al., 2016a, 2016b) to evaluate the SOC values of the predicted map in agricultural plots regarding this monitoring network that extends across France in a 16 km × 16 km square grid sampling.

4.4 RESULTS

In the following paragraphs, the three temporal bare soil mosaics (S2Bsoil_0, S2Bsoil_1, S2Bsoil_2) generated in this study are first presented. Subsequently, the SOC content prediction performance is presented according to the aforementioned mosaics and the influence of the most important environmental covariates on the modelling is discussed. Lastly, a predicted Top SOC content map (average predicted SOC values from the 50 models) at 25 m resolution per pixel and its uncertainty are provided and analyzed.

4.4.1 Maximum bare topsoil area mapped and description of spectral patterns

The mean percentage of bare soils on single dates between February and May was lower than the mean between July and November (~24% and ~41% of bare soil pixels, respectively) (Table S2a,b. supplementary material). S2Bsoil_0 with 57 images considered achieved the highest percentage of bare topsoil mapping (81.8%), followed by S2Bsoil_2 and S2Bsoil_1 (81.4% and 70.9%, respectively) (Figure 4.3). The difference between S2Bsoil_0 and S2Bsoil_2 (0.4%) showed that detecting a bare soil cover higher than 80% of the agricultural area was possible with a 32-dates time series. Furthermore, we found that all S2Bsoil obtained an acceptable potential to recover bare soil pixels, confirmed by the closeness of the scattering points to the 1:1 soil line (Baret et al., 1993). The coefficients of determination in the soil line analysis were high (R^2 > 0.8) (Figure 4.3); similar results were reported by Silvero et al. (2021b) who used Sentinel and Landsat images to extend the area of bare soil.

Table 4.2 displays the statistics of SOC content in the Beauce area for the sites of each bare soil mosaic. As expected, due to their percentage of bare soil, S2Bsoil_0 and S2Bsoil_2 obtained the highest number of sites from our database (353 and 350 sites, respectively), while S2Bsoil_1 used 304. The statistics among the three sets were similar,

SOC ranged from 3.5 to 27.8 $g.kg^{-1}$ (mean of 13.4 $g.kg^{-1}$; median of 13 $g.kg^{-1}$; standard deviation of 3 $g.kg^{-1}$). A skewness between 0.85 and 0.93 indicated that the data were positively but moderately skewed, while a kurtosis > 3 showed that the data had a heavy-tailed Leptokurtic kurtosis. A logarithmic transformation was used to convert the data to a normal distribution, however this did not improve the model performance. Therefore, we used the original SOC data for spatial modeling in this study.

Figure 4.4 shows for each bare soil mosaic (S2Bsoil_0, S2Bsoil_1 and S2Bsoil_2) the spectral patterns of median reflectance calculated from the quartile values of the SOC samples used for each S2Bsoil. As expected the spectral features of the first quartiles (i.e. the lowest SOC content ≤ 11.4 $g.kg^{-1}$) showed the highest reflectance values while the spectral features with the lowest reflectance corresponded to the quartiles with the highest SOC content. Generally, a difference between SOC quartiles was observed confirming the quality of each mosaic. However, in S2Bsoil_1 such difference was more noticeable especially in quartiles 3 and 4 whereas in S2Bsoil_0 and S2Bsoil_2 these two quartiles had quite similar values. The number of samples used for each bare soil mosaic in Figure 4 was lower than the number of samples mentioned in Section 4.3.1. (391 samples) because they were the available samples located in bare soil pixels for each mosaic (S2bsoil). That is, as each S2Bsoil was elaborated with images in different periods, the bare soil pixels and their location varied, then for each mosaic the number of samples varied and was always lower than 391.

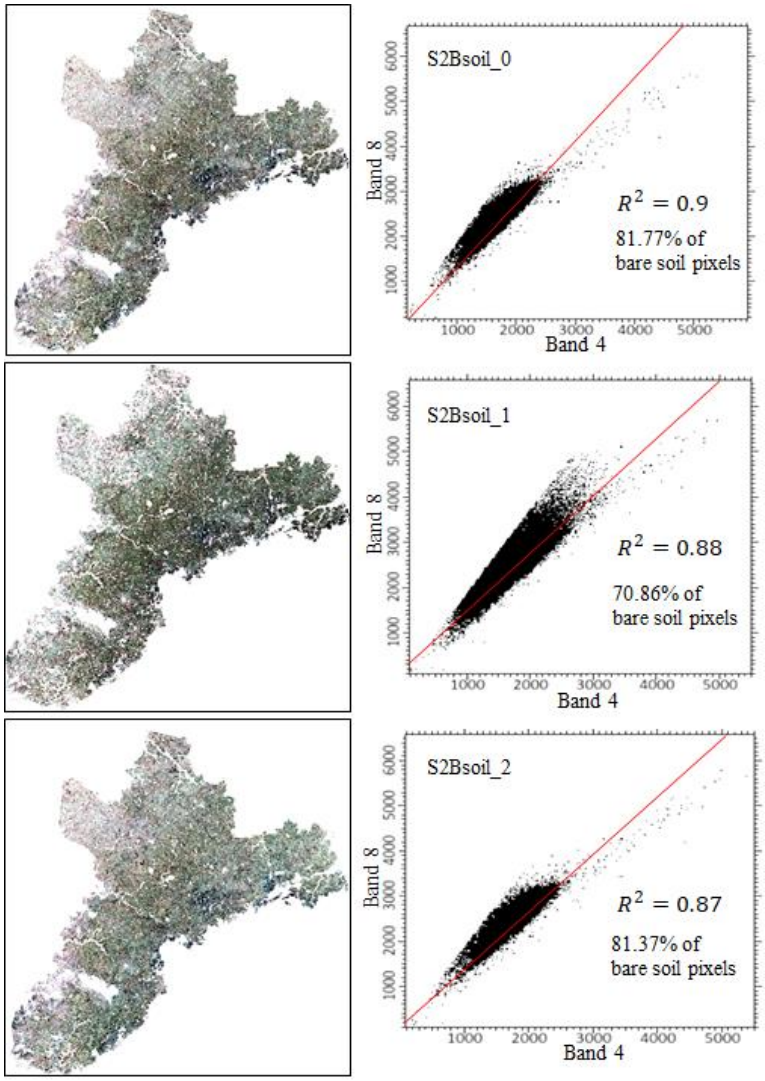


Figure 4.3. *Sentinel-2 temporal mosaics of bare soil images (S2Bsoil_0, S2Bsoil_1, and S2Bsoil_2), infrared false color composition (R,G,B: 8,4,3). Relationship between NIR (Band 8) and Red (Band 4) bands are presented to show the soil line for each S2Bsoil. The value of the coefficient of determination (R^2) was used as an indication of bare soil retrieval.*

Table 4.2. *SOC sample set statistics for each S2Bsoil used for modelling*

S2Bsoil	sites	SOC g.kg^{-1}								
		Minimum	Q1	Median	Mean	Q3	Maximum	SD	Skewness	Kurtosis
S2Bsoil_0	353	3.46	11.40	13	13.4	15	27.8	3.34	0.86	4.83
S2Bsoil_1	304	3.46	11.41	13	13.4	15	27.8	3.32	0.93	5.18
S2Bsoil_2	350	3.46	11.39	13	13.4	15	27.8	3.35	0.85	4.82

Q1, the first quartile; Q3, the third quartile.

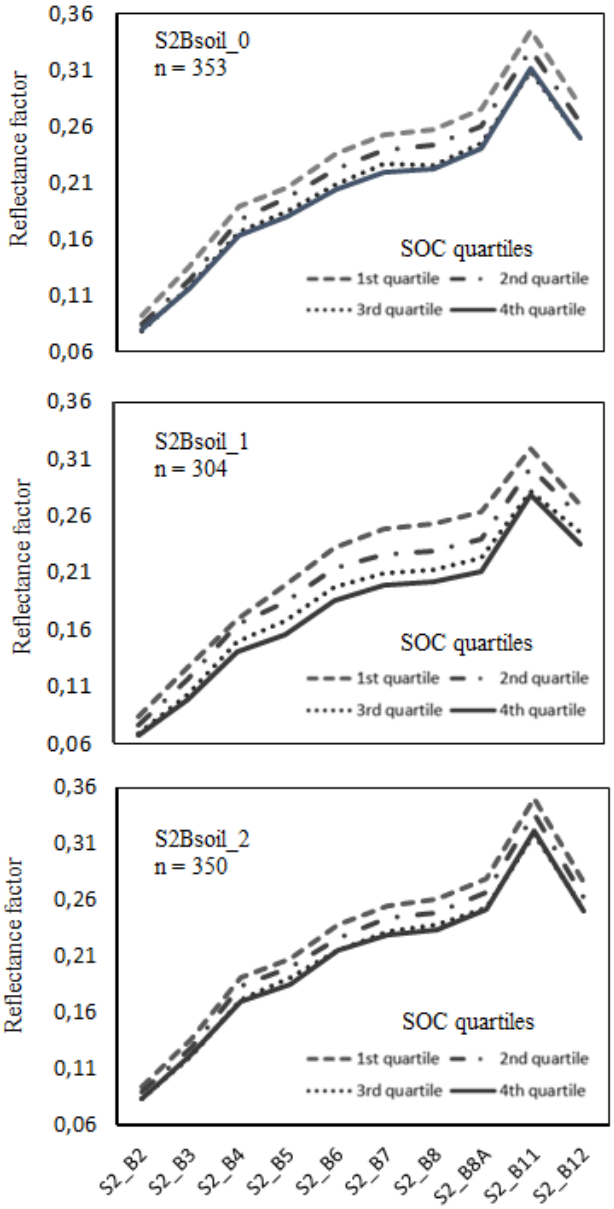


Figure 4.4. *Spectral patterns of S2Bsoil_0, S2Bsoil_1 and S2Bsoil_2 considering the median of the SOC sample values by quartiles.*

4.4.2 SOC model performance

The performance of the SOC content predictions varied according to the S2Bsoil and the environmental covariate datasets used. The best model performances were obtained using all covariates for each S2Bsoil (sets M4, e.g. $0.26 \leq R^2 \leq 0.33$ and $2.59\ g.kg^{-1} \leq RMSE \leq 2.75\ g.kg^{-1}$) (Table 4.3). We obtained decreasing performances when only the S2 bands were considered. All M1 models exhibited the lowest performance, $0.11 \leq R^2 \leq 0.18$, $RMSE \geq 2.97\ g.kg^{-1}$ and maximum bias of $-0.35\ g.kg^{-1}$. The performances of the models considering S2Bsoil_0 and S2Bsoil_2 showed a close result; however, the maximum RMSE reached ($3.17\ g.kg^{-1}$) was obtained using only the set of spectral bands (M1) of the bare soil mosaic produced from S2 images between July and November (S2Bsoil_2). The poor performance of the S2BSoil_2 models may presumably be associated with the SM measured at each single date. The SMPs between July and November of our time series showed a slightly higher mean value (SM ≈ 13.9 vol.%) compared to those between February and May (SM ≈ 13.2 vol.%) for S2Bsoil_1 (Figure S1a,b. supplementary material). In addition, the inter-day difference (ds2 - dsm) between SMPs and S2 images between July and November was higher with respect to the scenes considered between February and May, so it is possible that a bias of the real moisture values in the single dates of this particular period affected the modeling performance.

Table 4.3. *Model performance of Quantile random Forest to predict soil organic carbon by mosaics of bare soil over a 6-year period (The values presented correspond to the mean values of the 50 repetitions mentioned in section 4.3.7.)*

S2Bsoil	Modelling dataset	R^2	RMSE ($g.kg^{-1}$)	Bias	Concordance
S2Bsoil_0	M1	0.18	3.00	-0.33	0.32
	M2	0.19	2.98	-0.31	0.33
	M3	0.15	2.98	-0.30	0.29
	M4	0.26	2.75	-0.20	0.40
S2Bsoil_1	M1	0.19	2.97	-0.32	0.35
	M2	0.22	2.90	-0.30	0.35
	M3	0.22	2.79	-0.28	0.34
	M4	0.33	2.59	-0.22	0.42
S2Bsoil_2	M1	0.11	3.17	-0.35	0.25
	M2	0.11	3.14	-0.30	0.24
	M3	0.12	3.00	-0.29	0.25
	M4	0.27	2.71	-0.21	0.39

R^2, coefficient of determination; RMSE, root mean square error.

4.4.3 Influential covariates

Figure 4.5 displays the top 25 environmental covariates in the computed QRF models. This figure compares three models for each temporal bare soil mosaic using the M4 dataset. The number of contributing covariates originating from remote sensing (satellite S2, S2 indices and SM in addition to gamma-ray) was 13 or 14 out of 25 for the temporal mosaics of all dates (S2Bsoil_0) or of "late winter-spring" dates (S2Bsoil_1) while that of topography and position was slightly higher (15) for the temporal mosaic of "summer-autumn" (S2Bsoil_2) only. It is worth noting that satellite and gamma-ray covariates prevailed in the top five influent covariates, whatever S2 temporal mosaic, and that all S2 bands were amongst influent covariates. The most significant bands among the best performing models were from the near-infrared (b6, b7, b8 and b8a) and SWIR (b11, b12) regions, and to a lesser extent, from the visible region

through the CalcI index (b3). While all S2 bands turned out to be influent, only three out of eight soil spectral indices (CalcI, followed by BI then HI) had a predictive weight for Beauce. The best temporal series was composed of "late winter-spring" dates when bare soil is not likely to be disturbed by crop residues; the S1-derived soil moisture appears influent using only this temporal mosaic, which makes sense noting the closeness of the dates between the S2 images with which S2Bsoil_1 was constructed and the SMPs mentioned above in section 4.4.2.

Our results also indicate the fundamental role played by the information of neighboring distances of topography in the three models, particularly for the "summer-autumn" temporal mosaic: slope (with 200 m radius, SLOPE8); slope length (with 200 m radius, LS8); DEM (with 1000 m radius, DEM40) and MrVBF (with 1000 m radius, MRVBF40). The OGC at an angle of 135° (oblique135) was among the 25 most important covariates in in all bare soil mosaics.

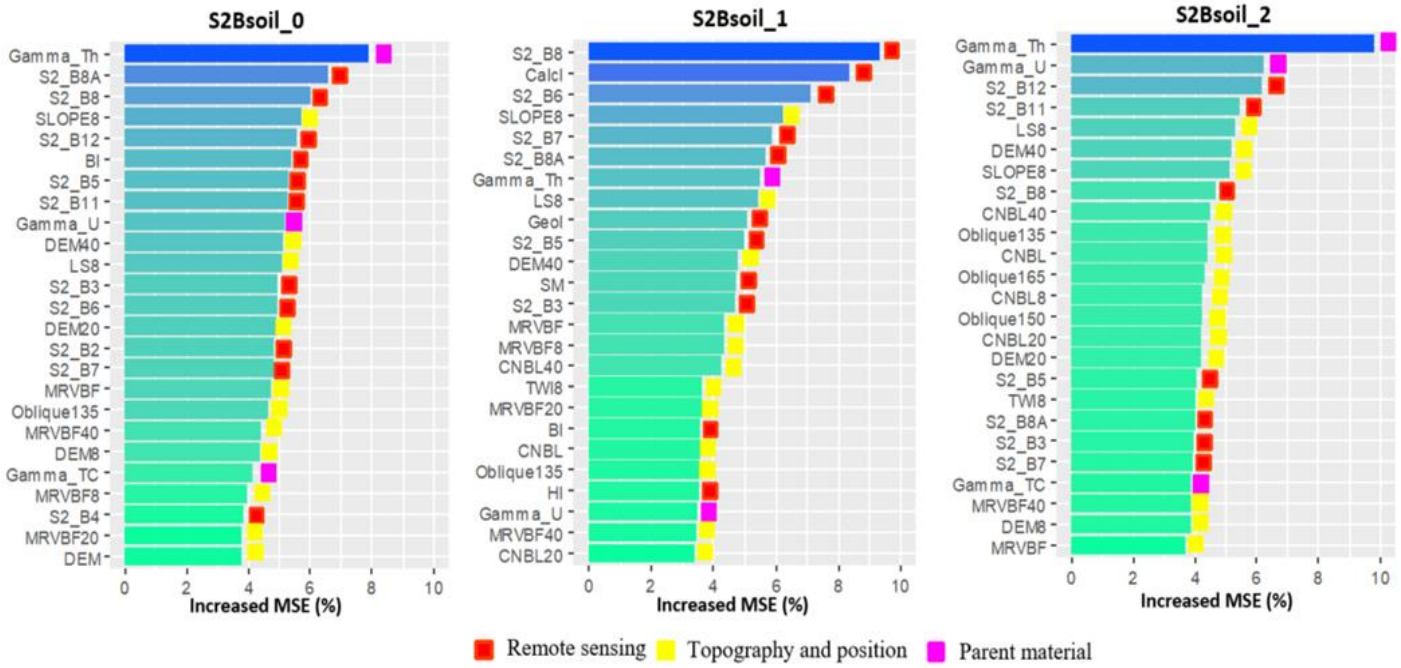


Figure 4.5. *Relative importance of the first 25 environmental covariates for SOC prediction (M4 datasets).*

Furthermore, both Thorium and Uranium surface concentrations derived from airborne gamma-ray surveys also played an important role in the prediction of SOC, especially for S2Bsoil_2 and S2Bsoil_0. In DSM, gamma-ray spectrometry can be used as substitute of lithological maps (Loiseau et al., 2020). A recent review (Reinhardt et al., 2019) pointed out the main applications of gamma rays in mapping some soil attributes (e.g. parent material, topsoil soil texture, clay mineralogy, weathering intensity, soil type). A strong limitation is that gamma-ray data are far from covering the whole world (and even the whole France) because of their cost and because they have been mainly used for mining purposes. Although the use of airborne gamma-ray data in soil science has been mainly explored to assist digital soil mapping and to understand landscape processes, lithology and mineralogy (Chen et al., 2021; Martelet et al., 2013; Wilford and Minty, 2006), there are still few papers that relate soil attributes to gamma radiation emissions at a rather broad-scale (van Egmond et al.,

2010; Van Der Klooster et al., 2011; Loiseau et al., 2020; Mello et al., 2022). Thorium is rather stable under conditions of soil at about or higher than pH>=7 where it can be adsorbed to the surface of clay minerals and organic matter (Von Gunten et al., 1996). Note that under intensive cultivation and over a calcareous bedrock, this is the case (Richer-de-Forges, 2008). Therefore, it is not surprising that gamma-ray data emitted by thorium contribute to this study by predicting SOC. Note that in Figure 4.1 in the southeast of the study area, high SOC values are observed in the sampling sites and in Figure 4.2b in the same part the thorium image shows the highest values. This zone is characterized by the presence of soils with clayey texture. An opposite effect is observed to the north in a large plateau with a deep loess thickness that was prone to a loss of clay content in topsoil by a process of illuviation (Chen et al., 2021), in this area the values of gamma radiation emitted by thorium are lower.

4.4.4 SOC variability and predicted spatial uncertainty

Figure 4.6a displays the predicted SOC map calculated from the S2Bsoil_1 bands and all covariates in the M4 data set. The mean and standard deviation of the predicted SOC content values were 13.2 $g.kg^{-1}$ and 1.6 $g.kg^{-1}$ respectively, i.e. quite similar to the mean and distribution of SOC observed values (Table 4.2 and Figure 4.1). The map of SOC clearly exhibits some gradients in SOC distribution. These gradients depict some general trends that could be linked to relationships between clay and SOC content. Indeed, the northern area, which is highly depleted in clay due to illuviation, has very low SOC contents. This area corresponds to a wide loessic plateau already identified by Chen et al. (2021) when studying the thickness of loess in the same region. The effect of local topography is also very striking with numerous high values close to the valleys.

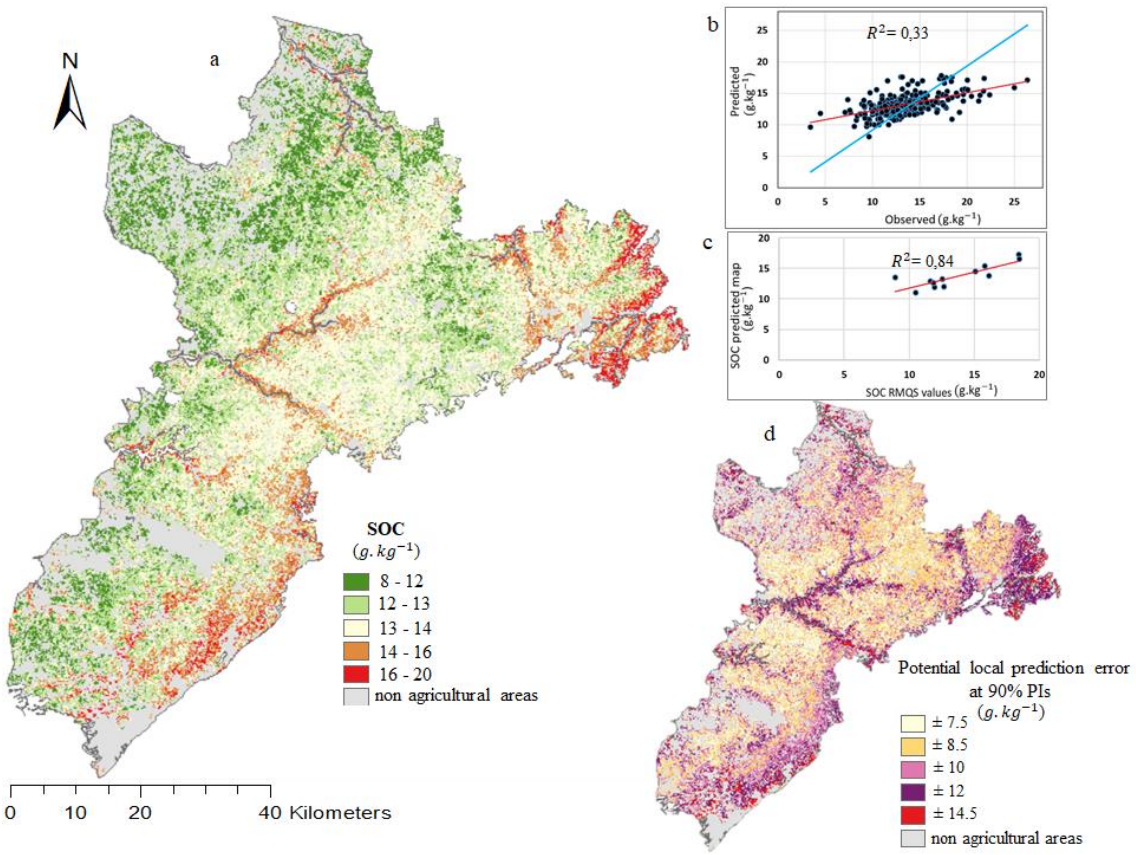


Figure 4.6. *a) SOC predicted map (S2Bsoil_1, using the M4 dataset); b) scatter plot of observed versus predicted values on the validation set (Red line corresponds to the line fitted to the points and blue line to the 1:1 line); c) scatter plot of SOC values from the RMQS monitoring network versus predicted map values; and d) map of potential local prediction error.*

Note, however, that the validation results are far from the 1:1 line (Figure 4.6b). In other words, there is a general tendency to overestimate the low SOC values and to underestimate the high ones. The low sample density of higher and lower SOC values in our set (Figure 4.1) might have influenced the prediction accuracy at the extreme values in the modelling. Note also that the same trend is observed for the 12 points of the systematic independent grid (Figure 4.6c). The remote sensing data and covariates used in our model were significant in mapping mainly mean SOC values in mid-elevation areas to lower values in flat areas over plateaus at high elevation (Figure

4.6d). Nevertheless, 90% prediction intervals were large. The wind direction of the Loess deposits observed mainly in the large plateau to the north was from the northwest (Bertran et al., 2016; Borderie et al., 2017; Chen et al., 2021) and interestingly, one of the most important covariates was oblique135. Figure 4.7a shows this covariate divided into four classes. Dividing our predicted SOC map and thorium gamma image into the same classes provides a striking matching (Figure 4.7b and 4.7c, respectively), in which a trend from lower SOC and thorium values from the northwest to higher values towards the southwest is observed.

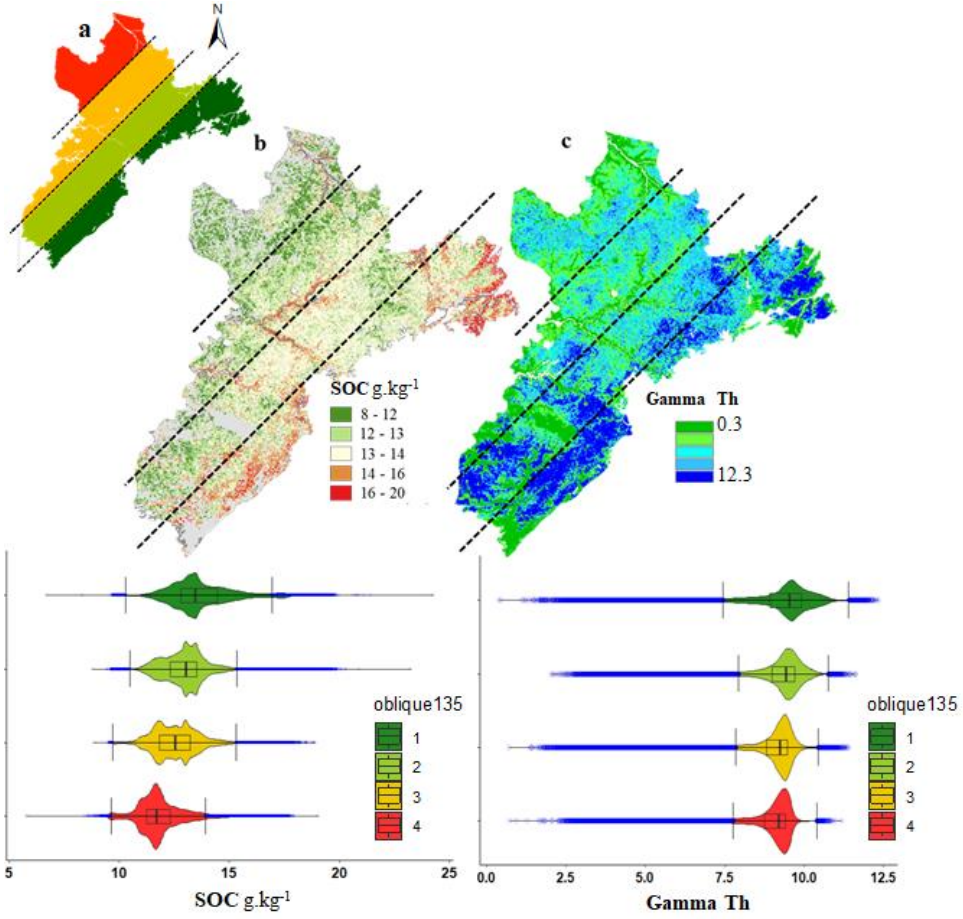


Figure 4.7. *a) Oblique geographic coordinates at angle of 135° (oblique135) divided into four classes; b) SOC content of the predicted map based on the oblique135 classes; and c) Thorium content of Gamma airborne image based on oblique135 classes.*

4.5 DISCUSSION

Over an intensively managed agricultural region such as Beauce, SOC content of the top horizon is rather low with a narrow range of values, which hampers a good prediction using spectral optical models only (Vaudour et al., 2019b, 2022), and indeed the use of a more diversified and relevant set of covariates enabled to reach the best performing model. This study demonstrates that it is worthwhile to include temporal mosaics of bare soils into such DSM approach, especially in conjunction with airborne gamma-ray and morphometric data.

4.5.1 The more relevant set of covariates in SOC modelling

4.5.1.1 The importance of both satellite-derived and morphometric covariates in SOC modelling

Previous studies have reported that the importance of S2 bands for predicting SOC can vary depending on the study area and soil conditions during the date of image acquisition (Castaldi et al., 2019b;2023; Urbina-Salazar et al., 2021; Vaudour et al., 2019b). Similar results about the importance of the S2 bands of a temporal mosaic of bare soil were found by Žížala et al. (2022), but these authors did not include spectral indices as covariates.

Behrens et al. (2019) stated that spatial context influences the content of soil attributes when considering long-range processes (Teleconnected systems). These authors suggested that including covariates derived from multi-scale neighboring information can enable the identification of spatial trends in the data using machine learning models to improve their accuracy. In our case, the importance of slope and MrVBF might be related to a long-range effect from lower SOC values at high elevation plateaus, medium values at mid elevation locations and to a higher increase at steep slopes surrounding valleys mainly near rivers. Previous studies have shown promising results on the performance of models including covariates of neighboring

information in DSM (Behrens et al., 2010; Chen et al., 2021). Recently, in a national-scale SOC mapping, Žížala et al. (2022) used a mixed approach of octave formation (Gaussian pyramids) for relief covariates and reported multi-scale derivatives of elevation among the most important covariates. Another method to fill the gap with respect to the ability to identify spatial trends over large regions that enable the estimation of soil attributes using Random Forest is the calculation of OGC that are used as covariates (Møller et al., 2020). The use of OGC (specifically, "oblique135") in Beauce confirms the results obtained by Møller et al. (2020) who reported in an agricultural field in Denmark that including OGC contributes to the prediction of organic matter due to the ability of these to identify from terrain features to management practices in a geographic space.

4.5.1.2 The major importance of airborne gamma-ray covariates in SOC modelling

Interestingly, our results emphasize the importance that airborne gamma-ray data played in SOC modelling. Our results confirm an indirect relationship between gamma-ray emissions and SOC. This confirms the results of Dierke and Werban (2013), who stated in a field site that it is possible to indirectly estimate organic carbon through thorium concentrations. In a nearby region of France, Loiseau et al. (2020) demonstrated that thorium was among the most important co-variates to predict clay content. Therefore, in soils that have been cultivated for many decades, it cannot be excluded that a large proportion of the remaining SOC in topsoil is controlled by stabilization by clay (Arrouays et al., 2006; Bruun et al., 2010; Chen et al., 2019b). Indeed, we observed a positive relationship between clay and SOC content in the Beauce French topsoils used in this study for plots for which both parameters were measured (Figure S2, supplementary material).

The results obtained with Uranium are more difficult to explain, since

the behaviour of Uranium in soils is quite complex: Uranium is fairly mobile and may significantly vary according to soil physico-chemical conditions, in particular redox state and saturation of soils plays an important role on the adsorption of Uranium on clays and organic matter. It is however difficult to relate U and SM contents since dedicated acquisition methodology and processing is applied to reduce as much as possible the effect of soil moisture on gamma-ray data (IAEA, 2003). Lastly, despite specific corrections taken in the preprocessing of gamma-ray data (Grasty, 1997) the presence of Radon gaz in the disintegration chain of Uranium also complexifies the reading of this variable. The relationship between SOC and Uranium content in soils is, however, well evidenced in our study, and further specific analysis would be needed to clearly explain this relation.

4.5.2 SOC variability and predicted spatial uncertainty

The validation by the 12 points grid should be taken with caution because it has been demonstrated that using a very small number of independent sites may lead to erroneous conclusions (Lagacherie et al., 2019). Large 90% prediction intervals are rather common for DSM SOC prediction at such resolution and over a rather wide area (Lemercier et al., 2022). It is quite likely that the uncertainty could be traced to the underestimation of extreme values mentioned above. It is very difficult in our study to establish one main reason that impacts the confidence level of prediction at sites with higher uncertainty. Still, management practices in agricultural plots, topography, soil physical conditions and historical soil-forming environmental processes in Beauce may influence SOC content and its detection via satellite (Dodin et al., 2021; Møller et al., 2020; Urbina-Salazar et al., 2021; Vaudour et al., 2019a). The SOC gradient that we observed along the study area might be influenced in some way by the direction of the winds that deposited the Loess. In other words, considering OGC it was possible to detect a spatial trend related to the SOC content (Møller et al., 2020).

4.5.3 Perspectives for SOC mapping over large agricultural regions

The availability of satellite imagery at no cost (e.g. Sentinel and Landsat) has been a turning point that has driven the use of remote sensing derived techniques and products in soil science for bare soil retrieval, attribute mapping and quantification (e.g. SCMaP, Rogge et al.(2018); GEOS3, Demattê et al. (2018); HISET, Heiden et al. (2022); R90, Castaldi et al. (2023) ; moisture maps, El Hajj et al. (2017)). The development of these products enables to obtain information about the soil surface (e.g., soil moisture maps) and to extend the percentage of bare soil. This last point is, however, questionable as cover cropping is more and more recommended to increase SOC inputs in soils and favor SOC sequestration. This is one paradox of science for SOC remote sensing: bare soils favor a better detection of SOC by remote sensing, but practices recommended to increase SOC include having soils continuously covered by vegetation. Our results obtained by evaluating S2Bsoil in two different periods of the year demonstrate that a given period, here the "late winter-spring" for Beauce, significantly favours the performance of topsoil prediction by DSM at regional scale while including different covariates in the modelling such as airborne gamma-ray data, DEM derivatives and coordinates. The uncertainty, however, remained quite high, which question the feasibility of transitioning from mapping to monitoring over short-term time ranges. An important finding in this study was to observe a spectral trend of neighboring pixel information in the entire region related to natural soil-forming processes. However, although SOC detection using spectral information from satellite images has grown exponentially in the last 5 years mainly in farms and small regions (areas with a median of 118 km^2) (Vaudour et al., 2022), there are still few studies for large agricultural regions that describe such trends in cultivated areas.

Previous studies in small regions reported that soil surface condition (e.g., SM, soil roughness, plant residues) and hence the date of image

acquisition should be considered as one of the main factors impacting SOC retrieval via satellite (Castaldi et al., 2019a;2023; Vaudour et al., 2019a). Therefore, increasing the number of soil surface condition variables for a single date or time series to be used as input covariates in models should be further explored in large agricultural regions. Temporal images of soil moisture are becoming available and are constantly being improved (El Hajj et al., 2017; Bazzi et al., 2019) while soil roughness products are still very limited and barely tested in small regions (Baghdadi et al., 2018). Information about land use history could also be interesting as well, but such information is much more tedious to gather at the scale of a region as large as Beauce. This information could have been crucial if some drastic changes in land use had happen since some decades (e.g., deforestation), but the Beauce region has been cultivated for a very long time. It is important to note that a large number of covariates providing information on soil surface condition might not necessarily improve modelling performance, but may be useful information to identify factors influencing SOC detection. Our results suggest that some additional co-variates could be useful such as for instance clay content maps. However, we should keep in mind that these maps are also not error free and that adding them as co-variates may induce some more error propagation. However, the continuous accumulation of soil data through French and EU ongoing programs will progressively help to get more and more precise maps and data for monitoring.

Regional models often require more in-depth study than on-farm or small-region models due to the short time step of SOC variability influenced by climate, landforms, parent material, management practices and crop rotation (Urbina-Salazar et al., 2021; Vaudour et al., 2021). Obviously, we also missed some information about long-term soil management and carbon inputs to soil in the past such as the application of EOM as amendment to increase the productivity of agricultural soils (Moinard et al., 2021). Green waste compost and

manure application changes SOC content as well as the spectral information detected by the satellite. These increasingly common practices are driving studies that aim to develop spectral indices in agricultural plots capable of detecting whether or not there was any application of any type of soil amendment (Dodin et al., 2021; Gomez et al., 2021). Therefore, further studies could include this information as a tool for increasing performance in SOC prediction in wide regions where these types of practices may be present. Getting information on agricultural practices and on crop yields and SOC inputs thanks to remote sensing is for sure a key input for SOC modelling, mapping and forecasting.

4.6 CONCLUSION

This study was conducted over a large area of intensive agriculture to evaluate the Sentinel-2 satellite's ability to predict SOC. Our approach was based on bare soil areas from temporal mosaics (S2Bsoil) over a 6-year period. Although the maximum bare soil coverage (81.8%) was reached using a series of 57 single dates, the performance for predicting SOC was not the best. The use of S2Bsoil by periods considering management practices and activities related to crop rotation (e.g., soil seedbed condition at the end of winter and potentially high presence of straw after harvest in late summer) slightly showed a difference in the accuracy of the models. Therefore, it is worth taking into account the soil management practices in a given region prior to building temporal bare soil mosaics over long time series to map SOC. We highlight that the use of SMPs is a valuable tool that helps to choose the best dates to create temporal bare soil mosaics, however, the SMPs uncertainty associated with the days of difference between the study dates and the products may induce biases leading to an increase in SOC prediction uncertainty.

The relative importance of environmental covariates indicated the weight of S2 bands jointly to airborne gamma ray imagery and

morphometric covariates in the SOC modelling of top horizon. Topography neighbor distances, OGC, and gamma thorium were also key covariates in this study for their contribution not only to SOC mapping but also for providing information related to the impact of long-range effects on SOC content in such a specific region.

The produced SOC map makes sense from a topographical and historical perspective related to past natural processes and the intense agricultural activity present for decades in the area. This map can be used in soil management decision making from a global vision in Beauce. However, for farmers, i.e. at a local farm scale, the uncertainty map shows that more detailed studies would be required. Further studies might use this information in better locating new soil sampling efforts to help reducing the error. Surveys of farmers practices would be a useful tool to identify management practices such as organic amendments that could distort SOC prediction at local scale. Conversely, methods to produce variables from neighboring information might be useful in regional scale machine learning models in order to better detect broad spatial trends related to SOC variability.

CHAPTER 5
DISCUSSION AND GENERAL CONCLUSIONS

5 DISCUSSION AND GENERAL CONCLUSIONS

Sustainability, climate change mitigation and food security are three challenges that humanity is currently facing to ensure the well-being and the right to a dignified life for future generations. Soil organic carbon (SOC) is an essential component that may help mitigate these challenges due to the imbalance and negative effects caused mainly by anthropogenic actions that endanger the quality of our agricultural soils. In this context, the development and technological progress materialized in satellite sensors has positioned them as a promising alternative for SOC monitoring at different spatial and temporal scales. ***Chapter 1*** introduced an overview of the development of satellite series such as Landsat and SPOT in terms of spatial and spectral resolution and SOC detection applications around the world. This is an indicator of the success of these missions in the observation and monitoring of the Earth's surface in soil science studies. However, the number of studies aiming to map SOC using purely spectral methods derived from satellite imagery in countries such as France is just increasing in relation to the total number of works analyzed at the national level in this thesis. Globally Vaudour et al. (2022) reported sixty-two satellite-derived SOC studies since the 1990s and most of the studies have been performed on small regions covering a few hundred km^2, few studies have explored the contribution of satellite image series over large and very large regions such as Beauce and Pyrenean piedmont. The general objective of this thesis was mainly based on determining the potential of the Sentinel-2 (S2) satellite to estimate SOC and to better understand its variability in relation to the morphometric, pedological, natural and historical context of each study area. In ***Chapter 2*** we dealt with the low number of soil samples available due to the limited area of bare soil available on the dates of the satellite images. However, this limitation was fundamental to structure the sequence of approaches and methods that were

subsequently applied (Table 2.3). The results at Pleine-Fougères not only suggested the need to develop a method of extending the area of bare topsoil in order to cover as many sites as possible with soil sampling, but also highlighted the interest of implementing the approach by single date with satellite data in France over different regions where this approach had not been addressed. In ***Chapter 3***, we addressed a purely spectral and mixed method using soil moisture products (SMPs) as covariate. All models showed variation in prediction performances mainly related to the date of acquisition of S2 images, soil texture, soil sampling period, soil moisture and landforms in Pyrenean piedmont. Image date and soil moisture were the most evident factors in model accuracy. However, our results suggested that single date imaging approaches should be further explored by considering not only components related to the satellite image characteristics but also to the study area, the samples used and the surface soil condition. We also showed that SOC spatial distribution could be related to some topographic and past land-use situations. In ***Chapter 4*** we produced S2 temporal mosaics of bare soil (S2Bsoil) in the Beauce region to evaluate possible effects on model performances using S2Bsoil over a full time series (2016-2021) and by periods. SMPs and several terrain-derived variables were used as covariates in the modelling. The results suggested that the selection of images for building bare soil temporal mosaics using S2 or other different sensors should be taken into account due to the variability that SOC may present in a given area as well as disturbing factors such as those discussed in ***Chapter 3*** that may affect the SOC prediction. The importance of the variables used in this chapter and the observed effects of some disturbing factors in Pyrenean piedmont and Beauce leaves remaining questions on how and what strategies to use to improve SOC prediction through the different satellite-based methods and approaches.

5.1 GENERAL PERSPECTIVES

Techniques for mapping SOC on bare soil with single-date multispectral images are increasingly being replaced by temporal composite images. Spatially this technique offers the advantage of obtaining a wider coverage of bare soil pixels which has enabled SOC mapping up to national scales (Wang et al., 2020; Safanelli et al., 2021; Žížala et al., 2022). The results of ***Chapter 2***, ***3*** and ***4*** suggested that despite the limitations that a single date approach may have (e.g., small area of bare soil available, low sample set for model calibration and validation), performing pre-analysis on single dates can contribute to understanding the causes affecting model performance and therefore selecting specific covariates and better prediction dates to produce bare soil temporal images. We will now focus on the questions mentioned in ***Chapter 1*** by discussing some factors that should be taken into account for further research: ***which is the optimal period to acquire satellite images for SOC estimation?*** The best date to find bare soil may vary from one place to another in our case for instance for Northern France, the best season to observe bare soil and obtain the best prediction performances was spring confirming previous field observations and earlier studies in the Ile-de-France region (Vaudour et al., 2019a; 2019b). However elsewhere the best season is autumn in the western Guanzhong Plain, China (Wang et al., 2021) or even in dry periods in winter in the southeast of São Paulo State, Brazil (Sayão et al., 2018). Vaudour et al. (2019a) reported a positive linear relationship between sun elevation angle and the performance of SOC prediction models using images mainly acquired in spring. Recently, Castaldi et al. (2023) reported in 10 different local scale areas around the northern hemisphere different months for each area as the most appropriate for bare soil retrieval using a two-year S2 image series. These authors, for instance, found that for two areas in Turkey depending on the approach used to elaborate bare soil composite images the best months were August, September and

October. In the United States in two areas evaluated the most suitable months were April, May and November; in three areas in Italy they were June, July and October and finally in Lithuania in three areas the months used in the bare soil composite images to detect SOC were March, April, June, September and October. In this thesis we evaluated spring images in ***chapters 2*** and ***3*** as well as broad periods in ***chapter 4*** for a large area. Therefore, it would be interesting that more studies continue to evaluate for different periods or seasons time series considering different agroecosystems and management practices. This would help to clarify some questions such as: **i)** what would be the main factors that would determine the best period per region? **ii)** would they have any relation with practices such as crop rotation or the application of organic amendments? **iii)** would image characteristics such as sun elevation angle show any tendency in different regions? These are some outstanding questions in different regional contexts that need to be addressed in the near future. The second question is ***can differences between the SOC sampling periods and the date of satellite imagery influence the model performances?*** The soil sampling period and its relationship to the performance of prediction models is not usually investigated in detail mainly in studies where legacy SOC values from different years are used. The results of ***Chapter 3*** suggested that although variations in SOC may be minimal in a given period, caution must be taken in sites where there might be significant variations over time for reasons such as; changes in land cover and land use, continuous planting of a particular crop or other management practices such as the application of organic amendments (e.g., animal manure, organic waste and even biochar). The application of organic amendments to agricultural soils has increased in recent years and the rate of application to croplands is expected to increase further not only to improve soil quality but also as a possible alternative to mitigate climate change. However, from a SOC remote sensing perspective, to what extent would the application of such amendments affect SOC prediction using satellite sensors?;

would SOC legacy data in eventual applications no longer be useful?, would new SOC resampling be necessary?

Strategies to improve SOC prediction using bare soil temporal images are currently needed. However, still very few studies have focused on that and most of them have mainly aimed at using threshold indices such as NDVI and NBR2 to filter pixels with bare soil information. In this context, ***What strategy should we use to retrieve bare soil over a satellite time series in order to improve SOC prediction in a given area?*** In Beauce, we considered the NDVI and NBR2 indices as well as soil moisture using the mean values of the soil moisture products (SMPs). Vaudour et al. (2021) used different combinations of index thresholds by date and by pixel on the Versailles plain. Other methodologies have started to be explored in order to improve the quality of bare soil composite images. For instance, Heiden et al. (2022) developed a methodology (HISET, Histogram SEparation Threshold Method) based on the SCMAP mapping processor, in which they tested quality scores and spectral index thresholds to validate bare soil pixels in a very large region covering the German federal state of Bavaria and some parts of the Czech Republic and Austria. However, this methodology has not yet been tested for determining SOC. At local scale, Castaldi et al. (2023) determined SOC by assessing the potential of four strategies based on the use of S2Bsoil. These authors found that the best performances corresponded to either the median value of reflectance for each band ("Median" approach) or in dry soil conditions excluding extreme values of reflectance ("R90" approach). A general question regarding these methodologies is to what extent can they be applied in other locations in terms of scale, climatic conditions and land management practices? It is expected that other methodologies based on new indexes or index thresholds will be tested in other regions with different environmental conditions, land cover and management practices that may allow establishing some standards for the use of bare soil composite images according to each

regional and even national context.

In a bare soil composite imaging approach using mixed methods, ***what would be the most relevant covariates to include in the models?*** we showed, especially in ***chapter 4***, that the SOC predictions using only bare soil remote sensing data can be noticeably improved at a regional scale, by adding other relevant co-variates in a more global DSM framework. We also showed that adding some relevant co-variates (eg., airborne gamma-ray images and oblique geographic coordinates) and looking at their relationships and spatial structure may help, not only to improve the prediction performance, but also to better understand the process at the origin of SOC spatial variability. This is indeed, one of the big challenges of DSM, i.e. not only to improve predictions but to better understand the process governing soil variability (Arrouays et al., 2020b, 2020c; Wadoux et al., 2020; Wadoux and Molnar, 2022).

The poor number or non-uniform distribution of available soil samples is a factor that might influence the performance of SOC detection models in terms of spatial and spectral representativeness. An alternative to address this is the "Bottom up" method, i.e. the use of soil spectral libraries (SSL) with multispectral or hyperspectral images (Castaldi et al., 2018; Aichi et al., 2021). SSL collect information on a large number of sampling sites that can cover unsampled locations for model refinement. In other words, these data may be used to calibrate spectral models and then predict more accurate SOC values in areas of bare soil from satellite imagery where no data are available or sampling was not possible. Although we do not use this method in this Thesis, it would be interesting to test it with Sentinel-2 data and SSL at national (Gogé et al., 2014; Clairotte et al., 2016) and regional (Thoisy et al., 2022) scales in France.

Finally, it is important to mention that one of the best ways to address the current challenges and inquiries in SOC remote sensing are field

observations. In ***Chapter 4*** we suggested the use of additional information such as the application of organic soil amendments to improve the performance of SOC determination. Indices that can be used to detect such practices could be tested in regions where satellite SOC prediction may be affected by such practices. More information about the soil surface condition is required to improve not only bare soil retrieval but also the accuracy of SOC prediction models. Recently Dvorakova et al. (2023) performed a bare soil composite image to produce SOC prediction maps. These authors in the field carried out soil sampling and took photographs to identify the different soil surface conditions coinciding with the same days that the S2 images were acquired. They finally determined that applying an NBR2 threshold <0.05 and normalizing the S2 spectra excluded residues and wet soils, and corrected the effect of soil crusts on the reflectance spectra. In other words, soil sampling and field observations are necessary to develop Bare soil temporal images in order to obtain more realistic SOC prediction performance in each study area.

5.2 FINAL CONSIDERATIONS

This thesis after a regional scope from a small region (Pleine-Fougères), to a large (Beauce) and very large region (Pyrenean piedmont), we show that SOC mapping via the sentinel-2 satellite is a relevant tool in digital soil mapping. However, the predictive potential can vary significantly from one given area to another and is linked both to the image characteristics, soil surface conditions and the relationship between SOC spatial variability and long-range terrain effects due to past processes. Further studies could consider the main findings of this thesis, addressing factors and strategies such as the ones mentioned in the previous section, considering the use of the newly launched satellite sensors (the Italian PRISMA (Pignatti et al., 2015; Angelopoulou et al., 2023), the German EnMap (Guanter et al., 2015; Chabrillat et al., 2022), the Chinese GaoFen-5 (Ye et al., 2020) and forthcoming missions : the European Copernicus Hyperspectral

Imaging Mission for the Environment planned to be launched in 2028-2030 (Rast et al., 2021), the US Surface Biology and Geology (Lee et al., 2015), the French BIODIVERSITY (Briottet et al., 2022). Furthermore, the increasing demand for satellites mainly in the security and telecommunications sector is currently driving the development of Microsatellites and high altitude pseudosatellites (HAPS) with single digit optical resolutions and a higher real-time imaging capacity than high altitude satellites. Both technologies might be complementary to the wide swath of satellite coverage and the narrow swath of unmanned aerial vehicles, making them potentially interesting as a tool to support digital soil mapping and precision agriculture at different scales. In that sense, for instance, the analysis of a large number of images and the increasing expansion of SSL confirm that deep learning algorithms capable of processing images and large-scale data should be considered and tested in the prediction of SOC content. Deep learning has been implemented to map clay content (Tziolas et al., 2020), however, studies addressing the use of deep learning to detect SOC from satellite series are still scarce (Vaudour et al., 2022). Therefore future studies may address SOC prediction in relation to dense SSL, satellite, microsatellite and HAPS images using deep learning. Nevertheless the free availability of microsatellite and HAPS images is questionable at least in the near future.

Finally, our study highlights that digital SOC mapping should be further addressed to regional contexts. At present several global projects and initiatives (e.g. GlobalSoilMap and SoilGrids) have mainly focused on SOC mapping at national and global scales. Information on the national or global SOC content is important, however, as we showed in this thesis the variability of SOC can change according to the study region and the prediction performances via satellite imagery can vary for several reasons. SOC variability in many regions around the world is currently unknown and detailed maps at resolutions suitable for end users such as farmers are in extreme demand. Projects

such as STEROPES that address specific regions not only in Europe but worldwide are fundamental to **i)** identify factors influencing model performance, **ii)** establish mapping methodologies adapted to different cropping systems and **iii)** support bottom-up national mapping. Future research projects should be aimed at the development of key products at local or regional scales for farmers. Production management zone maps (de Almeida et al., 2023), soil fertility index maps (Viscarra Rossel et al., 2010) and even carbon storage potential maps (Chen et al., 2019a) elaborated from soil attributes such as SOC or terrain and satellite data are some examples to address different applications related to risks and suitability of agricultural land for various crops; a gap between digital soil mapping and digital soil assessment that needs to be addressed (Carré et al., 2007; Kidd et al., 2015).

REFERENCES

Adhikari, K., Hartemink, A.E., 2016. Linking soils to ecosystem services - A global review. Geoderma 262, 101–111. https://doi.org/10.1016/j.geoderma.2015.08.009

Agbu, P.A.; Fehrenbacher, D.J.; Jansen, I.J.,1990. Soil Property Relationships with SPOT Satellite Digital Data in East Central Illinois. Soil Science Society of America Journal 54, 807–812, doi:10.2136/sssaj1990.03615995005400030031x.

AGRESTE, 2020. Spécialisation territoriale de la production agricole en 2020 (OTEX en 17 postes) – Recensement agricole 2020. https://stats.agriculture.gouv.fr/cartostat/#c=home (Accessed online on February 14, 2023)

Aichi, H., Fouad, Y., Lili Chabaane, Z., Sanaa, M., Walter, C., 2021. Soil total carbon mapping, in Djerid Arid area, using ASTER multispectral remote sensing data combined with laboratory spectral proximal sensing data. Arab J Geosci 14, 405. https://doi.org/10.1007/s12517-021-06698-z

Angelopoulou, T., Chabrillat, S., Pignatti, S., Milewski, R., Karyotis, K., Brell, M., Ruhtz, T., Bochtis, D., Zalidis, G., 2023. Evaluation of Airborne HySpex and Spaceborne PRISMA Hyperspectral Remote Sensing Data for Soil Organic Matter and Carbonates Estimation. Remote Sensing 15, 1106. https://doi.org/10.3390/rs15041106

Aubert, M.; Baghdadi, N.; Zribi, M.; Douaoui, A.; Loumagne, C.; Baup, F.; El Hajj, M.; Garrigues, S., 2011. Analysis of TerraSAR-X data sensitivity to bare soil moisture, roughness, composition and soil crust. Remote Sens. Environ. 115, 1801–1810. https://doi.org/10.1016/j.rse.2011.02.021.

Arrouays, D.; Baize, D.; Hardy, M., 1992. Les sols de "touyas" issus d'alluvions anciennes des gaves pyrénéens: Veracrisols. In-tégration au Référentiel Pédologique. Sci. Du Sol. 30, 227–247.

Arrouays, D.; Pelissier, P., 1994. Changes in carbon storage in temperate humic loamy soils after forest clearing and continuous corn cropping in France. Plant Soil, 160, 215–223. https://doi.org/10.1007/bf00010147.

Arrouays, D.; Vion, I.; Kicin, J.L., 1995a. Spatial analysis and modeling of topsoil carbon storage in forest humic loamy soils of France. Soil Sci. 159, 191–198.

Arrouays, D.; Balesdent, J.; Mariotti, A. ; Girardin, C., 1995b. Modelling organic carbon turnover in cleared temperate forest soils converted to maize cropping by using 13C natural abundance measurements. Plant Soil, 173, 191–196. https://doi.org/10.1007/bf00011455.

Arrouays, D. King, C. Vion, I. Le Bissonnais, Y., 1996. Detection of soil crusting risks related to low soil organic carbon contents by using discriminant analysis on thematic mapper data. Geocarto International, 11 (4):11-16.

Arrouays, D. Daroussin, J. Kicin, J.L. Hassika, P., 1998. Improving topsoil carbon storage prediction using a digital elevation model in temperate forest soils of France. Soil Science.

163:103-108.

Arrouays, D.; Saby, N.; Walter, C.; Lemercier, B.; Schvartz, C., 2006. Relationships between particle-size distribution and organic carbon in French arable topsoils. Soil Use Manag. 22, 48–51. https://doi.org/10.1111/j.1475-2743.2006.00020.x.

Arrouays, D., McKenzie, N.., Hempel, J., Richer-de-Forges, A.C., McBratney, A., 2014a. GlobalSoilMap: Basis of the Global Spatial Soil Information System. CRC Press Taylor & Francis Group.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L., 2014b. GlobalSoilMap. Toward a Fine-Resolution Global Grid of Soil Properties, Advances in Agronomy. https://doi.org/10.1016/B978-0-12-800137-0.00003-0

Arrouays, D., Horn, R., 2019. Soil carbon - 4 per Mille - an introduction. Soil Tillage Res. 188, 1—2. https://doi.org/10.1016/j.geoderma.2017.01.002

Arrouays, D., Richer-de-Forges, A.C., Héliès, F., Mulder, V.L., Saby, N.P.A., Chen, S., Martin, M.P., Román Dobarco, M., Follain, S., Jolivet, C., Laroche, B., Loiseau, T., Cousin, I., Lacoste, M., Ranjard, L., Toutain, B., Le Bas, C., Eglin, T., Bardy, M., Antoni, V., Meersmans, J., Ratié, C., Bispo, A., 2020a. Impacts of national scale digital soil mapping programs in France. Geoderma Reg. 23. https://doi.org/10.1016/j.geodrs.2020.e00337

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de- Forges, A., Morgan, Cristine L.S., Roudier, P., Poggio, L., Mulder, VL., 2020b. Impressions of digital soil maps: The good, the not so good, and making them ever better. Geoderma Regional 20, e00255. https://doi.org/10.1016/j.geodrs.2020.e00255.

Arrouays, D., Poggio, L., Salazar, O., Mulder, VL., 2020c. Digital soil mapping and GlobalSoilMap. Main advances and ways forward. Geoderma Regional 21, e00265. https://doi.org/10.1016/j.geodrs.2020.e00265

Baetens, L.; Desjardins, C.; Hagolle, O., 2019. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. Remote Sens. 11, 433. https://doi.org/10.3390/rs11040433.

Baghdadi, N.; Holah, N.; Zribi, M., 2006. Calibration of the Integral Equation Model for SAR data in C-band and HH and VV polarizations. Int. J. Remote Sens. 27, 805–816. https://doi.org/10.1080/01431160500212278.

Baghdadi, N.; El Hajj, M.; Zribi, M.; Bousbih, S., 2017. Calibration of the Water Cloud Model at C-Band for Winter Crop Fields and Grasslands. Remote Sens. 9, 969. https://doi.org/10.3390/rs9090969.

Baghdadi, N., Hajj, M. El, Choker, M., Zribi, M., Bazzi, H., Vaudour, E., Gilliot, J.M., Ebengo, D.M., 2018. Potential of Sentinel-1 images for estimating the soil roughness over bare agricultural soils. Water (Switzerland) 10, 1–14. https://doi.org/10.3390/w10020131

Baret, F., Jacquemoud, S., Hanocq, J.F., 1993. About the soil line concept in remote sensing. Adv. Sp. Res. 13, 281–284. https://doi.org/10.1016/0273-1177(93)90560-X

Barthès, B.G., Chotte, J., 2021. Infrared spectroscopy approaches support soil organic carbon estimations to evaluate land degradation. Land Degradation & Development 32, 310–322. https://doi.org/10.1002/ldr.3718

Bazzi, H.; Baghdadi, N.; El Hajj, M.; Zribi, M.; Belhouchette, H., 2019. A Comparison of Two Soil Moisture Products S2MP and Copernicus-SSM Over Southern France. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 3366–3375. https://doi.org/10.1109/jstars.2019.2927430.

Behrens, T., MacMillan, R.A., Viscarra Rossel, R.A., Schmidt, K., Lee, J., 2019. Teleconnections in spatial modelling. Geoderma 354, 113854. https://doi.org/10.1016/j.geoderma.2019.07.012

Behrens, T., Zhu, A.X., Schmidt, K., Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma 155, 175–185. https://doi.org/10.1016/j.geoderma.2009.07.010

Bellón, B., Bégué, A., Seen, D. Lo, de Almeida, C.A., Simões, M., 2017. A remote sensing approach for regional-scale mapping of agricultural land-use systems based on NDVI time series. Remote Sens. 9, 1–17. https://doi.org/10.3390/rs9060600

Ben-Dor, E., Goldlshleger, N., Benyamini, Y., Agassi, M., Blumberg, D.G., 2003. The Spectral Reflectance Properties of Soil Structural Crusts in the 1.2- to 2.5-µm Spectral Region. Soil Sci. Soc. Am. J. 67, 289–299. https://doi.org/10.2136/sssaj2003.2890

Ben-Dor, E., Inbar, Y., Chen, Y., 1997. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400-2500 nm) during a controlled decomposition process. Remote Sens. Environ. 61, 1–15. https://doi.org/10.1016/S0034-4257(96)00120-4

Ben-Dor, E., Banin, A., 1995. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. Soil Sci. Soc. Am. J. 59, 364–372. https://doi.org/10.2136/sssaj1995.03615995005900020014x

Berthier, L., Pitres, J.C., Vaudour, E., 2008. Prédiction spatiale des teneurs en carbone organique des sols par spectroscopie vi-sible-proche infrarouge et télédétection satellitale SPOT. Exemple au niveau d'un périmètre d'alimentation en eau po-table en Beauce. Etude et Gestion des Sols 15, 4, 161-172.

Bertran, P., Liard, M., Sitzia, L., Tissoux, H., 2016. A map of Pleistocene aeolian deposits in Western Europe, with special emphasis on France. J. Quat. Sci. 31, 844–856. https://doi.org/10.1002/jqs.2909

Besnard, E.; Chenu, C.; Balesdent, J.; Puget, P.; Arrouays, D., 1996. Fate of particulate organic matter in soil aggregates during cultivation. Eur. J. Soil Sci. 47, 495–503. https://doi.org/10.1111/j.1365-2389.1996.tb01849.x.

Biney, J.K.M., Saberioon, M., Borůvka, L., Houška, J. , Vašát, R. , Chapman, P., Coblinski, J.A., Klement, A., 2021. Exploring the Suitability of UAS-Based Multispectral Images for Estimating Soil Organic Carbon: Comparison with Proximal Soil Sensing and Spaceborne Imagery. Remote

Sensing, 13, 308. https://doi.org/10.3390/rs13020308

Blum, W.E.H., 2005. Functions of soil for society and the environment. Rev. Environ. Sci. Biotechnol. 4, 75–79. https://doi.org/10.1007/s11157-005-2236-x

Boettinger, J.L.; Ramsey, R.D.; Bodily, J.M.; Cole, N.J.; Kienast-Brown, S.; Nield, S.J.; Saunders, A.M.; Stum, A.K., 2008. Landsat Spectral Data for Digital Soil Mapping. In Digital Soil Mapping with Limited Data; Hartemink, A.E., McBratney, A., Mendonça-Santos, M.d.L., Eds.; Springer: Dordrecht, The Netherlands; pp. 193–202, ISBN 978-1-4020-8591-8.

Boiffin, J. La Dégradation Structurale des Couches Superficielles des Sols Sous l'action des Pluies. PhD Thesis, Sciences du Vi-vant [q-bio]. Institut National Agronomique Paris Grignon, Paris, France, 1984. (In French)

Borderie, Q., Chamaux, G., Roussaffa, H., Douard, M., Fencke, É., Rodot, M., Perrichon, P., Selles, H., 2017. La couverture lœssique d'Eure-et-Loir (France) :Potentiel pédo-sedimentaire et organisation spatiale. Quaternaire 28, 389–400. https://doi.org/10.4000/quaternaire.8331

Bouarfa, S., Brunel, L., Granier, J., Mailhol, J.C., Morardet, S., Ruelle, P., 2011. Évaluation en partenariat des stratégies d'irrigation en cas de restriction des prélèvements dans la nappe de Beauce (France). Cah. Agric. 20, 124–129. https://doi.org/10.1684/agr.2010.0461

Bouma, J., Montanarella, L., Evanylo, G., 2019. The challenge for the soil science community to contribute to the implementation of the UN Sustainable Development Goals. Soil Use Manag. 35, 538–546. https://doi.org/10.1111/sum.12518

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Briottet, X., Bajjouk, T., Chami, M., Delacourt, C., Féret, J.B., Jacquemoud, S., Minghelli, A., Sheeren, D., Weber, C., Fabre, S., Adeline, K., Vaudour, E., Luque, S., Deville, Y., Soudani, K., Verpoorter, C., 2022. BIODIVERSITY – A new space mission to monitor Earth ecosystems at fine scale. RFPT 224, 33–58. https://doi.org/10.52638/rfpt.2022.568

Bruun, T.B., Elberling, B., Christensen, B.T., 2010. Lability of soil organic carbon in tropical soils with different clay minerals. Soil Biol. Biochem. 42, 888–895. https://doi.org/10.1016/j.soilbio.2010.01.009

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: beyond DSM. Geoderma 142 (1–2), 69–79.

Castaldi, F., Chabrillat, S., Jones, A., Vreys, K., Bomans, B., van Wesemael, B., 2018. Soil Organic Carbon Estimation in Croplands by Hyperspectral Remote APEX Data Using the LUCAS Topsoil Database. Remote Sensing 10, 153. https://doi.org/10.3390/rs10020153

Castaldi, F.; Chabrillat, S.; Don, A.; Van Wesemael, B., 2019a. Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects. Remote Sens. 11, 2121. https://doi.org/10.3390/rs11182121.

Castaldi, F.; Hueni, A.; Chabrillat, S.; Ward, K.; Buttafuoco, G.; Bomans, B.; Vreys, K.; Brell, M.; van Wesemael, B., 2019b. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS J. Photogramm. Remote Sens. 147, 267–282.

https://doi.org/10.1016/j.isprsjprs.2018.11.026.

Castaldi, F., Koparan, Muhammed., Wetterlind, J., Žydelis, R., Vinci, I., Savaş, Ayşe., Kıvrak, C., Tunçay, T., Volungevičius, J., Obber, S., Ragazzi, F., Malo, D., Vaudour, E., 2023. Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands. ISPRS Journal of Photogrammetry and Remote Sensing. 2023, 199, 40-60. https://doi.org/10.1016/j.isprsjprs.2023.03.016

de Almeida, G. S., Rizzo, R., Amorim, M. T. A., dos Santos, N. V., Rosas, J. T. F., Campos, L. R., et al., 2023. Monitoring soil–plant interactions and maize yield by satellite vegetation indexes, soil electrical conductivity and management zones. Precision Agriculture, 1-21. https://doi.org/10.1007/s11119-023-09994-8

de Castro-Padilha, M.C., Vicente, L.E., Demattê, J.A.M., dos Santos Wendriner Loebmann, D.G., Vicente, A.K., Urbina-Salazar, D.F.U., Guimarães, C.C.B., 2020. Using Landsat and soil clay content to map soil organic carbon of oxisols and Ultisols near São Paulo, Brazil. Geoderma Reg. 21, e00253. https://doi.org/10.1016/j.geodrs.2020.e0025

de Jong, S.; Addink, E.; van Beek, L.; Duijsings, D., 2011. Physical characterization, spectral response and remotely sensed mapping of Mediterranean soil surface crusts. CATENA, 86, 24–35. https://doi.org/10.1016/j.catena.2011.01.018.

Chabrillat, S., Foerster, S., Steinberg, A., Segl, K., 2014. Prediction of common surface soil properties using airborne and simulated EnMAP hyperspectral images: Impact of soil algorithm and sensor characteristic. Int. Geosci. Remote Sens. Symp. 2914–2917. https://doi.org/10.1109/IGARSS.2014.6947086

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B., 2019. Imaging Spectroscopy for Soil Mapping and Monitoring. Surveys in Geophysics 40, 361–399. https://doi.org/10.1007/s10712-019-09524-0

Chabrillat, S., Segl, K., Foerster, S., Brell, M., Guanter, L., Schickling, A., Storch, T., Honold, H. P., and Fischer, S., 2022. EnMAP Pre-Launch and Start Phase: Mission Update. In IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium (pp. 5000-5003). (International Geoscience and Remote Sensing Symposium (IGARSS); Vol. 2022-July). Institute of Electrical and Electronics Engineers Inc.doi.org/10.1109/IGARSS46834.2022.9884773

Chen, S., Arrouays, D., Angers, Denis A., Chenu, C, Barre, P., Martin, M., Saby, N., Walter, C., 2019a. National estimation of soil organic carbon storage potential for arable soils: A data-driven approach coupled with carbon-landscape zones. Science of The Total Environment 666, 355–367.

Chen, S., Arrouays, D., Angers, D.A., Martin, M.P., Walter, C., 2019b. Soil carbon stocks under different land uses and the applicability of the soil carbon saturation concept. Soil Tillage Res. 188, 53–58. https://doi.org/10.1016/j.still.2018.11.001

Chen, S., Richer-de-Forges, A.C., Leatitia Mulder, V., Martelet, G., Loiseau, T., Lehmann, S., Arrouays, D., 2021. Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France. Catena 198. https://doi.org/10.1016/j.catena.2020.105062

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z.,

Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A.C., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. Geoderma 409, 115567. https://doi.org/10.1016/j.geoderma.2021.115567

Chenu, C., Angers, D. A., Barre, P., Derrien, D., Arrouays, D., Balesdent, J., 2019. Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. Soil Tillage Research 188, 41–52. https://doi.org/10.1016/j.still.2018.04.011

Cierniewski, J., and Courault, D., 1993. Bidirectional reflectance of bare soil surfaces in the visible and near-infrared range, Remote Sens. Rev. 7:321-339.

Clairotte, M., Grinand, C., Kouakoua, E., Thébault, A., Saby, N.P.A., Bernoux, M., Barthès, B.G., 2016. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. Geoderma 276, 41–52. https://doi.org/10.1016/j.geoderma.2016.04.021

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Costa, C.; Seabauer, M.; Schwarz, B.; Dittmer, K.; Wollenberg, E., 2021. Scaling Soil Organic Carbon Sequestration for Climate Change Mitigation; CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS): Wageningen, The Netherlands.

Crème, A.; Rumpel, C.; Malone, S.L.; Saby, N.P.A.; Vaudour, E.; Decau, M.-L.; Chabbi, A, 2020. Monitoring Grassland Man-agement Effects on Soil Organic Carbon—A Matter of Scale. Agronomy, 10, 2016, doi:10.3390/agronomy10122016.

Demattê, J.A.M., Campos, R.C., Alves, M.C., Fiorio, P.R., Nanni, M.R., 2004. Visible-NIR reflectance: A new approach on soil evaluation. Geoderma 121, 95–112. https://doi.org/10.1016/j.geoderma.2003.09.012

Demattê, J.A.M., Sayão, V.M., Rizzo, R., Fongaro, C.T., 2017. Soil class and attribute dynamics and their relationship with natural vegetation based on satellite remote sensing. Geoderma 302, 39–51. https://doi.org/10.1016/j.geoderma.2017.04.019

Demattê, J.A.M., Fongaro, C.T., Rizzo, R., Safanelli, J.L., 2018. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens. Environ. 212, 161–175. https://doi.org/10.1016/j.rse.2018.04.047

Demattê, J.A.M., Dotto, A.C., Bedin, L.G., Sayão, V.M., Souza, A.B. e., 2019a. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. Geoderma 337, 111–121. https://doi.org/10.1016/j.geoderma.2018.09.010

Demattê, J.A., Dotto, A.C., Paiva, A.F., Sato, M.V., Dalmolin, R.S., Maria do Socorro, B., da Silva, E.B., Nanni, M.R., ten Caten, A., Noronha, N.C., et al., 2019b. The Brazilian soil spectral library (bssl): A general view, application and challenges. Geoderma 354, https://doi.org/10.1016/j.geoderma.2019.05.043.

Dexter, A.R., Richard, G., Arrouays, D., Czyz, E.A., Jolivet, C., Duval, O., 2008. Complexed organic

matter controls soil physical properties. Geoderma 144, 620-627. https://doi.org/10.1016/j.geoderma.2008.01.022

Diek, S.; Fornallaz, F.; Schaepman, M.E.; De Jong, R., 2017. Barest Pixel Composite for Agricultural Areas Using Landsat Time Series. Remote Sens. 9, 1245. https://doi.org/10.3390/rs9121245.

Dierke, C., Werban, U., 2013. Relationships between gamma-ray data and soil properties at an agricultural test site. Geoderma 199, 90–98. https://doi.org/10.1016/j.geoderma.2012.10.017

Dilts, T.E. Topography Tools for ArcGIS 10.3 and Earlier. Available online: https://www.arcgis.com/home/item.html?id=b13b3b40fa3c43d4a23a1a09c5fe96b9 (accessed on 10 September 2021).

Dodin, M., Smith, H.D., Levavasseur, F., Hadjar, D., Houot, S., Vaudour, E., 2021. Potential of sentinel-2 satellite images for monitoring green waste compost and manure amendments in temperate cropland. Remote Sens. 13. https://doi.org/10.3390/rs13091616

Dvorakova, K.; Shi, P.; Limbourg, Q.; van Wesemael, B., 2020. Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues. Remote Sensing, 12, 1913. https://doi.org/10.3390/rs12121913

Dvorakova, K.; Heiden, U.; van Wesemael, B., 2021. Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. Remote Sensing, 13, 1791. https://doi.org/10.3390/rs13091791

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., Os, G., Wesemael, B., 2023. Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties. Geoderma 429, https://doi.org/10.1016/j.geoderma.2022.116128.

El Hajj, M.; Baghdadi, N.; Zribi, M.; Bazzi, H., 2017. Synergic Use of Sentinel-1 and Sentinel-2 Images for Operational Soil Moisture Mapping at High Spatial Resolution over Agricultural Areas. Remote Sens. 9, 1292. https://doi.org/10.3390/rs9121292.

Ellili-Bargaoui, Y., Walter, C., Michot, D., Pichelin, P., Lemercier, B., 2019. Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale. Geoderma 351, 1–8. https://doi.org/10.1016/j.geoderma.2019.03.005.

Ellili-Bargaoui, Y., Walter, C., Michot, D., Lemercier, B., 2020. Mapping soil properties at multiple depths from disaggregated legacy soil maps in the Brittany region. France. Geoderma Regional 23, https://doi.org/10.1016/j.geodrs.2020.e00342.

Escadafal, R., Girard, M.C., Courault, D., 1989. Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data. Remote Sens. Environ. 27, 37–46. https://doi.org/10.1016/0034-4257(89)90035-7.

Escadafal, R., 1989a. Caractérisation de la surface des sols arides par observations de terrain et par télédétection. Applications : exemple de la région de Tataouine (Tunisie). Thèse de doctorat, Univ. Pierre et Marie Curie, Paris 6, 316 p.

Escadafal, R., 1989b Remote Sensing of Arid Soil Surface Color with Landsat Thematic Mapper. Adv. Space Res. 9, 159–163.

FAO; ITPS. Global Soil Organic Map V1.5: Technical Report; FAO: Quebec City, QC, Canada; ITPS: Den Haag, The Netherlands , 2020; ISBN 9789251321447

Follain, S., Walter, C., Legout, A., Lemercier, B., Dutin, G., 2007. Induced effects of hedgerow networks on soil organic carbon storage within an agricultural landscape. Geoderma 142, 80–95. https://doi.org/10.1016/j.geoderma.2007.08.002

Gasmi, A.; Gomez, C.; Chehbouni, A.; Dhiba, D.; El Gharous, M., 2022. Using PRISMA hyperspectral satellite imagery and GIS approaches for soil fertility mapping (FertiMap) in northern Morocco. Remote Sens. 14, https://doi.org/10.3390/rs14164080

Geladi, P.; Kowalski, B.R., 1986. Partial least-squares regression: A tutorial. Anal. Chim. Acta, 185, 1–17. https://doi.org/10.1016/0003-2670(86)80028-9.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sens. Environ. 218, 89–103. https://doi.org/10.1016/j.rse.2018.09.015

Girard, M.-C., 1978. Emploi de la télédétection pour l'étude de l'humidité des sols. La Houille Blanche 533–539. https://doi.org/10.1051/lhb/1978044

Girard, M.C., Girard, C.M., 1989. Télédétection appliquée zones tempérées et intertropicales. Masson, Paris, 260 p.

Girard M.C., Girard, CM., 2003. Processing of remote sensing data., Oxford & IBH Publishing Co., New Delhi, 486 p.

GIS Sol. 2011. L'état des sols de France. Groupement d'intérêt scientifique sur les sols, 188 p. https://www.gissol.fr/rapports/Rapport_HD.pdf (Accessed online on February 14, 2023)

Gogé, F., Gomez, C., Jolivet, C., Joffre, R., 2014. Which strategy is best to predict soil properties of a local site from a national Vis–NIR database? Geoderma 213, 1–9. https://doi.org/10.1016/j.geoderma.2013.07.016

Gomez, C.; Viscarra Rossel, R.A.; McBratney, A.B., 2008. Soil Organic Carbon Prediction by Hyperspectral Remote Sensing and Field Vis-NIR Spectroscopy: An Australian Case Study. Geoderma 146, 403–411, doi:10.1016/j.geoderma.2008.06.011

Gomez, C., Lagacherie, P., Coulouma, G., 2012. Regional predictions of eight common soil properties and their spatial structures from hyperspectral Vis-NIR data. Geoderma 189–190, 176–185. https://doi.org/10.1016/j.geoderma.2012.05.023

Gomez, C., Dharumarajan, S., Lagacherie, P., Riotte, J., Ferrant, S., Sekhar, M., Ruiz, L., 2021. Mapping of tank silt application using Sentinel-2 images over the Berambadi catchment (India). Geoderma Régional, 25. https://doi.org/10.1016/j.geodrs.2021.e00389

Gomez, C., Vaudour, E., Féret, J.-B., de Boissieu, F., Dharumarajan, S., 2022. Topsoil clay content mapping in croplands from Sentinel-2 data: Influence of atmospheric correction methods across a season time series. Geoderma 423, 115959. https://doi.org/10.1016/j.geoderma.2022.115959

Grasty, R.L., 1997. Radon emanation and soil moisture effects on airborne gamma-ray measurements. Geophysics 62, 1379–1385. https://doi.org/10.1190/1.1444242

Grinand, C., Arrouays, D., Laroche, B., Martin, M.P., 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. Geoderma 143, 180–190. https://doi.org/10.1016/j.geoderma.2007.11.004

Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. Soil Sci. Soc. Am. J. 75, 1201–1213. https://doi.org/10.2136/sssaj2011.0025

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van der Linden, S., Leitão, J.P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP spaceborne imaging spectroscopy mission for Earth observation. Remote Sens. 7. https://doi.org/10.3390/rs70708830

Guevara, M.; Arroyo, C.; Brunsell, N.; Cruz, C.O.; Domke, G.; Equihua, J.; Etchevers, J.; Hayes, D.; Hengl, T.; Ibelles, A.; et al., 2020. Soil Organic Carbon Across Mexico and the Conterminous United States (1991–2010). Glob. Biogeochem. Cycles, 34, e2019GB006219. https://doi.org/10.1029/2019gb006219.

Guiresse, A.; Cambou, E.; Collin Bellier, C.; Denjean, A.; Falba, P.; Guigues, E.; Mouclier, M.; Muller, N.; Nesling, E.; Party, J.-P.; et al., 2014. Les Pédo-paysages des plaines centrales de Midi-Pyrénées. Etude Gest. Des Sols, 21, 77–84.

Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V., 2002. Gene Selection for Cancer Classification using Support Vector Machines. Mach. Learn. 46, 389–422. Available online: https://link.springer.com/article/10.1023/A:1012487302797 (accessed on 4 April 2023).

Hadjimitsis, D.G., Papadavid, G., Agapiou, A., Themistocleous, K., Hadjimitsis, M.G., Retalis, A., Michaelides, S., Chrysoulakis, N., Toulios, L., Clayton, C.R.I., 2010. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. Nat. Hazards Earth Syst. Sci. 10, 89–95

Hartemink, A.E., Krasilnikov, P., Bockheim, J.G., 2013. Soil maps of the world. Geoderma, 207, 256-267. https://doi.org/10.1016/j.geoderma.2013.05.003

Heiden, U., d'Angelo, P., Schwind, P., Karlshöfer, P., Müller, R., Zepp, S., Wiesmeier, M., Reinartz, P., 2022. Soil Reflectance Composites—Improved Thresholding and Performance Evaluation. Remote Sensing, 14(18), 4526. https://doi.org/10.3390/rs14184526

Hengl, T.; De Jesus, J.M.; MacMillan, R.A.; Batjes, N.H.; Heuvelink, G.B.M.; Ribeiro, E.; Samuel-Rosa, A.; Kempen, B.; Leenaars, J.G.; Walsh, M.G.; et al., 2014. SoilGrids1km—Global Soil Information Based on Automated Mapping. PLoS ONE, 9, e105992. https://doi.org/10.1371/journal.pone.0105992.

Hengl, T.; Heuvelink, G.B.M.; Kempen, B.; Leenaars, J.G.B.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; Macmillan, R.A.; De Jesus, J.M.; Tamene, L.; et al., 2015. Mapping soil properties of Africa at 250

m resolution: Random forests significantly improve current predictions. PLoS ONE, 10, e0125814. https://doi.org/10.1371/journal.pone.0125814.

Hengl, T., De Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning, PLoS ONE. https://doi.org/10.1371/journal.pone.0169748

Huang, X., Senthilkumar, S., Kravchenko, A., Thelen, K., Qi, J., 2007. Total carbon mapping in glacial till soils using near-infrared spectroscopy, Landsat imagery and topographical information. Geoderma 141, 34–42. https://doi.org/10.1016/j.geoderma.2007.04.023

IAEA, 2003. Guidelines for radioelement mapping using gamma ray spectrometry data. IAEA-TECDOC-1363.

IGN, 2011. IGN (Institut Géographique National) . URL https://geoservices.ign.fr/sites/default/files/2021-07/DC_BDALTI_2-0.pdf

INSEE, 2021, *La France et ses territoires*, Paris https://www.insee.fr/fr/statistiques/5039859?sommaire=5040030#:~:text=En%20France%2C%20en%202019%2C%20la,cultures%20permanentes%20(figure%201) (Accessed online on February 14, 2023)

IUSS Working Group WRB., 2015. World Reference Base for Soil Resources 2014, update 2015 International soil classification system for naming soils and creating legends for soil maps. FAO, Rome

Jaber, S.M.; Lant, C.L.; Al-Qinna, M.I., 2011 Estimating spatial variations in soil organic carbon using satellite hyperspectral data and map algebra. International Journal of Remote Sensing 32, 5077–5103. https://doi.org/10.1080/01431161.2010.494637

Jacquemoud, S., Baret, F., & Hanocq, J. F., 1992. Modeling spectral and directional soil reflectance. Remote sensing of Environment, 41, 123−132.

Jarmer, T.; Hill, J.; Lavée, H.; Sarah, P., 2010. Mapping Topsoil Organic Carbon in Non-agricultural Semi-arid and Arid Eco-systems of Israel. Photogrammetric Engineering & Remote Sensing 76, 85–94. https://doi.org/10.14358/PERS.76.1.85

Jolivet, C.; Arrouays, D.; Andreux, F.; Lévêque, J., 1997. Soil organic carbon dynamics in cleared temperate forest spodosols converted to maize cropping. Plant Soil, 191, 225–231. https://doi.org/10.1023/a:1004294822799.

Jolivet, C., Arrouays, D., Boulonne, L., Ratié, C., Saby, N., 2006. Le réseau de mesures de la qualité des sols de France (RMQS)-Etat d'avancement et premiers résultats. Etude Gest. des sols 13, 149-164 (in French).

Jones, A.; Fernandez Ugalde, O.; Scarpa, S., 2020. LUCAS 2015 Topsoil Survey; EUR 30332 EN, JRC121325; Publications Office of the European Union: Luxembourg; ISBN 978-92-76-21080-1. https://doi.org/10.2760/616084

Kidd, D., Webb, M., Malone, B., Minasny, B., McBratney, A., 2015. Digital soil assessment of agricultural suitability, versatility and capital in Tasmania. Australia. Geoderma. Regional 6, 7-21. http://dx.doi.org/10.1016/j.geodrs.2015.08.005

King, D., Daroussin, J., Tavernier, R., 1994. Development of a soil geographic database from the Soil Map of the European Communities. Catena 21, 37–56. https://doi.org/10.1016/0341-8162(94)90030-2

Koch, A., Mcbratney, A., Adams, M., Field, D., Hill, R., Crawford, J., Minasny, B., Lal, R., Abbott, L., O'Donnell, A., Angers, D., Baldock, J., Barbier, E., Binkley, D., Parton, W., Wall, D.H., Bird, M., Bouma, J., Chenu, C., Flora, C.B., Goulding, K., Grunwald, S., Hempel, J., Jastrow, J., Lehmann, J., Lorenz, K., Morgan, C.L., Rice, C.W., Whitehead, D., Young, I., Zimmermann, M., 2013. Soil Security: Solving the Global Soil Crisis. Glob. Policy 4, 434–441. https://doi.org/10.1111/1758-5899.12096

Kuhn, M., 2022. caret: classification and Regression Training. R Packag. version 6.0-93. https://doi.org/https://cran.r-project.org/web/packages/caret/index.html

Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. J. Stat. Softw. 2010, 36, 1–13.

Lacoste, M., Lemercier, B., Walter, C., 2011. Regional mapping of soil parent material by machine learning based on punctual training data. Geomorphology 133 (1–2), 90–99.

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014a. High resolution 3d mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311. https://doi.org/10.1016/j.geoderma.2013.07.002

Lacoste, M., Martin, M.P., Saby, N.P.A., Paroissien, J.B., Lehmann, S., Richer de Forges, A.C., Arrouays, D., 2014b. Carbon content and stocks in the O-horizons of French forest soils. Globalsoilmap: Basis of the Global Spatial Soil Information System. Crc Press-Taylor & Francis Group, Boca Raton, pp. 91–97

Lacoste, M., Viaud, V., Michot, D., Walter, C., 2016. Model-based evaluation of impact of soil redistribution on soil organic carbon stocks in a temperate hedgerow landscape. Earth Surf. Process. Landforms 41, 1536–1549. https://doi.org/10.1002/esp.3925.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M., Saby, N.P.A., 2019. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma 337, 1320–1328. https://doi.org/10.1016/j.geoderma.2018.08.024

Lal, R., 2007. Carbon management in agricultural soils. Mitig. Adapt. Strateg. Glob. Chang. 12, 303–322. https://doi.org/10.1007/s11027-006-9036-7

Lal, R., 2008. Carbon sequestration. Philos. Trans. R. Soc. B Biol. Sci. 363, 815–830. https://doi.org/10.1098/rstb.2007.2185

Lal, R., Bouma, J., Brevik, E., Dawson, L., Field, D.J., Glaser, B., Hatano, R., Hartemink, A.E., Kosaki, T., Lascelles, B., Monger, C., Muggler, C., Ndzana, G.M., Norra, S., Pan, X., Paradelo, R., Reyes-Sánchez, L.B., Sandén, T., Singh, B.R., Spiegel, H., Yanai, J., Zhang, J., 2021. Soils and sustainable

development goals of the United Nations: An International Union of Soil Sciences perspective. Geoderma Reg. 25. https://doi.org/10.1016/j.geodrs.2021.e00398

LE Bissonnais, Y.; Arrouays, D., 1997. Aggregate stability and assessment of soil crustability and erodibility: II. Application to humic loamy soils with various organic carbon contents. Eur. J. Soil Sci. 48, 39–48. https://doi.org/10.1111/j.1365-2389.1997.tb00183.x.

Lee, Christine M., Cable, Morgan L., Hook, Simon J., Green, Robert O., Ustin, Susan L., Mandl, Daniel J., Middleton, Elizabeth M., 2015. An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. Remote Sens. Environ. (0034-4257) 167, 6–19. https://doi.org/10.1016/j.rse.2015.06.012.

Lemercier, B., Lagacherie, P., Amelin, J., Sauter, J., Richer-de-Forges, A.C., Arrouays, D., 2022. Multiscale evaluations of global, national and regional digital soil mapping products in France across regions and soil properties. Geoderma, 425, https://doi.org/10.1016/j.geoderma.2022.116052.

Lin, L.I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45, 255–268. https://doi.org/10.2307/2532051

Loiseau, T., Chen, S., Mulder, V.L., Román Dobarco, M., Richer-de-Forges, A.C., Lehmann, S., Bourennane, H., Saby, N.P.A., Martin, M.P., Vaudour, E., Gomez, C., Lagacherie, P., Arrouays, D., 2019. Satellite data integration for soil clay content modelling at a national scale. Int. J. Appl. Earth Obs. Geoinf. 82. https://doi.org/10.1016/j.jag.2019.101905

Loiseau, T., Richer-de-Forges, A.C., Martelet, G., Bialkowski, A., Nehlig, P., Arrouays, D., 2020. Could airborne gamma-spectrometric data replace lithological maps as co-variates for digital soil mapping of topsoil particle-size distribution? A case study in Western France. Geoderma Reg. 22, e00295. https://doi.org/10.1016/j.geodrs.2020.e00295

Loizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G., 2018. Prisma: The Italian hyperspectral mission. In Proceedings of the IGARSS IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 175–178.

Lombardo, L., Saia, S., Schillaci, C., Mai, P.M., Huser, R., 2018. Modeling soil organic carbon with Quantile Regression: Dissecting predictors' effects on carbon stocks. Geoderma 318, 148–159. https://doi.org/10.1016/j.geoderma.2017.12.011

Macaire, J.M. Etude sédimentologique des formations superficielles sur le tracé de l'autoroute A10 entre Artenay et Meung sur Loire. (PhD thesis), University of Orleans, France (1971)

Martelet, G., Drufin, S., Tourliere, B., Saby, N.P.A., Perrin, J., Deparis, J., Prognon, F., Jolivet, C., Ratié, C., Arrouays, D., 2013. Regional Regolith Parameter Prediction Using the Proxy of Airborne Gamma Ray Spectrometry. Vadose Zo. J. 12, vzj2013.01.0003. https://doi.org/10.2136/vzj2013.01.0003

Martelet, G., Nehlig, P., Arrouays, D., Messner, F., Tourlière, B., Laroche, B., Deparis, J., Saby, N., Richer-de-Forges, A.C., Jolivet, C., Ratié, C., 2014. Airborne gamma-ray spectrometry: potential for regolith-soil mapping and characterization, in: GlobalSoilMap: Basis of the Global Spatial Soil Information System, Taylor & Francis, CRC Press, London (2014). pp. 401–408.

Martin, M.P.; Wattenbach, M.; Smith, P.; Meersmans, J.; Jolivet, C.; Boulonne, L.; Arrouays, D., 2011. Spatial Distribution of Soil Organic Carbon Stocks in France. Biogeosciences 8, 1053–1065, doi:10.5194/bg-8-1053-2011.

Martin, M.P.; Dimassi, B.; Dobarco, M.R.; Guenet, B.; Arrouays, D.; Angers, D.A.; Blache, F.; Huard, F.; Soussana, J.; Pellerin, S., 2021. Feasibility of the 4 per 1000 aspirational target for soil carbon: A case study for France. Glob. Chang. Biol. 27, 2458–2477. https://doi.org/10.1111/gcb.15547.

Mathieu, R.; Pouget, M.; Cervelle, B.; Escadafal, R., 1998. Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment. Remote Sens. Environ. 66, 17–28.

Matinfar, H.R.; Maghsodi, Z.; Mousavi, S.R.; Rahmani, A., 2021. Evaluation and Prediction of Topsoil organic carbon using Machine learning and hybrid models at a Field-scale. CATENA 202, 105258. https://doi.org/10.1016/j.catena.2021.105258

McBratney, A.B., Mendoca Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/https://doi.org/10.1016/S0016-7061(03)00223-4

Meersmans, J., Martin, M., Lacarce, E., De Baets, S., Jolivet, C., Boulonne, L., Lehmann, S., Saby, N., Bispo, A., Arrouays, D., 2012. A high resolution map of French soil organic carbon. Agron. Sustain. Dev. 32 (4), 841–851

Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.

Meinshausen, N., 2017. quantregForest: Quantile Regression Forests. R Packag. version 1.3-7.

Meliyo, J.L.; Msanya, M.B.; Kimaro, D.N.; Massawe, J.B.H.; Hieronimo, P.; Mulungu, L.S.; Deckers, J.; Gulinck, H., 2016. Variability of soil organic carbon with landforms and land use in the Usambara Mountains of Tanzania. J. Soil Sci. Environ. Manag. 7, 123–132. https://doi.org/10.5897/JSSEM2016.0557.

Mello, D.C. De, Veloso, G.V., Lana, M.G. De, Mello, F.A.D.O., Poppiel, R.R., Cabrero, D.R.O., Di Raimo, L.A.D.L., Schaefer, C.E.G.R., Filho, E.I.F., Leite, E.P., Demattê, J.A.M., 2022. A new methodological framework for geophysical sensor combinations associated with machine learning algorithms to understand soil attributes. Geosci. Model Dev. 15, 1219–1246. https://doi.org/10.5194/gmd-15-1219-2022

Mello, F.A.O.; Demattê, J.A.M.; Rizzo, R.; Dotto, A.C.; Poppiel, R.R.; Mendes, W.D.S.; Guimarães, C.C.B., 2020. Expert-based maps and highly detailed surface drainage models to support digital soil mapping. Geoderma, 384, 114779. https://doi.org/10.1016/j.geoderma.2020.114779.

Mevik, B.-H.; Wehrens, R., 2007. The pls package: Partial least squares and principal component regression. J. Stat. Softw. 18, 1–24.

Minasny, B., McBratney, A.B., Hartemink, A.E., 2010. Global pedodiversity, taxonomic distance, and World Reference Base. Geoderma 155, 132–139.

Minasny, B.; McBratney, A.; Bellon-Maurel, V.; Roger, J.-M.; Gobrecht, A.; Ferrand, L.; Joalland, S., 2011. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the

prediction of soil organic carbon. Geoderma, 167–168, 118–124. https://doi.org/10.1016/j.geoderma.2011.09.008.

Minasny, B.; Whelan, B.M.; Triantafilis, J.; McBratney, A.B., 2013. Pedometrics Research in the Vadose Zone-Review and Perspectives. Vadose Zone J. 12, vzj2012.0141. https://doi.org/10.2136/vzj2012.0141.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.C., Vågen, T.G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

Minty, B., 1997. Fundamentals of airborne gamma-ray spectrometry. AGSO J. Aust. Geol. Geophys. 17, 39–50.

Mirzaee, S. ; Ghorbani-Dashtaki, S. ; Mohammadi, J. ; Asadi, H. ; Asadzadeh, F., 2016. Spatial variability of soil organic matter using remote sensing data. CATENA, 145, 118–127. https://doi.org/10.1016/j.catena.2016.05.023

Mishra, U., Lal, R., Liu, D.S., Van Meirvenne, M., 2010. Predicting the spatial variation of the soil organic carbon pool at a regional scale. Soil Sci. Soc. Am. J. 74, 906–914.

Moeys, J. The soil texture wizard: R functions for plotting, classifying, transforming and exploring soil texture data. R pack-age, version 1.2.13, R vignette, 2018; 104 . https://cran.r-project.org/web/packages/soiltexture/vignettes/soiltexture_vignette.pdf (accessed on 12 December 2021)

Mohamed, E.S. ; Baroudy, A.A.E. ; El-Beshbeshy, T. ; Emam, M. ; Belal, A.A. ; Elfadaly, A. ; Aldosari, A.A. ; Ali, A.M. ; Lasaponara, R., 2020. Vis-NIR Spectroscopy and Satellite Landsat-8 OLI Data to Map Soil Nutrients in Arid Conditions: A Case Study of the Northwest Coast of Egypt. Remote Sensing, 12, 3716. https://doi.org/10.3390/rs12223716

Moinard, V., Levavasseur, F., Houot, S., 2021. Current and potential recycling of exogenous organic matter as fertilizers and amendments in a French peri-urban territory. Resour. Conserv. Recycl. 169, 105523. https://doi.org/10.1016/j.resconrec.2021.105523

Montanarella, L., Pennock, D.J., McKenzie, N., Badraoui, M., Chude, V., Baptista, I., Mamo, T., Yemefack, M., Aulakh, M.S., Yagi, K., Hong, S.Y., Vijarnsorn, P., Zhang, G.L., Arrouays, D., Black, H., Krasilnikov, P., Sobocká, J., Alegre, J., Henriquez, C.R., Mendonça-Santos, M. de L., Taboada, M., Espinosa-Victoria, D., AlShankiti, A., AlaviPanah, S.K., Mustafa Elsheikh, E.A. El, Hempel, J., Arbestain, M.C., Nachtergaele, F., Vargas, R., 2016. World's soils are under threat. Soil 2, 79–82. https://doi.org/10.5194/soil-2-79-2016

Møller, A., Beucher, A.M., Pouladi, N., Humlekrog Greve, M., 2020. Oblique geographic coordinates as covariates for digital soil mapping. Soil 6, 269–289. https://doi.org/10.5194/soil-6-269-2020

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016a. GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. Sci. Total Environ. 573, 1352–1369. https://doi.org/10.1016/j.scitotenv.2016.07.066

Mulder, V.; Lacoste, M.; Richer-De-Forges, A.; Martin, M.; Arrouays, D., 2016b. National versus global modelling the 3D distribution of soil organic carbon in mainland France. Geoderma, 263, 16–34. https://doi.org/10.1016/j.geoderma.2015.08.035.

Mzid, N.; Castaldi, F.; Tolomio, M.; Pascucci, S.; Casa, R.; Pignatti, S., 2022. Evaluation of Agricultural Bare Soil Properties Retrieval from Landsat 8, Sentinel-2 and PRISMA Satellite Data. Remote Sensing 14. 714. https://doi.org/10.3390/rs14030714

Nanni, M.R.; Demattê, J.A.M., 2006 Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. Soil Sci. Soc. Am. J. 70, 393–407, doi:10.2136/sssaj2003.0285.

Nield, S.J., Boettinger, J.L., Ramsey, R.D., 2007. Digitally Mapping Gypsic and Natric Soil Areas Using Landsat ETM Data. Soil Sci. Soc. Am. J. 71, 245–252. https://doi.org/10.2136/sssaj2006-0049

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E. Ben, Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Adv. Agron. 132, 139–159. https://doi.org/10.1016/bs.agron.2015.02.002

Nowkandeh, S.M.; Homaee, M.; Noroozi, A.A., 2013. Mapping Soil Organic Matter Using Hyperion Images. International Journal of Agronomy and Plant Production 4 (8), 1753-1759.

Orgiazzi, A.; Ballabio, C.; Panagos, P.; Jones, A.; Fernández-Ugalde, O.. 2018, LUCAS Soil, the Largest Expandable Soil Data-set for Europe: A Review. Eur J Soil Sci 2018, 69, 140–153, doi:10.1111/ejss.12499.

Ou, D.; Tan, K.; Lai, J.; Jia, X.; Wang, X.; Chen, Y.; Li, J., 2020. Semi-supervised DNN regression on airborne hyperspectral imagery for improved spatial soil properties prediction. Geoderma, 385, 114875. https://doi.org/10.1016/j.geoderma.2020.114875.

Padarian, J., Minasny, B., McBratney, A.B., 2017. Chile and the Chilean soil grid: A contribution to GlobalSoilMap. Geoderma Reg. 9, 17–28. https://doi.org/10.1016/j.geodrs.2016.12.001

Paroissien, J.., Saby, N., de Forges, A.., Arrouays, D., Louis, B., 2014. Populating soil maps with legacy data from a soil testing databases., in: GlobalSoilMap: Basis of the Global Spatial Soil Information System. Taylor & Francis, CRC Press, London (2014), pp. 319–324.

Patton, N.R.; Lohse, K.A.; Seyfried, M.S.; Godsey, S.E.; Parsons, S.B., 2019. Topographic controls of soil organic carbon on soil-mantled landscapes. Sci. Rep. 9, 6390. https://doi.org/10.1038/s41598-019-42556-5.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G.P., Smith, P., 2016. Climate-smart soils. Nature 532, 49–57. https://doi.org/10.1038/nature17174

Pignatti, S.; Acito, N.; Amato, U.; Casa, R.; Castaldi, F.; Coluzzi, R.; De Bonis, R.; Diani, M.; Imbrenda, V.; Laneve, G.; et al. 2015., Environmental products overview of the Italian hyperspectral prisma mission: The SAP4PRISMA project. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31, 3997–4000.

Poeplau, C.; Bolinder, M.A.; Kätterer, T., 2016. Towards an Unbiased Method for Quantifying Treatment Effects on Soil Carbon in Long-Term Experiments Considering Initial within-Field Variation. Geoderma. 267, 41–47, doi:10.1016/j.geoderma.2015.12.026.

Poggio, L.; Gimona, A.; Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma. 209–210, 1–14. https://doi.org/10.1016/j.geoderma.2013.05.029

Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D., 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. SOIL, 7, 217–240. https://doi.org/10.5194/soil-7-217-2021

Potapov, P., Turubanova, S., Hansen, M.C., Tyukavina, A., Zalles, V., Khan, A., Song, X.P., Pickens, A., Shen, Q., Cortez, J., 2022. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. Nat. Food 3, 19–28. https://doi.org/10.1038/s43016-021-00429-z

Pouget, M.; Madeira Netto, J.; Le Floc'h, E.; Kamal, S.,1991. Caractéristiques spectrales des surfaces sableuses de la région côtière Nord-Ouest de l'Egypte: Application aux données satellitaires SPOT. In Caractérisation et Suivi des Milieux Terrestres en Régions Arides et Tropicales; Pouget, M., Ed.; ORSTOM: Paris, France; pp. 27–38, ISBN 2-7099-1051-9. Available online: https: //horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_6/colloques2/34605.pdf (accessed on 4 April 2023).

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramankutty, N., Evan, A.T., Monfreda, C., Foley, J.A., 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. Global Biogeochem. Cycles 22, 1–19. https://doi.org/10.1029/2007GB002952

Rast, M., Nieke, J., Adams, J., Isola, C., Gascon, F., 2021. Copernicus Hyperspectral Imaging Mission for the Environment (Chime), 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, pp. 108-111, doi: 10.1109/IGARSS47720.2021.9553319.

Redon, P.-O.; Bur, T.; Guiresse, M.; Probst, J.-L.; Toiser, A.; Revel, J.-C.; Jolivet, C.; Probst, A., 2013. Modelling trace metal background to evaluate anthropogenic contamination in arable soils of south-western France. Geoderma, 206, 112–122. https://doi.org/10.1016/j.geoderma.2013.04.023.

Reinhardt, N.; Herrmann, L. Gamma-ray spectrometry as versatile tool in soil science: A critical review. J. Plant Nutr. Soil Sci. 2019, 182, 9–27. https://doi.org/10.1002/jpln.201700447

Richer de Forges, A.C., 2008. Référentiel pédologique de la Région Centre. Notice explicative de la carte des pédopaysages du Loiret à 1/50 000. INRA, éditions QUAE, Versailles, France.

270 p. (in French).

Rienzi, E.A.; Mijatovic, B.; Mueller, T.G.; Matocha, C.J.; Sikora, F.J.; Castrignanò, A., 2014. Prediction of Soil Organic Carbon under Varying Moisture Levels using Reflectance Spectroscopy. Soil Sci. Soc. Am. J., 78, 958–967. https://doi.org/10.2136/sssaj2013.09.0408.

Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). Remote Sens. Environ. 205, 1–17. https://doi.org/10.1016/j.rse.2017.11.004

Safanelli, J.L.; Demattê, J.A.; Chabrillat, S.; Poppiel, R.R.; Rizzo, R.; Dotto, A.C.; Silvero, N.E.; Mendes, W.D.S.; Bonfatti, B.R.; Ruiz, L.F.; et al. Leveraging the application of Earth observation data for mapping cropland soils in Brazil. Geoderma 2021, 396, https://doi.org/10.1016/j.geoderma.2021.115042.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., De Lourdes Mendonça-Santos, M., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vågen, T.G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.L., 2009. Digital soil map of the world. Science (80-.). 325, 680–681. https://doi.org/10.1126/science.1175084

Sayão, V.M., Demattê, J.A., Bedin, L.G., Nanni, M.R., Rizzo, R., 2018. Satellite land surface temperature and reflectance related with soil attributes. Geoderma 325, 125–140. https://doi.org/10.1016/j.geoderma.2018.03.026

Schlesinger, W., 1984. Soil organic matter: A source of atmospheric CO2. Role Terr. Veg. Glob. Carbon Cycle Meas. Remote Sens. 4, 111–127.

Silvero, N.E.Q.; Raimo, L.A.D.L.D.; Pereira, G.S.; de Magalhães, L.P.; Terra, F.D.S.; Dassan, M.A.A.; Urbina-Salazar, D.F.U.; Demattê, J.A., 2020. Effects of water, organic matter, and iron forms in mid-IR spectra of soils: Assessments from laboratory to satellite-simulated data. Geoderma, 375, 114480. https://doi.org/10.1016/j.geoderma.2020.114480.

Silvero, N.E.Q., Demattе, J.A.M., Vieira, J. De S., Mello, F.A. De O., Amorim, M.T.A., Poppiel, R.R., Mendes, W. De S., Bonfatti, B.R., 2021a. Soil property maps with satellite images at multiple scales and its impact on management and classification. Geoderma 397, 11508. https://doi.org/10.1016/j.geoderma.2021.115089

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., 2021b. Soil Variability and Quantification Based on Sentinel-2 and Landsat-8 Bare Soil Images: A Comparison. Remote Sensing of Environment, 252, 112117, doi:10.1016/j.rse.2020.112117.

Smith, P.; Bustamante, M.; Ahammad, H.; Clark, H.; Dong, H.; Elsiddig, E.A.; Haberl, H.; Harper, H.; House, J.; Jafari, M.; et al., 2014. Agriculture, Forestry and Other Land Use (AFOLU). In Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. In Climate Change 2014 Mitigation of Climate Change; IPCC: Geneva, Switzerland; pp. 811–922.

Smith, P.; Soussana, J.-F.; Angers, D.; Schipper, L.; Chenu, C.; Rasse, D.P.; Batjes, N.H.; van

Egmond, F.; McNeill, S.; Kuhnert, M.; et al., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob. Chang. Biol. 26, 219–241. https://doi.org/10.1111/gcb.14815.

Stamnes, K., Li, W., Eide, H., & Stamnes, J. J.,2005. Challenges in atmospheric correction of satellite imagery. Optical Engineering, 44, 041003-1−041003-9

Steinberg, A.; Chabrillat, S.; Stevens, A.; Segl, K.; Foerster, S., 2016. Prediction of Common Surface Soil Properties Based on Vis-NIR Airborne and Simulated EnMAP Imaging Spectroscopy Data: Prediction Accuracy and Influence of Spatial Res-olution. Remote Sensing 8, 613. https://doi.org/10.3390/rs8070613

Stevens, A.; van Wesemael, B.; Bartholomeus, H.; Rosillon, D.; Tychon, B.; Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma, 144, 395–404. https://doi.org/10.1016/j.geoderma.2007.12.009.

Stevens, A.; Udelhoven, T.; Denis, A.; Tychon, B.; Lioy, R.; Hoffmann, L.; van Wesemael, B., 2010. Measuring soil organic carbon in croplands at regional scale using airborne imaging spectroscopy. Geoderma, 158, 32–45. https://doi.org/10.1016/j.geoderma.2009.11.032.

Stoner, E.R., Baumgardner, M.F., 1981. Characteristic Variations in Reflectance of Surface Soils. Soil Sci. Soc. Am. J. 45, 1161–1165. https://doi.org/10.2136/sssaj1981.03615995004500060031x

Sullivan, D.G.; Shaw, J.N.; Rickman, D., 2005. IKONOS Imagery to Estimate Surface Soil Property Variability in Two Ala-bama Physiographies. Soil Sci. Soc. Am. J. 69, 1789–1798, doi:10.2136/sssaj2005.0071.

Tan, Y.; Jiang, Q.; Yu, L.; Liu, H.; Zhang, B., 2021. Reducing the Moisture Effect and Improving the Prediction of Soil Organic Matter with VIS-NIR Spectroscopy in Black Soil Area. IEEE Access, 9, 5895–5905. https://doi.org/10.1109/access.2020.3048794.

Thoisy, J., Mistou, M.N., Etayo, A., Vaudour, E., Latrille, E., Rossard, V., Girardin, C., 2022. Spectra transfer model of NIR spectra of soil samples, https://doi.org/10.15454/LXKFAS, Recherche Data Gouv, V1

Thompson, J.A.; Pena-Yewtukhiw, E.M.; Grove, J.H., 2006. Soil–landscape modeling across a physiographic region: Topographic patterns and model transportability. Geoderma, 133, 57–70. https://doi.org/10.1016/j.geoderma.2006.03.037.

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2019. A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved VIS-NIR-SWIR soil properties estimation. Geoderma, 340. https://doi.org/10.1016/j.geoderma.2018.12.044

Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. Remote Sens. 12, 1389. https://doi.org/10.3390/RS12091389.

Urbina-Salazar, D.F. ; Demattê, J.A.M. ; Vicente, L.E. ; Guimarães, C.C.B. ; Sayão, V.M. ; Cerri,

C.E.P. ; de C. Padilha, M.C. ; Mendes, W.D.S., 2020. Emissivity of agricultural soil attributes in southeastern Brazil via terrestrial and satellite sensors. Geoderma 361, 114038. https://doi.org/10.1016/j.geoderma.2019.114038

Urbina-Salazar, D., Vaudour, E., Baghdadi, N., Ceschia, E., Richer-de-Forges, A.C., Lehmann, S., Arrouays, D., 2021. Using Sentinel-2 images for soil organic carbon content mapping in croplands of southwestern France. The usefulness of Sentinel-1/2 derived moisture maps and mismatches between sentinel images and sampling dates. Remote Sens. https://doi.org/10.3390/rs13245115.

Urbina-Salazar, D.; Vaudour, E.; Richer-de-Forges, A.C.; Chen, S.; Martelet, G.; Baghdadi, N.; Arrouays, D., 2023. Sentinel-2 and Sentinel-1 Bare Soil Temporal Mosaics of 6-Year Periods for Soil Organic Carbon Content Mapping in Central France. Remote Sens. 15, 2410. https://doi.org/10.3390/rs15092410

Van Der Klooster, E., Van Egmond, F.M., Sonneveld, M.P.W., 2011. Mapping soil clay contents in Dutch marine districts using gamma-ray spectrometry. Eur. J. Soil Sci. 62, 743–753. https://doi.org/10.1111/j.1365-2389.2011.01381.x

van Egmond, F.M., Loonstra, E.H., Limburg, J., 2010. Gamma Ray Sensor for Topsoil Mapping: The Mole, in: Proximal Soil Sensing. Springer Netherlands, Dordrecht, pp. 323–332. https://doi.org/10.1007/978-90-481-8859-8_27

Vaudour E., Girard M.C., 2010. Pédologie. Chap. 23. In : Girard MC & Girard CM, Traitement des données de télédétection, Dunod, Paris

Vaudour, E.; Bel, L.; Gilliot, J.M.; Coquet, Y.; Hadjar, D.; Cambier, P.; Michelin, J.; Houot, S., 2013. Potential of SPOT Multi-spectral Satellite Images for Mapping Topsoil Organic Carbon Content over Peri-Urban Croplands. Soil Science Society of America Journal 77, 2122. https://doi.org/10.2136/sssaj2013.02.0062

Vaudour, E., Gilliot, J.M., Bel, L., Bréchet, L., Hamiache, J., Hadjar, D., Lemonnier, Y., 2014. Uncertainty of soil reflectance retrieval from SPOT and RapidEye multispectral satellite images using a per-pixel bootstrapped empirical line atmospheric correction over an agricultural region. International Journal of Applied Earth Observation and Geoinformation 26, 217–234. https://doi.org/10.1016/j.jag.2013.07.003

Vaudour, E., Gilliot, J.M., Bel, L., Lefevre, J., Chehdi, K., 2016. Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. Int. J. Appl. Earth Obs. Geoinf 49, 24–38. https://doi.org/10.1016/j.jag.2016.01.005

Vaudour, E.; Gomez, C.; Loiseau, T.; Baghdadi, N.; Loubet, B.; Arrouays, D.; Ali, L.; Lagacherie, P., 2019a. The Impact of Acquisition Date on the Prediction Performance of Topsoil Organic Carbon from Sentinel-2 for Croplands. Remote Sens. 11, 2143. https://doi.org/10.3390/rs11182143.

Vaudour, E.; Gomez, C.; Fouad, Y.; Lagacherie, P., 2019b. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. Remote Sens. Environ. 223, 21–33. https://doi.org/10.1016/j.rse.2019.01.006.

Vaudour, E.; Gomez, C.; Lagacherie, P.; Loiseau, T.; Baghdadi, N.; Urbina-Salazar, D.; Loubet, B.; Arrouays, D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. Int. J. Appl. Earth Obs. Geoinf. 96, 102277. https://doi.org/10.1016/j.jag.2020.102277.

Vaudour, E., Gholizadeh, A., Castaldi, F., Saberioon, M., Borůvka, L., Urbina-Salazar, D., et al., 2022. Satellite imagery to map topsoil organic carbon content over cultivated areas: an overview. Remote Sens. 14, 2917. https://doi.org/10.3390/rs14122917

Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). Geoderma Regional 4, 20–30. https://doi.org/10.1016/j.geodrs.2014.11.003

Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma 291, 55–64. https://doi.org/10.1016/j.geoderma.2016.12.017

Verley, F., 2020. Lessons from twenty years of local volumetric groundwater management: the case of the Beauce Aquifer, Central France, in: Sustainable Groundwater Management. Springer, pp. 93–108. https://doi.org/10.1142/s2382624x22800017

Viel, E. Etude des processus de transport des solutés hors équilibre physique : application à la zone non saturée des calcaires de Beauce. (PhD thesis), University of Orleans, France (2016)

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131, 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

Viscarra Rossel, R.A., Rizzo, R., Demattê, J.A.M., Behrens, T., 2010. Spatial modeling of a soil fertility index using visible-near-infrared spectra and terrain attributes. Soil Sci. Soc. Am. J. 74 (4), 1293–1300.

Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res. 53, 845–864. https://doi.org/10.1071/SR14366

Von Gunten, H.R., Surbeck, H., Rössler, E., 1996. Uranium series disequilibrium and high thorium and radium enrichments in Karst formations. Environ. Sci. Technol. 30, 1268–1274. https://doi.org/10.1021/es950473j

Wadoux, Alexandre M.J.-C., Minasny, Budiman, McBratney, Alex B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth-Science Reviews 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359.

Wadoux, A.M.-C., Molnar, C., 2022. Beyond prediction: methods for interpreting complex models of soil variation. Geoderma 422, 115953. https://doi.org/10.1016/j.geoderma.2022.115953

Wang, X., Zhang, Y., Atkinson, P.M., Yao, H., 2020. Predicting soil organic carbon content in Spain by combining landsat TM and ALOS PALSAR images. Int. J. Appl. Earth Obs. Geoinf. 92,

https://doi.org/10.1016/j.jag.2020.102182.

Wang, K.; Qi, Y.; Guo, W.; Zhang, J.; Chang, Q., 2021. Retrieval and Mapping of Soil Organic Carbon Using Sentinel-2A Spectral Images from Bare Cropland in Autumn. Remote Sensing, 13, 1072. https://doi.org/10.3390/rs13061072

Ward, K.J.; Chabrillat, S.; Brell, M.; Castaldi, F.; Spengler, D.; Foerster, S., 2020. Mapping Soil Organic Carbon for Airborne and Simulated EnMAP Imagery Using the LUCAS Soil Database and a Local PLSR. Remote Sensing 12, 3451, doi:10.3390/rs12203451.

Wilford, J., Minty, B., 2006. Chapter 16 The Use of Airborne Gamma-ray Imagery for Mapping Soils and Understanding Landscape Processes. Dev. Soil Sci. 31. https://doi.org/10.1016/S0166-2481(06)31016-1

Wold, S., 1978. Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. Technometrics, 20, 397–405. https://doi.org/10.1080/00401706.1978.10489693.

Wold, S.; Sjöström, M.; Eriksson, L., 2001. PLS-regression: A basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130. https://doi.org/10.1016/s0169-7439(01)00155-1.

Xiao, J., Shen, Y., Tateishi, R., Bayaer, W., 2006. Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. Int. J. Remote Sens. 27, 2411–2422. https://doi.org/10.1080/01431160600554363

Ye, B., Tian, S., Cheng, Q., Ge, Y., 2020. Application of lithological mapping based on advanced hyperspectral imager (AHSI) imagery onboard Gaofen-5 (GF-5) satellite. Remote Sensing 12(23), 3990.

Zaouche, M.; Bel, L.; Vaudour, E., 2017. Geostatistical Mapping of Topsoil Organic Carbon and Uncertainty Assessment in Western Paris Croplands (France). Geoderma Regional 10, 126–137, doi:10.1016/j.geodrs.2017.07.002.

Zepp, S.; Heiden, U.; Bachmann, M.; Wiesmeier, M.; Steininger, M.; van Wesemael, B., 2021. Estimation of Soil Organic Carbon Contents in Croplands of Bavaria from SCMaP Soil Reflectance Composites. Remote Sensing, 13, 3141. https://doi.org/10.3390/rs13163141

Zhou, T.; Geng, Y.; Chen, J.; Liu, M.; Haase, D.; Lausch, A., 2020a. Mapping soil organic carbon content using mul-ti-source remote sensing variables in the Heihe River Basin in China. Ecological Indicators 114, 106288. https://doi.org/10.1016/j.ecolind.2020.106288

Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A., 2020b. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. Sci. Total Environ. 729, 138244. https://doi.org/10.1016/j.scitotenv.2020.138244.

Zhou, T.; Geng, Y.; Ji, C.; Xu, X.; Wang, H.; Pan, J.; Bumberger, J.; Haase, D.; Lausch, A., 2021. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. Sci. Total Environ. 755, 142661. https://doi.org/10.1016/j.scitotenv.2020.142661.

Žížala, D.; Minařík, R.; Zádorová, T., 2019. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. Remote Sensing 11, 2947. https://doi.org/10.3390/rs11242947

Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Reyes Rojas, J., Penížek, V., Zádorová, T., 2022. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. Catena 212. https://doi.org/10.1016/j.catena.2022.106024

SUPPLEMENTARY MATERIAL

***Table S1a**. Soil moisture products (SMPs) derived from S1/S2 between February and May 2017-2021.*

			Soil Moisture (vol.%)			
S2 Acquisition Date (d_{s2}) (year/month/day)	**SMPs Date (d_{sm})** (year/month/day)	**Days** $\lvert d_{s2} - d_{sm} \rvert$	**Min**	**$\bar{x}$**	**Max**	**SD**
20170327	20170327	0	0.4	15.13	30.6	7.7
20170409	20170408	1	0.2	10.8	29.4	7.35
20180225	20180226	1	0.6	12.1	29.2	4
20180421	20180421	0	0.2	13.3	28.8	7.5
20180504	20180503	1	0.2	12.4	29.4	6.4
20190215	20190212	3	0.2	22.3	29.6	3.4
20190225	20190224	1	0.2	13.3	28.8	7.5
20190401	20190401	0	0.2	11.5	29	5.7
20190411	20190410	1	0.2	13.8	29.2	7
20190419	20190419/20[a]	0	0.2	9.3	28.6	5.8
20190514	20190514/13[a]	0	0.2	15	29	8.1
20200319	20200321	2	0.2	18.1	28.6	5.3
20200324	20200323	1	2	18.8	29.2	5.7
20200326	20200327	1	0.2	13.1	28.6	5.6
20200405	20200404	1	0.4	13	28.6	6.15
20200410	20200410	0	0.2	11.1	28	5.6
20200415	20200414	1	0.2	11.8	28.4	6.6
20200423	20200422	1	0.2	8.7	28.2	5.6
20200520	20200520	0	0.2	15.3	28.8	8.8
20200528	20200528	0	0.2	15	30.4	9.8
20210224	20210222	2	2.4	26.5	31	2.1
20210227	20210226	1	18.2	30	37.2	3
20210301	20210303	2	0.4	17.5	29.4	5
20210306	20210306	0	0.8	15.8	29.4	5.5
20210324	20210323	1	0.4	15.9	28.6	6.16
20210329	20210329	0	0.2	14.8	28.6	6.6
20210423	20210423	0	0.2	7.8	28.4	5
20210425	20210426	1	0.2	7.8	28.4	4.8

[a]SMP was a mosaic of two different dates. The images highlighted in blue ($\bar{x}$ ≥ 20vol.%) were not considered to obtain S2Bsoil_0, S2Bsoil_1 and S2Bsoil_2.

***Table S1b**. Soil moisture products (SMPs) derived from S1/S2 between July and November 2016-2021.*

S2 Acquisition Date (d_{s2}) (year/month/day)	SMPs Date (d_{sm}) (year/month/day)	Days $\lvert d_{s2} - d_{sm} \rvert$	Soil Moisture (vol.%)			
			Min	$\bar{x}$	Max	SD
20160812	20160811	1	0.2	15,2	29.8	6.3
20160822	20160823	1	0.8	15,6	30.4	5.8
20161130	20161127	3	2.6	26	30.6	2.9
20170822	20170824	2	0.2	15,4	30.6	6.1
20170827	20170824	3	0.2	15,4	30.6	6.1
20170829	20170830	1	0.6	24,5	30.8	4.2
20170921	20170923	2	1.8	18	31	5.6
20171016	20171017	1	0.4	16,9	30	6.3
20171107	20171104	3	2	23	31.2	5.2
20180725	20180720	5	0.4	14,7	30.4	6.8
20180804	20180801	3	1.6	13,4	30.2	6.6
20180819	20180819	0	0.2	13,9	30.2	6
20180901	20180831	1	1.2	13,8	30.6	6
20180911	20180912	1	2	13	30.4	5.4
20180926	20180930	4	0.2	11,6	28.6	4.9
20181008	20181006	2	1.2	12,3	30.2	5.2
20181018	20181018	0	0.4	12,1	29.4	5.2
20181021	20181024	3	0.2	12	30	5.4
20181225	20181223	2	2.6	27,5	31	1.4
20190723	20190721	2	0.2	11,7	29.2	7.1
20190725	20190725	0	0.2	10,5	28.8	9
20190906	20190907	1	0.2	12,4	29.4	5.7
20190916	20190916	0	0.2	11,9	29.6	5.6
20190918	20190919	1	0.6	12,1	29	5.3
20190921	20190919	2	0.6	12,1	29	5.3
20200722	20200719	3	0.2	12,6	28.8	8.5
20200727	20200724	3	0.2	13	30.6	7.4
20200806	20200805	1	0.6	13,2	30.6	7.2
20200915	20200917	2	0.8	13,7	28.8	5.5
20200917	20200917	0	0.8	13,7	28.8	5.5
20200922	20200922	0	0.8	12	29.6	5.6
20201104	20201103	1	18.8	31,2	38.4	3.2
20201106	20201106	0	2	20	31.4	4.7
20201121	20201122	1	0.2	21	28.6	4.1
20210719	20210719	0	0.8	15,5	31.2	6.6
20210722	20210722	0	0.6	15,7	31.2	6.5
20210902	20210902	0	1.6	15,8	29.4	5.4
20210907	20210907	0	0.6	18	28.6	5.4
20210922	20210923	1	1.8	20,2	31.4	5.9
20211022	20211023	1	18.8	28	37.8	3.8
20211109	20211110	1	0.2	20	30.8	4.4
20211129	20211129	0	0.2	18,6	28.6	3.6
20211221	20211222	1	1.2	20,9	28.4	3.6

The images highlighted in blue ($\bar{x} \geq 20$vol.%) were not considered to obtain S2Bsoil_0, S2Bsoil_1 and S2Bsoil_2.

Table S2a. *Sentinel-2 imagery and percentage of bare soil between February and May 2017-2021.*

	S2 Acquisition Date (year/month/day)	number of bare soil pixels	Bare soil pixels (%)
1	20170327	2206062	28,5
2	20170409	1535143	19,8
3	20180225	2459774	31,8
4	20180421	1233166	15,9
5	20180504	1295620	16,7
6	20190225	3815710	49,3
7	20190401	2064454	26,7
8	20190411	1576575	20,4
9	20190419	1489898	19,2
10	20190514	1362221	17,6
11	20200319	1819914	23,5
12	20200324	1968727	25,4
13	20200326	2260217	29,2
14	20200405	2215266	28,6
15	20200410	2088088	27,0
16	20200415	2014895	26,0
17	20200423	1655253	21,4
18	20200520	1292335	16,7
19	20200528	779879	10,1
20	20210301	2461543	31,8
21	20210306	2372825	30,7
22	20210324	1773662	22,9
23	20210329	1754860	22,7
24	20210423	1512342	19,5
25	20210425	1491054	19,3

Table S2b. *Sentinel-2 imagery and percentage of bare soil between July and November 2016-2021.*

	S2 Acquisition Date (year/month/day)	number of bare soil pixels	Bare soil pixels (%)
1	20160812	2139160	27,6
2	20160822	3210660	41,5
3	20170822	2792977	36,1
4	20170827	2307958	29,8
5	20170921	2234418	28,9
6	20171016	2995449	38,7
7	20180725	1748347	22,6
8	20180804	3418456	44,2
9	20180819	4029990	52,1
10	20180901	4685783	60,5
11	20180911	4677533	60,4
12	20180926	5116533	66,1
13	20181008	4610693	59,6
14	20181018	5145886	66,5
15	20181021	5118788	66,1
16	20190723	482202	6,2
17	20190725	563092	7,3
18	20190906	4188705	54,1
19	20190916	4202362	54,3
20	20190918	4538944	58,6
21	20190921	4241709	54,8
22	20200722	1267751	16,4
23	20200727	1943647	25,1
24	20200806	2701928	34,9
25	20200915	3334918	43,1
26	20200917	4894517	63,2
27	20200922	4992960	64,5
28	20210719	486412	6,3
29	20210722	588435	7,6
30	20210902	3869188	50,0
31	20210907	4044489	52,2
32	20211129	2101360	27,1

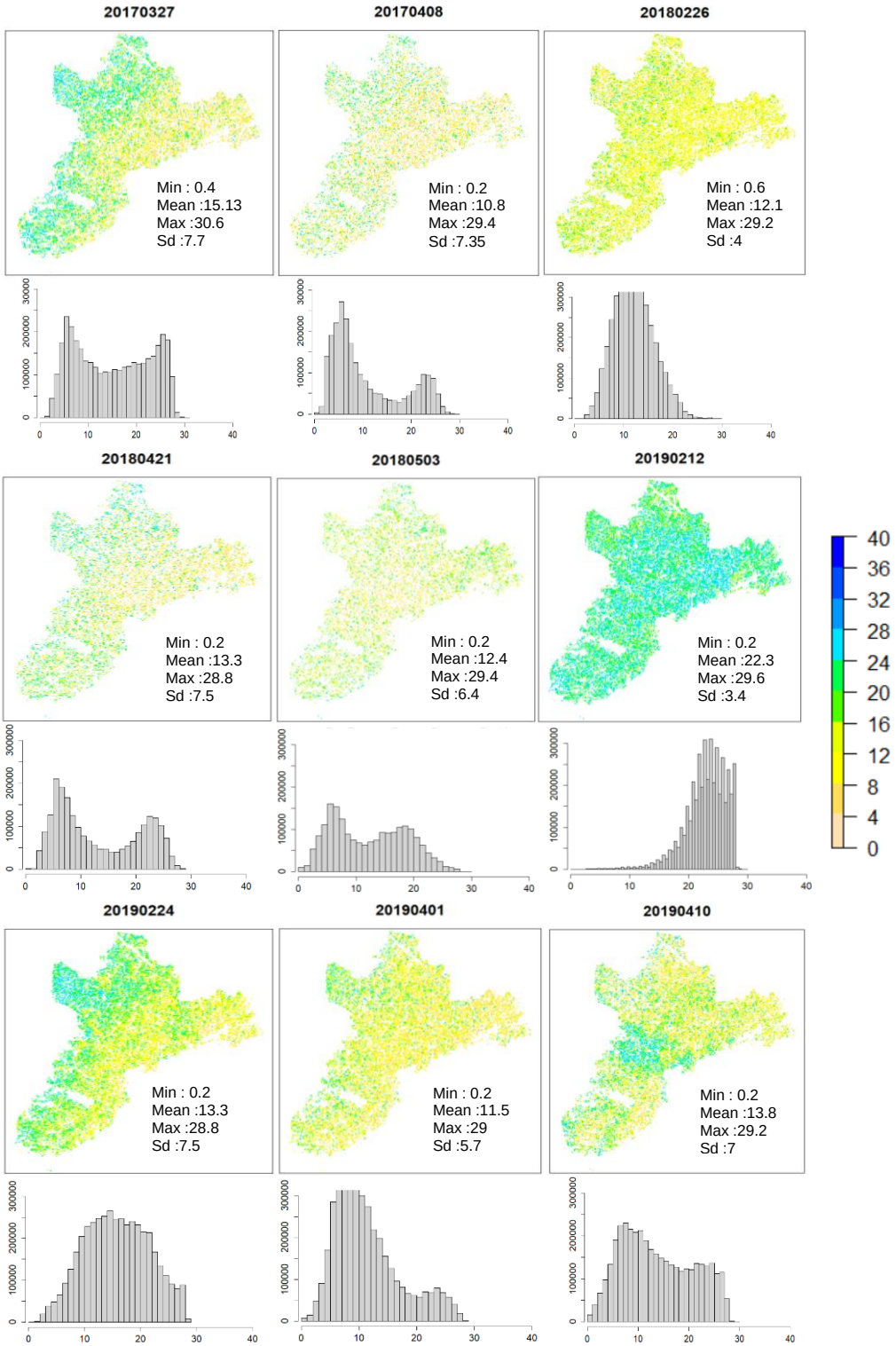


Figure S1a. *Soil moisture products and their respective histograms between February and May 2017-2021. (values are in vol.%)*

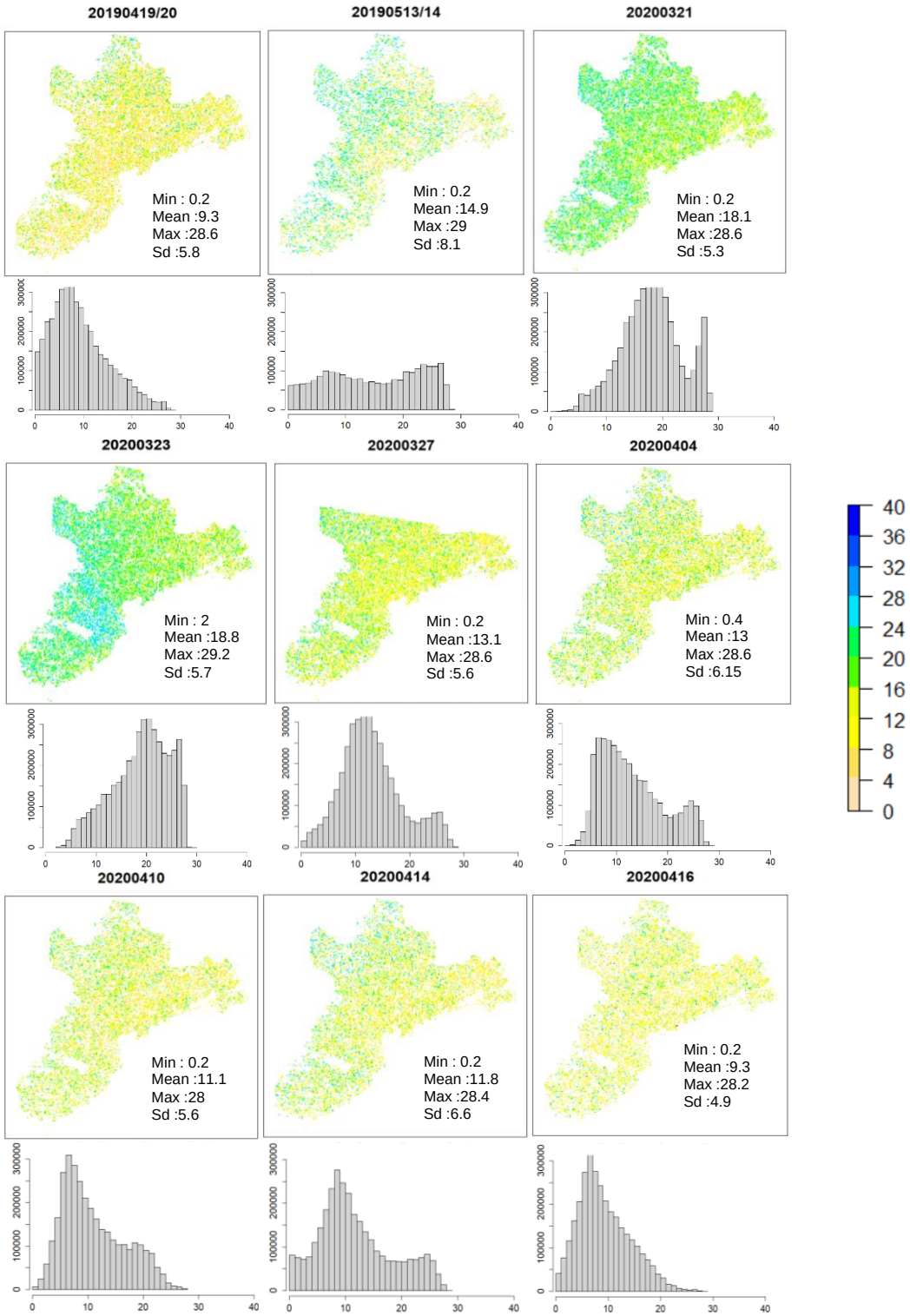


***Figure S1a**. continue*

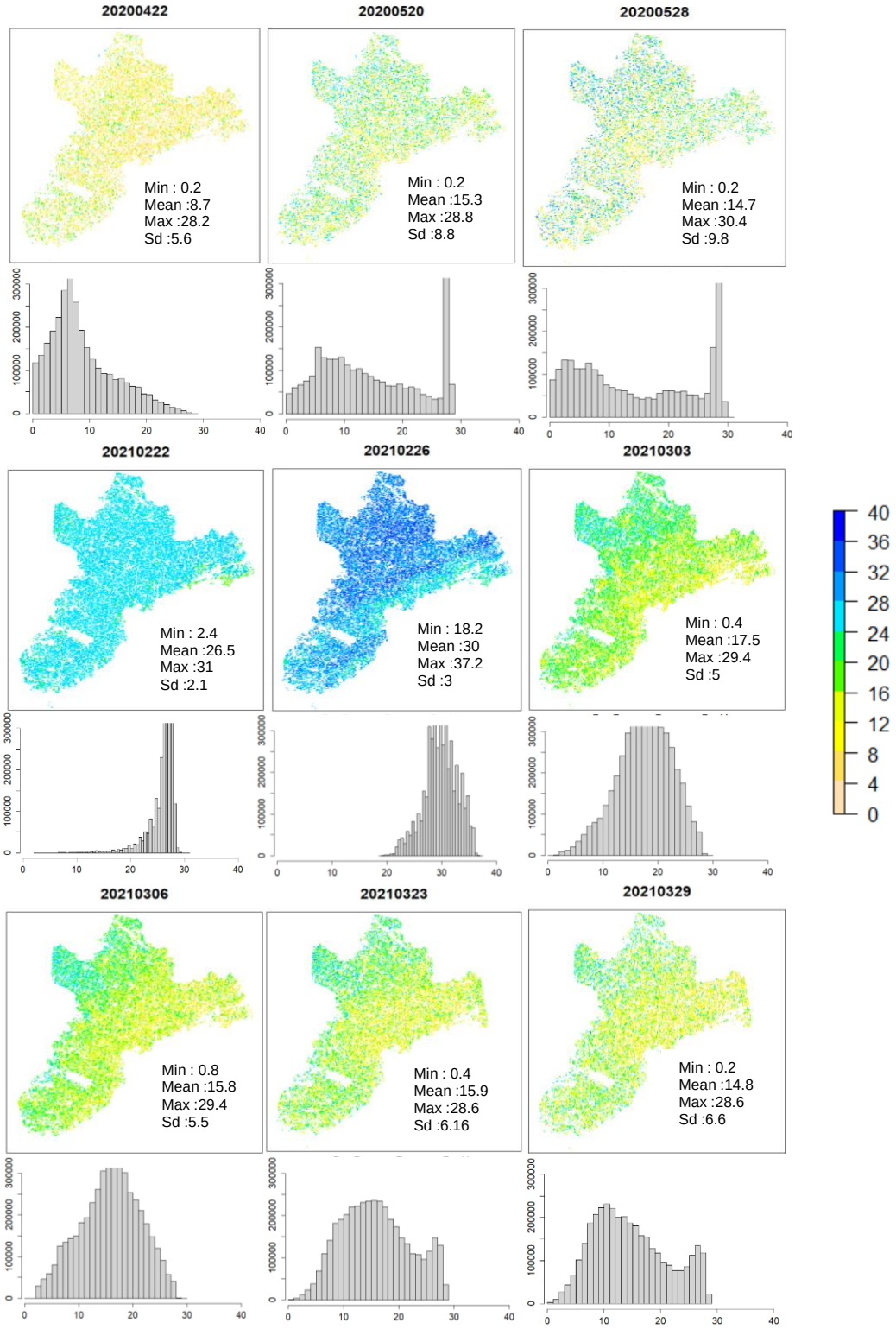


***Figure S1a**. continue*

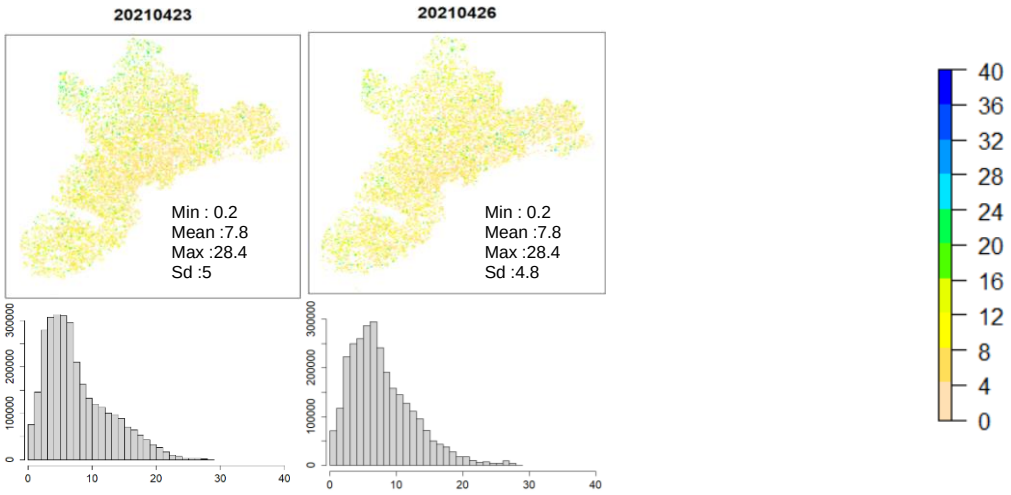


Figure S1a. *Soil moisture products and their respective histograms between February and May 2017-2021. (values are in vol.%)*

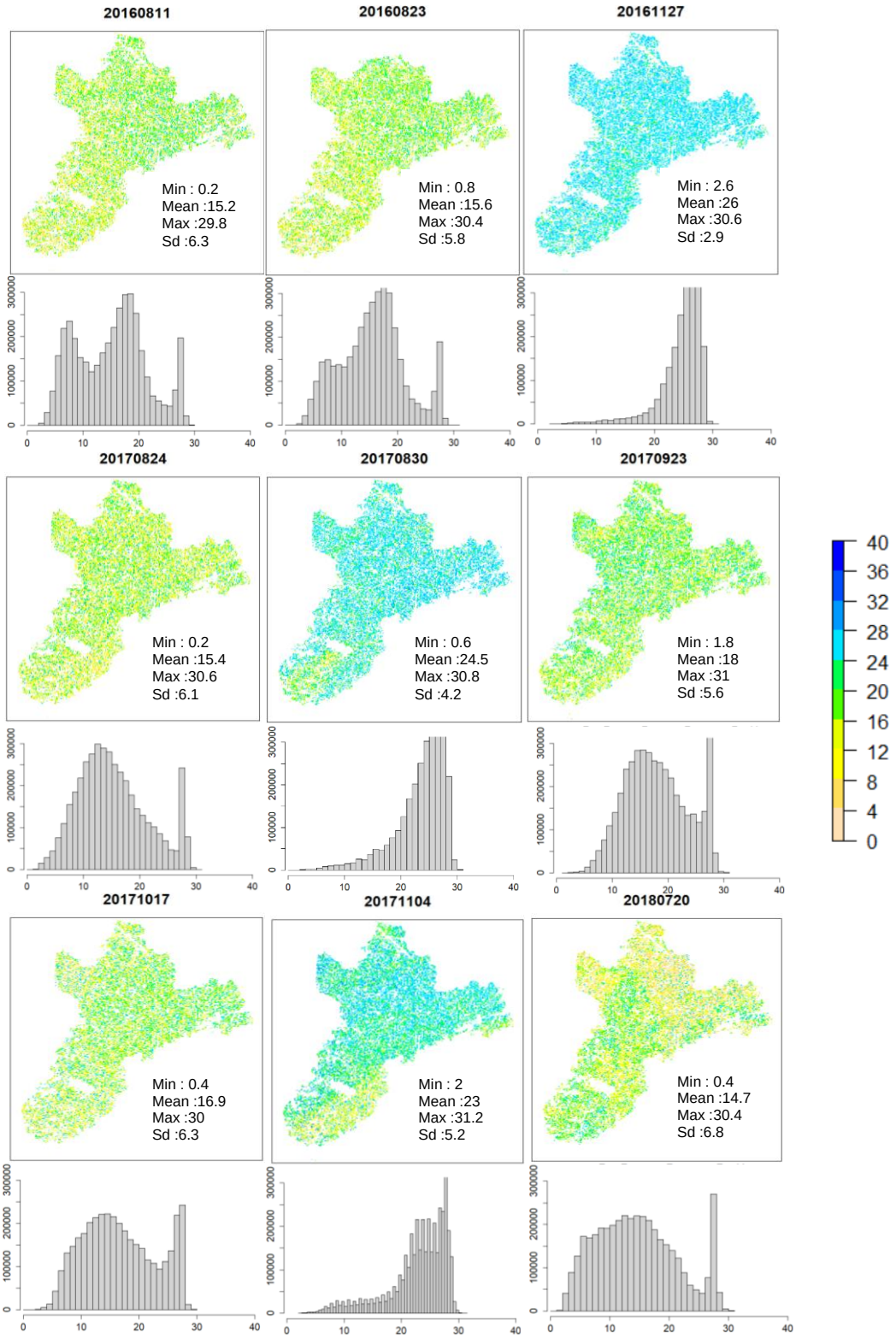


Figure S1b. *Soil moisture products and their respective histograms between July and November 2016-2021.(values are in vol.%)*

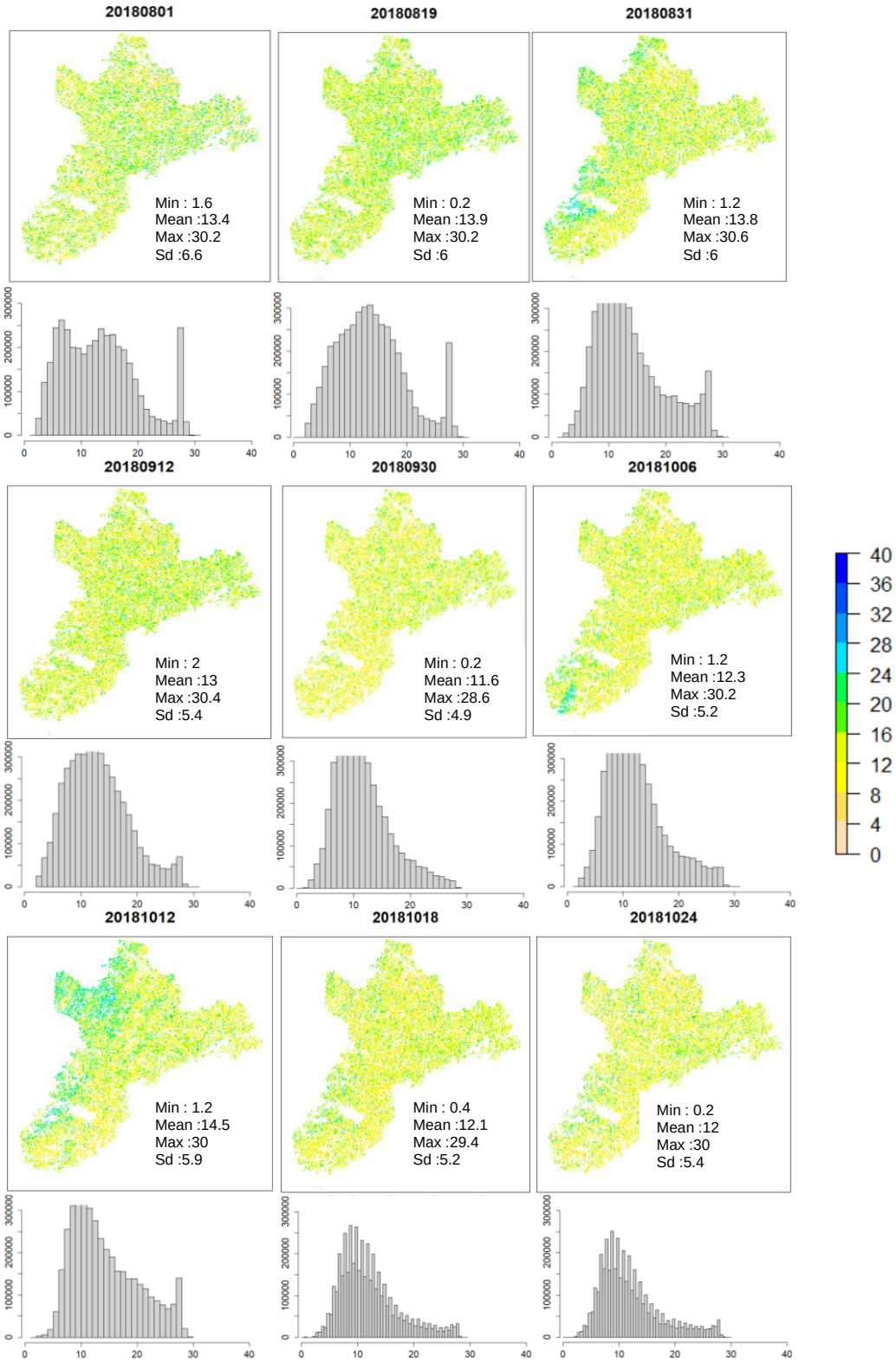


***Figure S1b**. continue*

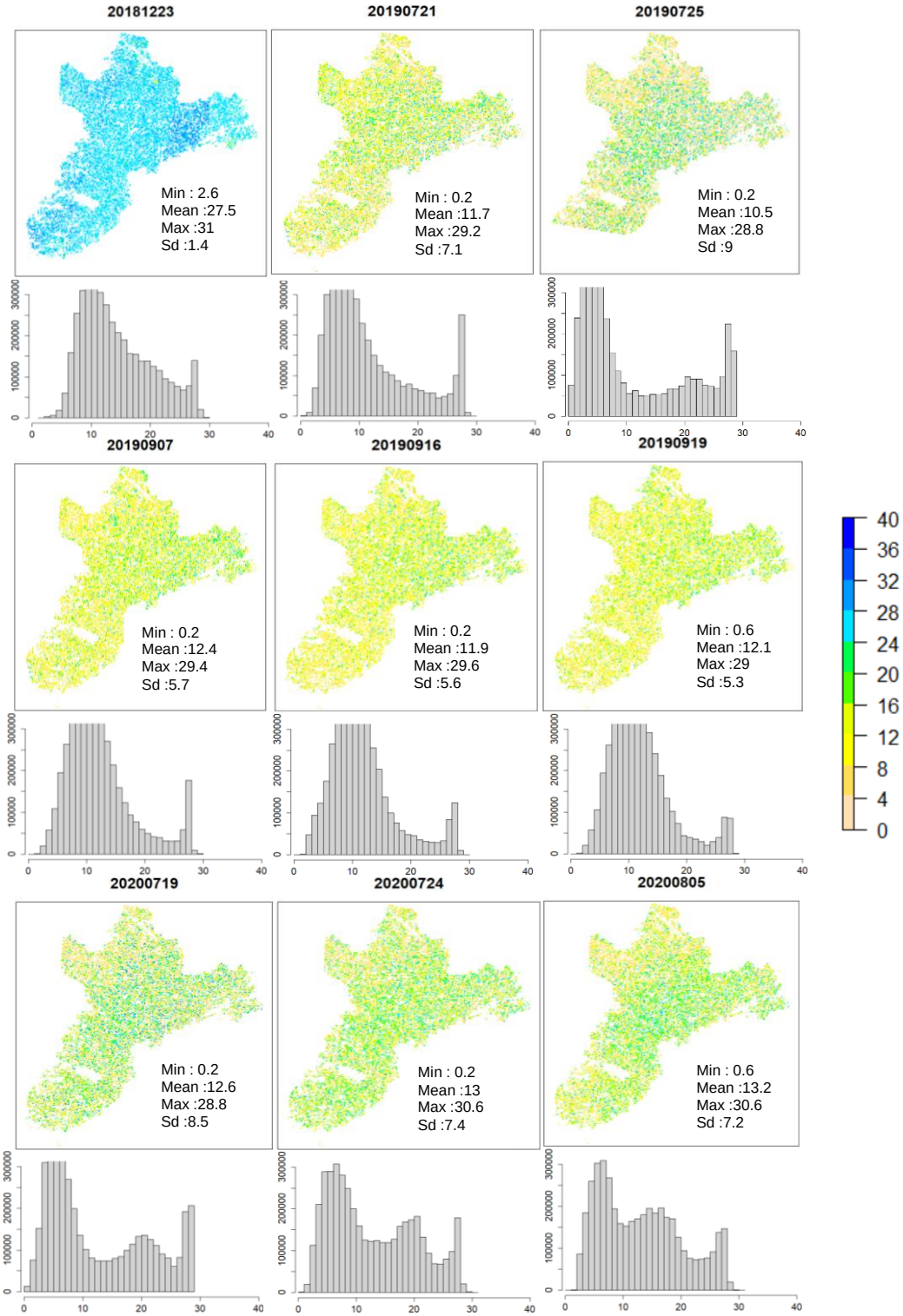


***Figure S1b**. continue*

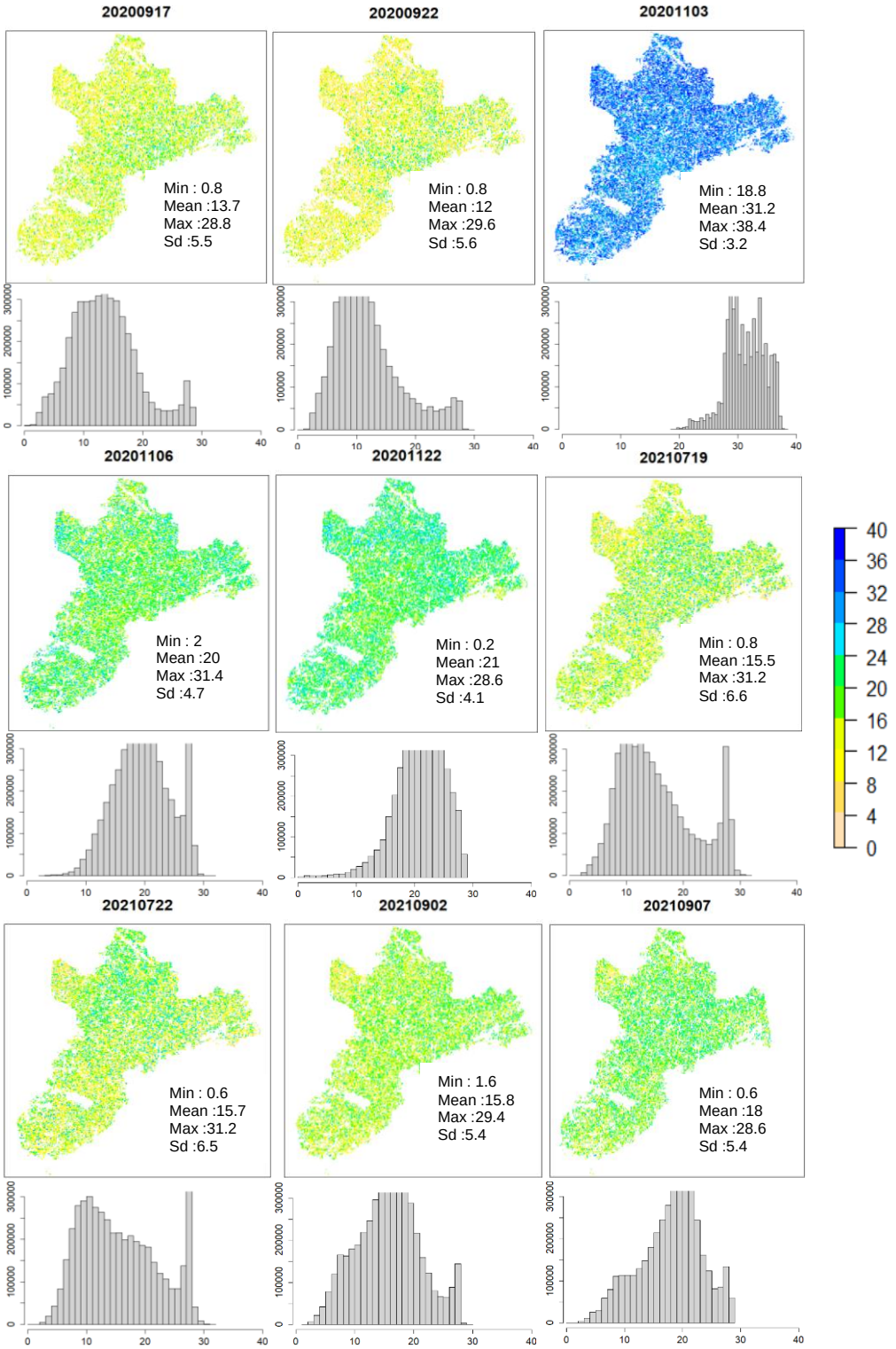


***Figure S1b**. continue*

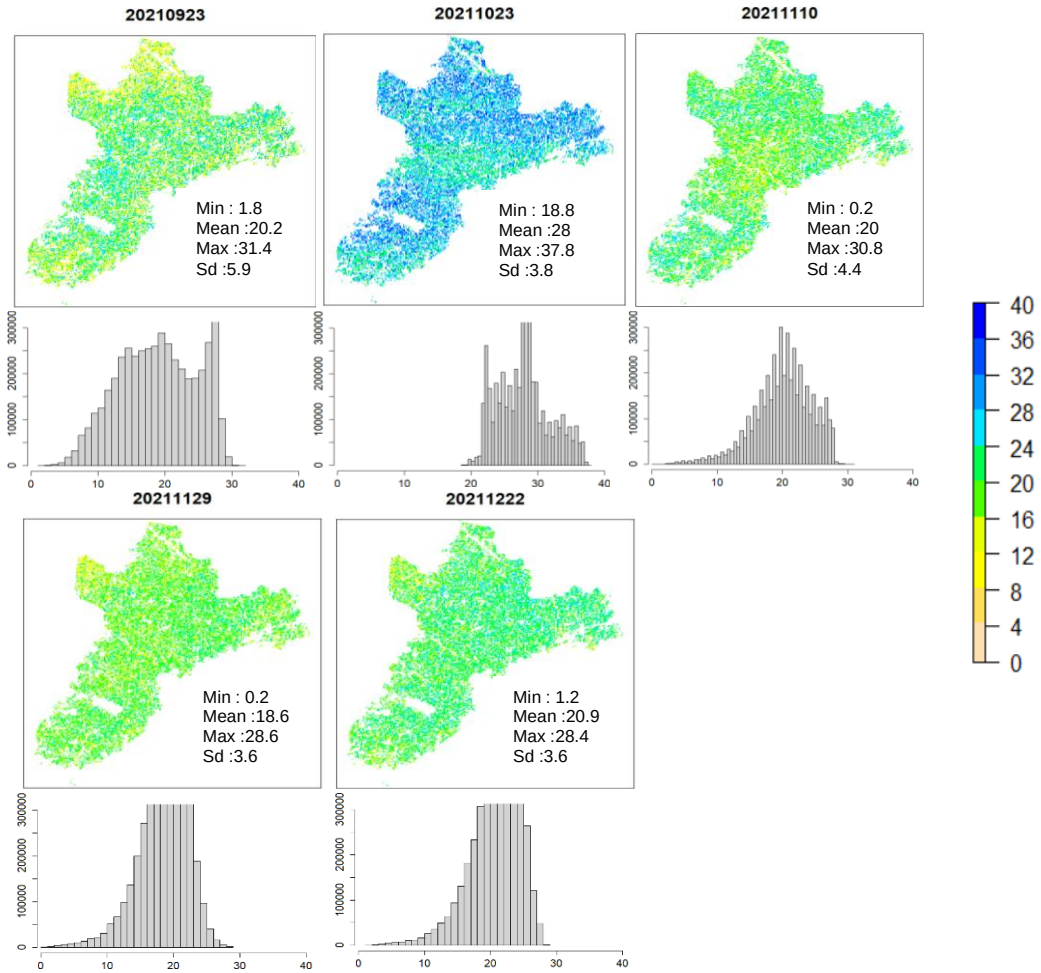


***Figure S1b**. Soil moisture products and their respective histograms between July and November 2016-2021.(values are in vol.%)*

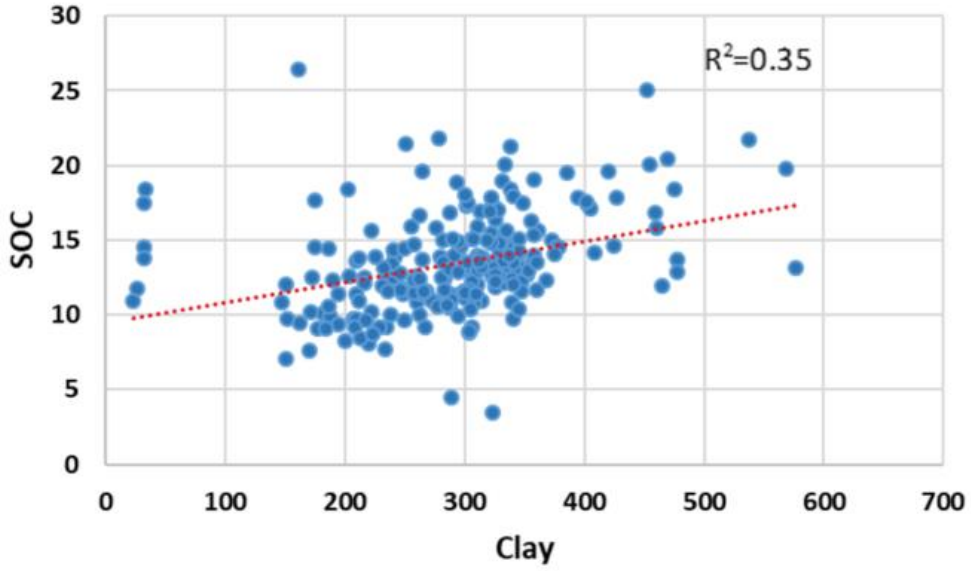


Figure S2. *Relationship between SOC and clay from sampling locations in Beauce*

RÉSUMÉ EN FRANÇAIS

La croissance rapide de la population entraîne des défis liés à la durabilité, aux modifications du climat et à la sécurité alimentaire. Les sols agricoles au sein d'une exploitation ou à l'échelon régional, national, et même mondial, constituent la matière première pour relever ces défis. Les sols agricoles représentent environ 15 à 20 % de la surface terrestre (Ramankutty et al., 2008). La bonne gestion de ces derniers est d'une grande importance pour les générations futures et la connaissance de la variabilité spatiale des différentes propriétés du sol à différents niveaux d'organisation spatiale revêt un intérêt majeur. Le carbone organique du sol (COS) est une composante essentielle qui peut contribuer à atténuer les émissions anthropiques de CO_2 atmosphérique et à assurer la sécurité alimentaire (Minasny et al., 2017 ; Arrouays and Horn, 2019), tout en restaurant la qualité de nos sols agricoles. Dans cette thèse, mon objectif était donc de déterminer le COS dans les sols agricoles de trois régions de la France métropolitaine (Bretagne, Occitanie et Centre Val de Loire) en utilisant principalement les images satellitaires Sentinel-2 et Sentinel-1. Ces régions ont été choisies en raison de leur représentativité au niveau national en termes de types de sols, de climat, de reliefs, de matériau parental, de types de cultures et de pratiques de gestion de l'utilisation des sols. Dans le ***chapitre 1***, nous présentons une vue d'ensemble de l'utilisation de l'imagerie satellitaire, et ses différentes résolutions spatiale et spectrale, pour détecter le COS dans le monde, ainsi que des séries satellitaires telle qu'au moyen de Landsat et de SPOT. Les principales méthodes de quantification du carbone prenant en compte les données spectrales issues des satellites, telles que les modèles de prédiction purement spectraux, les modèles incluant les données des banques de données spectrales (méthode "Bottom up") et les modèles de prédiction mixtes, sont mentionnées. Les chapitres suivants décriront, analyseront et discuteront plus en détail les possibles freins

à la détection du carbone liées aux effets atmosphériques, à la résolution spatiale et spectrale, à la disponibilité du sol nu et à la date d'acquisition des images, à la date d'analyse des échantillons de sol utilisés dans l'étalonnage du modèle, ainsi qu'à la composition de l'état de surface du sol, notamment l'humidité du sol, la rugosité du sol et les résidus végétaux. Le nombre d'études utilisant des données spectrales pour déterminer les propriétés du sol, tout comme celui des capteurs satellitaires, augmentent considérablement en science du sol. Dans ce premier chapitre, nous constatons que le nombre d'études visant à cartographier le COS et utilisant des méthodes purement spectrales dérivées de l'imagerie satellitaire est en augmentation en France par rapport au nombre total de travaux analysés au niveau national dans cette thèse. Au niveau international, Vaudour et al. (2022) ont rapporté soixante-deux études prédisant le COS à partir des données satellitaires depuis les années 1990 : la plupart de ces études ont été conduites dans de petites régions couvrant quelques centaines de km^2, tandis que peu d'études ont exploré la contribution des séries d'images satellitaires sur de grandes (< 10 000 km^2) et très grandes (> 10 000 km^2) régions. Le ***chapitre 2*** est dédié à la présentation des trois zones d'étude : Pleine-Fougères (10 km^2), Beauce (4838 km^2) et Piémont pyrénéen (22177 km^2). Il décrit l'ensemble des données pédologiques provenant de différentes bases de données (DoneSol, RMQS, LUCAS) ainsi que des campagnes d'échantillonnage des sols réalisées à Pleine-Fougères et dans le Piémont pyrénéen. Les approches prenant en compte les données spectrales d'images mono-date et les séries temporelles d'images de sols nus utilisées pour chacune des zones d'étude y sont ensuite présentées. Une analyse préliminaire a été réalisée en utilisant des modèles purement spectraux issus d'images Sentinel-2 mono-date à Pleine-Fougères. Le faible nombre d'échantillons de sol disponibles du fait de la superficie limitée de sol nu disponible aux dates d'acquisition des images a été le principal facteur limitant de l'étalonnage des modèles de prédiction. Le constat de ce verrou méthodologique a été fondamental afin de

structurer la séquence d'approches et de méthodes qui a été développée par la suite (tableau 2.3). Les résultats obtenus à Pleine-Fougères ont non seulement suggéré la nécessité de développer une méthode permettant d'étendre la surface de la couche arable nue afin de couvrir autant de points que possible disposant d'échantillons de sol, mais aussi mis en évidence l'intérêt de comparer les deux approches fondées sur une (mono-date) ou sur plusieurs dates combinées (mosaïque temporelle) en France dans différentes régions où ces approches n'ont pas été abordées. Le ***chapitre 3*** est focalisé sur le Piémont pyrénéen, commence par une méthode purement spectrale puis développe une méthode mixte utilisant les produits d'humidité du sol (SMPs) comme covariable. Tous les modèles ont montré des variations dans les performances de prédiction, principalement liées à la date d'acquisition des images Sentinel-2, à la texture du sol, à la période d'échantillonnage du sol, à l'humidité du sol et aux unités topographiques. La date de l'image et l'humidité du sol étaient les facteurs les plus évidents de la précision du modèle. Cependant, nos résultats suggèrent que les approches d'imagerie mono-date devraient être davantage explorées en considérant non seulement les caractéristiques de l'image satellitaire mais encore celles de la zone d'étude, des échantillons utilisés (représentativité spatiale et antériorité de collecte) et des états de surface du sol. Nous montrons en outre que la distribution spatiale du COS pourrait être liée à certaines situations topographiques et à l'utilisation historique des sols. Enfin, les images Sentinel-2 nous ont permis d'identifier des valeurs élevées de carbone sur des unités de paysage/sol spécifiques à cette région. Le ***chapitre 4*** a visé à augmenter le nombre d'échantillons en sols nus pour les modèles de prédiction, et développé à cette fin une méthode de mosaïquage temporel des sols nus s'appuyant sur les séries temporelles Sentinel-2 (S2Bsoil) dans la région de la Beauce. Dans ce chapitre, nous avons évalué la performance du modèle en utilisant S2Bsoil i) sur une série temporelle complète de 57 images entre 2016 et 2021 ; ii) sur une série temporelle

partielle de 25 images acquises de la fin de l'hiver au printemps ; iii) sur une série temporelle partielle de 32 images acquises de l'été à l'automne. Le SMPs, l'imagerie gamma aéroportée et plusieurs covariables morphométriques (issues d'un modèle numérique d'altitude et de ses dérivées) et lithologiques ont été utilisés dans la modélisation. Les résultats de ce chapitre suggèrent l'importance de sélectionner une période d'acquisition appropriée pour construire des mosaïques temporelles de sols nus : cela tient non seulement à la variabilité que le COS peut présenter dans une région donnée, mais encore aux facteurs perturbants de la prédiction du COS liés à la date d'acquisition de l'image et à l'état de la surface du sol. Les résultats obtenus dans cette région d'étude ont également montré l'importance, dans la détection du carbone au niveau régional, de la relation entre la variabilité spatiale du carbone et les covariables telles que les images gamma du thorium, les coordonnées obliques et celles calculées à partir des informations topographiques voisines, d'autant que cette région d'agriculture intense est caractérisée par des formations superficielles éoliennes de lœss, organisée selon des directions spécifiques. Enfin, le ***chapitre 5*** porte sur un certain nombre de questions essentielles de ce travail où nous discutons des conclusions tirées de cette thèse et soulignons les principales perspectives à considérer pour la télédétection du carbone du sol via l'imagerie satellitaire dans un avenir proche.

Mémoires du campus de Grignon

Domaine où a été fondée l'Ecole Nationale Supérieure Agronomique de Grignon, la plus ancienne de France. Fondé en 1826 tout près de Versailles avec une grande superficie de terres agricoles, des cultures et une variation de différents types de sols. Avec un château construit en 1636 (celui des photos), C'était l'alma mater des étudiants passionnés pela agronomie et de sciences connexes et où j'ai eu l'opportunité de travailler et de vivre pendant presque toute ma thèse. Je suis tellement reconnaissant pour chaque expérience et chaque apprentissage dans cet endroit. Les photos que j'ai prises ne sont qu'un simple hommage à ce lieu qui, même s'il n'est plus en activité, est un symbole de l'agriculture et de la recherche en France.

Printemps en mai 2020

L'hiver en février 2021

L'automne en novembre 2020