

Article

Radiative Transfer Model-Integrated Approach for Hyperspectral Simulation of Mixed Soil-Vegetation Scenarios and Soil Organic Carbon Estimation

Asmaa Abdelbaki ^{1,2,*}, Robert Milewski ¹, Mohammadmehdi Saberioon ¹, Katja Berger ¹, José A. M. Demattê ³ and Sabine Chabrillat ^{1,4}

¹ GFZ Helmholtz Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany; milewski@gfz.de (R.M.); saberioon@gfz.de (M.S.); katja.berger@gfz.de (K.B.); chabri@gfz.de (S.C.)

² Soils and Water Department, Faculty of Agriculture, Fayoum University, Fayoum 63514, Egypt

³ Department of Soil Science, Luiz de Queiroz College of Agriculture, University of São Paulo (ESALQ/USP), Av. Pádua Dias 11, CP9, Piracicaba 13418-900, SP, Brazil; jamdemat@usp.br

⁴ Institute of Earth System Science, Leibniz University Hannover, Herrenhäuser Straße 2, 30419 Hannover, Germany

* Correspondence: asmaa@gfz.de or ama09@fayoum.edu.eg

Abstract

Soils serve as critical carbon reservoirs, playing an essential role in climate change mitigation and agricultural sustainability. Accurate soil property determination relies on soil spectral reflectance data from Earth observation (EO), but current vegetation models often oversimplify soil conditions. This study introduces a novel approach that combines radiative transfer models (RTMs) with open-access soil spectral libraries to address this challenge. Focusing on conditions of low soil moisture content (SMC), photosynthetic vegetation (PV), and non-photosynthetic vegetation (NPV), the coupled Marmit-Leaf-Canopy (MLC) model is used to simulate early crop growth stages. The MLC model, which integrates MARMIT and PRO4SAIL2, enables the generation of mixed soil-vegetation scenarios. A simulated EO disturbed soil spectral library (DSSL) was created, significantly expanding the EU LUCAS cropland soil spectral library. A 1D convolutional neural network (1D-CNN) was trained on this database to predict Soil Organic Carbon (SOC) content. The results demonstrated relatively high SOC prediction accuracy compared to previous approaches that rely only on RTMs and/or machine learning approaches. Incorporating soil moisture content significantly improved performance over bare soil alone, yielding an R^2 of 0.86 and RMSE of 4.05 g/kg, compared to R^2 = 0.71 and RMSE = 6.01 g/kg for bare soil. Adding PV slightly reduced accuracy (R^2 = 0.71, RMSE = 6.31 g/kg), while the inclusion of NPV alongside moisture led to modest improvement (R^2 = 0.74, RMSE = 5.84 g/kg). The most comprehensive model, incorporating bare soil, SMC, PV, and NPV, achieved a balanced performance (R^2 = 0.76, RMSE = 5.49 g/kg), highlighting the importance of accounting for all surface components in SOC estimation. While further validation with additional scenarios and SOC prediction methods is needed, these findings demonstrate, for the first time, using radiative-transfer simulations of mixed vegetation-soil-water environments, that an EO-DSSL approach enhances machine learning-based SOC modeling from EO data, improving SOC mapping accuracy. This innovative framework could significantly improve global-scale SOC predictions, supporting the design of next-generation EO products for more accurate carbon monitoring.



Academic Editor: Konstantinos X. Soulis

Received: 18 May 2025

Revised: 20 June 2025

Accepted: 2 July 2025

Published: 9 July 2025

Citation: Abdelbaki, A.; Milewski, R.; Saberioon, M.; Berger, K.; Demattê, J.A.M.; Chabrillat, S. Radiative Transfer Model-Integrated Approach for Hyperspectral Simulation of Mixed Soil-Vegetation Scenarios and Soil Organic Carbon Estimation. *Remote Sens.* **2025**, *17*, 2355. <https://doi.org/10.3390/rs17142355>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: soil spectroscopy; MARMIT; PROSPECT4SAIL2; convolutional neural network (CNN); hybrid inversion model; LUCAS-SSL data

1. Introduction

Accurate estimation of soil organic carbon (SOC) content is essential for understanding ecosystem functions and supporting sustainable land management, especially given the significant potential of soils to restore large quantities of carbon, thereby contributing to the mitigation of climate change [1,2].

Soil spectral reflectance also serves as a critical baseline for interpreting vegetation signals [3]. Yet, its inherent variability, influenced by factors such as moisture and mineral composition, can significantly impact the accuracy of vegetation properties assessments [4]. Among the key soil properties affecting this reflectance, SOC stands out due to its profound influence on soil structure, fertility, and microbial activity [5]. Understanding the interplay between SOC and the spectral signature is further complicated by the presence of both photosynthetic vegetation (PV), actively absorbing light for growth, and non-photosynthetic vegetation (NPV), contributing organic matter to the soil [6]. Therefore, accurately disentangling the spectral contributions of soil, PV, and NPV is essential for reliable SOC estimation and for developing effective strategies for soil health management and carbon sequestration [7].

The significant spatial variability of soils presents a major challenge to reliable SOC assessment, as it impacts both SOC distribution and the spectral reflectance of vegetated surfaces [8]. Two primary approaches exist for estimating SOC using remote sensing data (Verrelst et al 2015): (1) Statistical methods: These rely on empirical relationships or machine learning to correlate spectral features with target variables. While effective, they often face limitations in generalizability across diverse conditions and require extensive ground-truth data. (2) Physically based methods: These employ radiative transfer models (RTMs) to simulate light interactions within vegetation canopies and soil layers, offering a more mechanistic understanding but often requiring complex parameterization [9]. The effectiveness of RTMs depends on accurate soil reflectance data to separate canopy and background signals, improving the retrieval of biophysical parameters [10,11]. Recently, hybrid approaches combining the physical realism of RTMs with the flexibility and scalability of statistical techniques have gained traction, demonstrating significant potential for improving predictive accuracy and computational efficiency in large-scale applications [12–15]. Advanced machine learning algorithms, such as deep neural networks, have been increasingly employed to accelerate predictions by learning from RTM-generated datasets, enabling rapid application across global, regional, and local scales. For example, Kattenborn et al. [16] used convolutional neural networks (CNNs) trained on RTM-generated datasets to estimate vegetation traits globally, while Zhang et al. [17] applied deep learning models for leaf biochemical property retrieval at the regional level.

A promising approach for monitoring vegetation and soil properties across scales is the use of full-range hyperspectral remote sensing data, which provides detailed spectral information for characterizing plant and soil traits [18,19]. Spaceborne missions such as EnMAP [20,21] and PRISMA [22] have demonstrated the potential of hyperspectral imaging; however, their limited temporal coverage and tasking constraints restrict their use for consistent, broad-scale applications. In contrast, upcoming missions like CHIME [23] and SBG [24], designed for routine global observations, are expected to deliver consistent, high-resolution hyperspectral data. Robust modeling frameworks will be essential to fully leverage these capabilities [25,26]. Radiative Transfer Models (RTMs) simulate solar radiation interactions with vegetation and soil surfaces, offering a non-invasive, cost-effective alternative to field-based assessments. The fine spectral resolution of hyperspectral data enables discrimination between bare soil, photosynthetic vegetation (PV), and non-photosynthetic vegetation (NPV), which is critical for accurate soil organic carbon (SOC) estimation. Combining VNIR and SWIR bands enhances model performance, with

SWIR wavelengths being particularly sensitive to SOC [27,28]. Additionally, hyperspectral imagery supports the separation of PV and NPV components, which is vital for semi-bare soil modeling and for improving algorithms such as HYSOMA and ENSOMAP [29,30].

Vegetation RTMs, such as PROSAIL [31], SLC [32], SCOPE [33], FLIGHT [34], and DART [35], are widely used to simulate the interaction between vegetation and remote sensing data. While advanced 3D models like DART and FLIGHT are capable of incorporating detailed soil and surface parameterizations, they are typically applied in vegetation-focused contexts and require extensive input data, making their operational use for soil property retrievals, such as SOC, less common. Moreover, many RTMs still rely on simplified soil representations or assume homogeneous backgrounds, which limits their ability to capture spatial variability in SOC, texture, moisture, and vegetation cover. Soil-specific RTMs, including Hapke-based models [36] and SOILSPECT [31], have improved soil reflectance modeling, while others like BSM [37] and Kubelka–Munk-based models [38] enhance moisture estimation. The MARMIT model [39] advances the simulation of soil moisture for dry, measured soils and is further extended by accounting for particle–water interactions, improving its applicability under varying soil moisture conditions. Nonetheless, most soil RTMs have been validated under controlled conditions and focus on single-variable estimation, such as moisture. A key challenge remains their limited capacity to simulate the spectral complexity of mixed pixels, where bare soil, PV, and NPV coexist, thus constraining accurate SOC retrieval. Addressing this limitation is critical for leveraging upcoming spaceborne hyperspectral missions for global soil monitoring.

Therefore, this study aims to analyze how surface disturbances (e.g., soil moisture, green/dry crop residues) influence soil reflectance and SOC prediction accuracy. Hence, the study is based on the following hypothesis: H1: Incorporating surface disturbances (moisture, residues) improves SOC prediction accuracy. H0 (baseline): Models calibrated only on dry (without disturbances) yield inferior predictions.

2. Materials and Methods

2.1. Overview of the Marmit–Leaf–Canopy Model Structure

The Marmit–Leaf–Canopy (MLC) RTM is designed to construct the Disturbed Soil Spectral Library (DSSL), an EO-simulated soil spectral database incorporating SMC, PV, and NPV coverage. It consists of three key sub-modules: MARMIT for soil moisture, PROSPECT4 for leaf optics, and 4SAIL2 for canopy architecture, collectively forming the PRO4SAIL2 module within the SLC model [32], as shown in Figure 1. These modules are integrated with the Hapke-based soil BRDF model to simulate hyperspectral images and retrieve plant biophysical and biochemical variables, with an emphasis on the role of soil background in enhancing canopy retrieval accuracy [40]. The original Hapke model, validated with GER-SIRIS spectrometer measurements, is replaced by the multilayer radiative transfer model (MARMIT) to more accurately simulate the effects of SMC on soil reflectance due to limitations in its uniform surface assumption [39]. MARMIT is parameterized using spectral data from a wetting experiment conducted on different agricultural soils in Europe. The PROSPECT4 and 4SAIL2 modules generate top-of-canopy reflectance spectra between 400 and 2500 nm for vegetation with green, brown, or mixed leaf types under varying soil conditions. The vegetation types are defined separately, with LAI quantified independently for green and brown vegetation (Table 1). A total of 822,572 DSSL simulations are generated by varying SMC, PV, and NPV. These simulations are categorized into four scenarios for evaluating SOC prediction, including the measured bare soil (Table 2). Descriptions of each module are provided below.

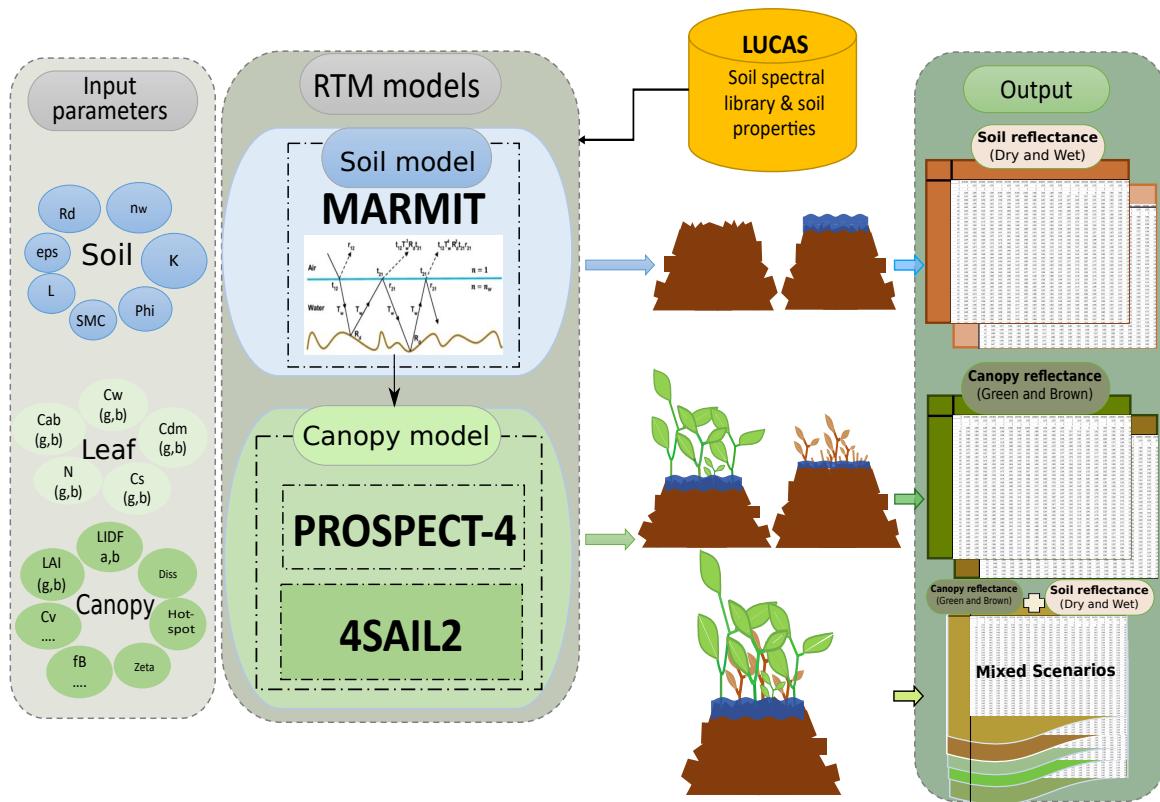


Figure 1. Conceptual framework of the developed MLC model for simulating soil and canopy reflectance under varying conditions.

2.1.1. MARMIT Model

The Multilayer Radiative Transfer model (MARMIT) of the soil reflectance model (original version) is well-suited for this study due to its simplicity, efficiency, and low computational demand. This ‘equivalent slab’ radiative transfer model, rooted in Ångström [41]’s ‘wet soil darkening’ concept, was refined to include spectral reflectance and water’s refractive index [42]. However, this model assumption of negligible water absorption in the visible and near-infrared (VIS-NIR) spectral regions limited its applicability to the shortwave infrared (SWIR), where water exhibits strong absorption bands. Bach and Mauser addressed this limitation in 1994 by introducing water absorption into the reflectance model using the Beer–Lambert law [43].

The MARMIT model represents a further advancement, explicitly accounting for light transmittance across the water–air interface. The model calculates total wet soil reflectance (R_{ws}) as a geometric series of multiple reflections and refractions at the surface:

$$R_{ws} = r_{12} + t_{12} \cdot r_{21} \cdot T_w \cdot R_d / (1 - r_{21} \cdot R_d \cdot T_w), \quad (1)$$

where r_{12} and t_{12} are Fresnel reflection and transmission coefficients at the air–water interface, r_{21} and t_{21} are the corresponding coefficients at the water–air interface, R_d is dry soil reflectance, and T_w is water transmittance calculated using the Beer–Lambert law:

$$T_w = \exp(-\alpha B \cdot L), \quad (2)$$

where αB is the water absorption coefficient and L is the equivalent water layer thickness. The model employs average Fresnel coefficients for unpolarized light.

To simulate mixed wet and dry soil conditions, MARMIT introduces an efficiency parameter (ε) representing the wet soil fraction. The resulting soil surface reflectance (R_{mod}) is a linear combination of wet and dry soil reflectances:

$$R_{\text{mod}} = \varepsilon \cdot R_{\text{ws}} + (1 - \varepsilon) \cdot R_{\text{d}} \quad (3)$$

2.1.2. PROSPECT4 Model

The Leaf Optical Properties Spectra model (PROSPECT-4), an extension of the original PROSPECT model [44], simulates leaf optical properties [45]. The PROSPECT model was initially developed based on the assumption that leaves behave as homogeneous plates with multiple scattering and absorption processes, primarily accounting for chlorophyll content and leaf structure. PROSPECT-4 improves upon this by introducing water content (C_w) as an additional input variable, enhancing its capability to analyze water stress and plant physiology. The PROSPECT-4 model simulates directional-hemispherical reflectance and transmittance for a single leaf, with input variables including the leaf structure parameter (N) and leaf biochemical constituents such as leaf chlorophyll content (LCC), leaf dry matter content (Cm), leaf water content (Cw), and leaf senescent matter content (Cs). PROSPECT-4 further differs by combining chlorophyll and carotenoid variables and incorporating brown leaves to represent leaf senescence, enabling the simulation of reflectance for both green and brown pigments. This version, integrated into the MLC model, is widely used to simulate leaf reflectance under various physiological conditions.

2.1.3. 4SAIL2 Model

Building upon the Scattering by Arbitrary Inclined Leaves (SAIL) canopy model [46], 4SAIL2 simulates canopy reflectance (400–2500 nm) for various sun–target–sensor geometries by incorporating leaf properties from PROSPECT4 [32]. Unlike its predecessor, it employs a 4-stream method for improved accuracy, making it suitable for diverse canopies like row crops and forests. 4SAIL2 introduces a double-layer model for green and brown leaves to improve reflectance simulations for heterogeneous canopies. Although the structural properties are assumed to be identical, their LAIs can differ, with the leaf angle distribution (LIDF) described by parameters a and b . The green and brown leaf LAI division is controlled by the fraction of brown leaves (f_B) and the dissociation factor (D), where $D = 1$ indicates complete separation and $D = 0$ indicates a homogeneous mixture. The tree shape factor (Zeta) is based on crown diameter to height ratio, with crop height and crown diameter values.

The vertical projection of vegetation elements (f_{Cover}) can be estimated from the SLC model, using the gap fraction theory and considering LAI, LIDF parameters, and crown coverage as shown in the equation below:

$$f_{\text{Cover}} = C_v * \left(1 - e^{-k * \text{LAI}}\right), \quad (4)$$

where k is the extinction coefficient in the vertical direction, and C_v is the vertically projected crown cover fraction.

Table 1. Overview of the model input variables of look-up tables (LUTs) used for generating wet soil and canopy spectra for green and brown.

| Variable | Unit | PV | NPV | Source of Information |
|--|----------------------------|--------------------|--------------|---------------------------|
| Soil Variables (MARMIT) | | | | |
| Reflectance bare soil (R_b) | Unitless | | LUCAS data | Jones et al. [47] |
| Refractive index of water (n_w) | Unitless | | - | Bablett et al. [39] |
| Specific absorption coefficient of water (K) | cm^{-1} | | - | Bablett et al. [39] |
| Steepness of curve (ψ) | Unitless | | - | Bablett et al. [39] |
| Wet soil surface ratio (ϵ) | Unitless | | - | MARMIT model |
| Thickness of water layer (L) | cm | | - | MARMIT model |
| Soil moisture content (SMC_g) | Unitless | 0.015, 0.035, 0.07 | | Prior knowledge |
| Leaf Variables (PROSPECT-4) | | | | |
| Internal leaf structure (N) | Unitless | | 1.5 | Kooistra and Clevers [48] |
| Leaf chlorophyll content (LCC) | $\mu\text{g cm}^{-2}$ | 80 | 0 | Prior knowledge |
| Water content (C_w) | cm | 0.0317 | 0.001 | Kooistra and Clevers [48] |
| Dry matter content (C_m) | g cm^{-2} | 0.005 | 0.02 | Botha et al. [49] |
| Senescent material (C_s) | Unitless | 0 | 1 | Wang et al. [50] |
| Canopy Variables (4SAIL2) | | | | |
| Leaf area index (LAI) | $\text{m}^2 \text{m}^{-2}$ | | 0.05 to 1 | Prior knowledge |
| Leaf inclination distribution (LIDFa/b) | Unitless | | 1 (a), 0 (b) | Wang et al. [50] |
| Hotspot coefficient (hot) | m m^{-1} | | 0.05 | Casa and Jones [51] |
| Vertical crown cover (C_v) | Unitless | | 1 | Prior knowledge |
| Tree shape factor (ζ) | Unitless | 0.3 | 0 | Abdelbaki et al. [52] |
| Layer dissociation factor (D) | Unitless | | 1 | Prior knowledge |
| Fraction of brown vegetation (f_b) | Unitless | 0 | 1 | Prior knowledge |
| Solar zenith angle (θ_s) | Degree | | 35 | Abdelbaki et al. [52] |
| Viewing zenith angle (θ_o) | Degree | | 0 | Abdelbaki et al. [52] |
| Relative azimuth angle (ψ) | Degree | | 0 | Abdelbaki et al. [52] |

Table 2. Summary of MLC variables, scenarios, and resulting sample sizes.

| ID | Variables | Modeling Sample Size |
|--|---|----------------------|
| Scenario 1 | | |
| Bare soils | - | 8941 |
| Scenario 2 | | |
| (Bare Soils + Bare Soils \times SMC) | SMC: 0.015, 0.035, 0.07 | 35,764 |
| Scenario 3 | | |
| (Bare Soils + Bare Soils \times SMC \times LAI-PV) | LAI: 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 | 393,404 |
| Scenario 4 | | |
| (Bare Soils + Bare Soils \times SMC \times LAI-NPV) | LAI: 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1 | 393,404 |
| Scenario 5—DSSL | | |
| (Bare Soils + Bare Soils \times 3SMC + Bare Soils \times 3SMC \times 11PV + Bare Soils \times 3SMC \times 11NPV) | All previous variables | 822,572 |

2.2. LUCAS Database Description

The Land Use/Land Cover Area Frame Survey (LUCAS) topsoil database is a comprehensive resource of soil information across Europe, supporting research in soil science, geochemistry, biology, and ecology. Freely accessible through the European Soil Data Cen-

ter (ESDAC), it also informs environmental and agricultural policy decisions [47]. Initiated between 2009 and 2012, the database covered 23 EU Member States and expanded to all 28 EU regions and areas above 1000 m altitude. The dataset includes 21,859 soil samples collected to a depth of 20 cm, with 540 points dedicated to soil biodiversity analysis across croplands, woodlands, shrublands, wetlands, bare land, and artificial land. However, potential sampling biases may affect the spatial variability representation. Each sample was analyzed for chemical and physical properties, including soil organic carbon (SOC) content, quantified via dry combustion at 900 °C, and complemented with spectral measurements (400–2500 nm, 2 nm resolution).

This research focuses on the 8941 cropland records, representing a wide range of soil taxonomic classes, horizons, and textures across European countries (Figure 2). The SOC content of the cropland points averages 17.6 g/kg (median = 14.3 g/kg), ranging from 0.10 g/kg to 519.10 g/kg. This soil spectral library (SSL) of the LUCAS dataset is used as input and background for the proposed model to simulate SMC, PV, and NPV.

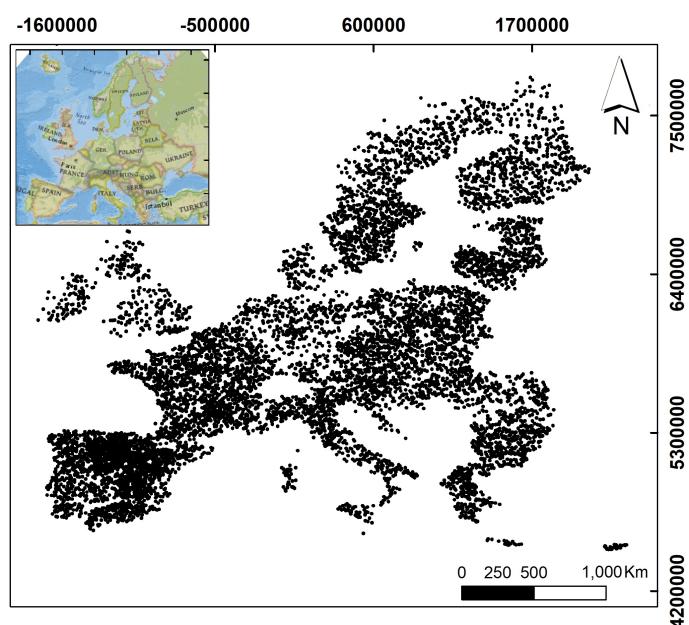


Figure 2. Geographic distribution of soil samples in the LUCAS Soil Spectral Library.

2.3. Deep Learning Spectral Modeling for SOC Estimation

Following the integration of the MLC model, the hybrid model implementation involved two key steps. First, the MLC-RTM model was employed as a forward model to simulate reflectance from soil and canopy. Second, the resulting simulated reflectance data, along with associated SOC values, served as input for the deep learning algorithm. A convolutional neural network (CNN), a prominent machine learning architecture, is utilized effectively in the prediction of SOC [53,54].

2.3.1. 1D-CNN Model Architecture

Convolutional neural networks (CNNs) are deep learning models with one or more convolutional layers. A typical CNN includes an input layer, multiple hidden layers (convolutional, pooling, and fully connected layers), and an output layer. This study uses a CNN architecture based on [54]. The input layer receives 2100 bands (400–2499 nm, 1 nm intervals) in a 2D format. Three convolutional blocks, each with increasing filters (32, 64, 128) and followed by max-pooling, extract features. These features are passed through two fully connected layers with ReLU activation and a dropout layer to reduce overfitting. The

output layer contains a single neuron for regression, and the loss function is computed for training, as shown in Table 3. The model was implemented in MATLAB R2022b on Windows (GPU) with 64 GB of memory.

Table 3. Description of the CNN network architecture along with the various layers.

| Layer Type | Filters | Kernel Size | Width | Activation Function |
|-------------------------------------|---------|--------------|-------|---------------------|
| Convolutional + Batch Normalization | 32 | 3×1 | 20 | ReLU |
| Max Pooling | — | 2×1 | — | — |
| Convolutional + Batch Normalization | 64 | 3×1 | 20 | ReLU |
| Max Pooling | — | 2×1 | — | — |
| Convolutional + Batch Normalization | 128 | 3×1 | 20 | ReLU |
| Max Pooling | — | 2×1 | — | — |
| Fully Connected | — | — | 256 | ReLU |
| Fully Connected | — | — | 64 | ReLU |
| Dropout | — | — | 0.2 | — |

2.3.2. Data Handling and Model Evaluation Metrics

A standard data partitioning and evaluation methodology is used to predict SOC from simulated DSSL and SSL data (Table 4). The datasets (DSSL and SSL) are classified into organic and mineral soils using a threshold of 120 g/kg; samples exceeding this value are considered organic, while the remainder are classified as mineral [55]. For convolutional neural network (CNN) analysis, we exclude organic soil samples ($SOC > 120$ g/kg) due to their limited number, focusing solely on mineral soils for prediction using DSSL (Table 1). Each mineral soil subset is then randomly split into training (80%), validation (10%), and testing (10%) sets for robust evaluation. To avoid overfitting and enhance model performance, 5-fold cross-validation is applied on the combined training and validation sets. This involves splitting the data into five folds, training on four, and evaluating on the fifth, repeated five times to create five trained models. Performance is assessed using the coefficient of determination (R^2), root mean square error (RMSE), residual prediction deviation (RPD), and ratio of performance to inter-quartile range (RPIQ), where higher R^2 and RPIQ, and lower RMSE, indicate better performance.

Table 4. Scenarios -based data splitting in CNNs.

| Scenarios | Number of Samples | Training | Validation | Testing |
|--|-------------------|----------|------------|---------|
| 1—Bare soil-SSL (Baseline) | | | | |
| (A) All measured datasets (organic and mineral soils) | 8941 | 7152 | 894 | 894 |
| (B) Mineral soils | 8896 | 7116 | 890 | 890 |
| (C) Organic soils | 45 | — | — | — |
| 2—Mixed scenario DSSL | | | | |
| (A) All simulated datasets (organic and mineral soils) | 822,572 | 658,057 | 82,257 | 82,258 |
| (B) Simulated datasets based on mineral soils | 818,432 | 654,745 | 81,843 | 81,844 |
| (C) Simulated datasets based on organic soils | 4140 | 3312 | 414 | 414 |
| (D) Bare soil and SMC (mineral soils) | 35,764 | 28,467 | 3558 | 3559 |
| (E) Bare soil, SMC, and PV (mineral soils) | 393,404 | 313,139 | 39,142 | 39,143 |
| (F) Bare soil, SMC, and NPV (mineral soils) | 393,404 | 313,139 | 39,142 | 39,143 |

3. Results

3.1. Descriptive Statistics of LUCAS Bare and Dry Soil Database

Table 5 summarizes the statistical properties of the dataset. It includes a wide range of soil attributes, such as SOC content, which varies substantially from 0.1 to 560.2 g/kg. The high coefficient of variation (CV) observed for variables like OC, $CaCO_3$, and N content

reflects the considerable variability within the LUCAS dataset, which is collected from diverse locations across Europe. This variation likely stems from differences in soil types, land management practices, topography, climate, and land use. Such heterogeneity is crucial for developing robust and generalizable soil models. The OC statistics exhibit a high standard deviation, a large CV, and a positively skewed distribution, highlighting the inherent variability in soil and land characteristics. Although the data are skewed, no transformation is applied, as preserving this variability is important for augmenting the representation of mineral and organic matter in the dataset. Figure 3 shows a boxplot illustrating the distribution of OC normalized to a scale between 0 and 1. The boxplot displays the interquartile range with the median marked, and red points indicate outliers beyond $1.5 \times \text{IQR}$. This histogram illustrates the overall distribution of the original (unnormalized) values.

Table 5. Statistical analysis of topsoil properties from the LUCAS 2015.

| Properties | Samples | Mean | Median | Std. Dev. | Max | Min | CV (%) | Kurtosis | Skewness |
|--|---------|-------|--------|-----------|--------|------|--------|----------|----------|
| Clay (%) | 830 | 25.53 | 25.00 | 11.71 | 45.86 | 2.00 | 45.86 | 2.72 | 0.38 |
| Sand (%) | 830 | 33.90 | 31.00 | 18.47 | 93.00 | 2.00 | 54.47 | 2.69 | 0.63 |
| Silt (%) | 830 | 40.55 | 40.00 | 11.55 | 67.00 | 5.00 | 28.48 | 2.76 | -0.08 |
| pH | 8941 | 6.86 | 7.09 | 1.07 | 9.63 | 3.58 | 15.59 | 2.12 | -0.50 |
| OC (g kg^{-1}) | 8941 | 17.62 | 14.30 | 19.15 | 519.10 | 0.10 | 108.63 | 196.00 | 10.97 |
| CaCO_3 (g kg^{-1}) | 8941 | 84.86 | 2.00 | 157.48 | 976.00 | 0.00 | 185.59 | 7.63 | 82.21 |
| EC (mS/m) | 8941 | 22.12 | 17.47 | 22.95 | 383.00 | 0.45 | 103.71 | 61.17 | 6.51 |

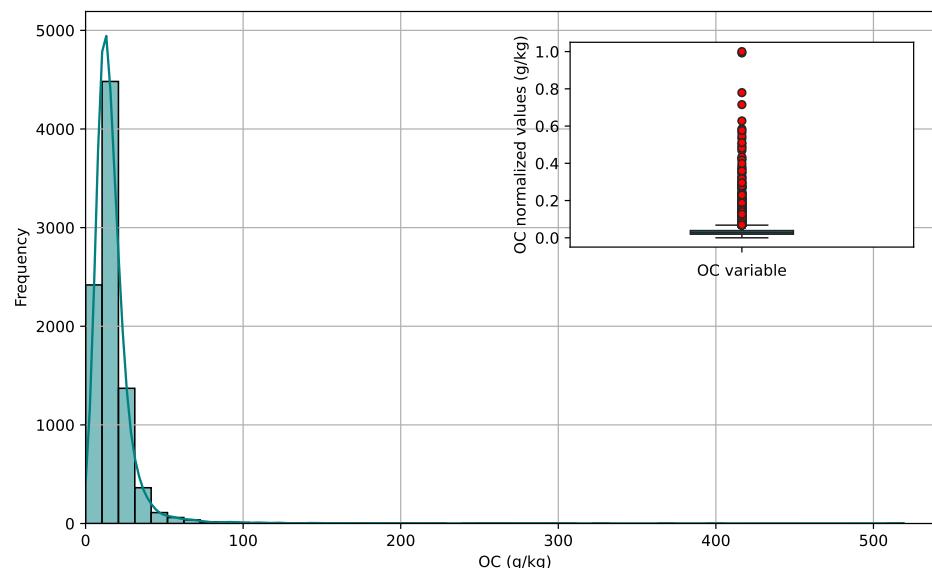


Figure 3. Boxplots and histograms showing the distribution of selected soil variables from the LUCAS dataset (organic carbon), normalized to a 0–1 scale. The boxplots display the interquartile range (IQR) with the median marked by a line. Points represent outliers beyond $1.5 \times \text{IQR}$ from the quartiles.

3.2. Simulation of MLC Model

Figure 4 illustrates the influence of varying soil organic carbon (SOC) content on soil reflectance across the 400–2500 nm spectral range. In panel (a), spectral reflectance curves corresponding to four different SOC levels (2.1, 4, 6.6, and 8.8%) reveal that while the general shape of the spectra remains consistent, subtle variations are evident with increasing SOC. Notably, the spectrum for the lowest SOC level (2.1%) exhibits a slight upward convexity near 800 nm, likely due to the reduced organic matter not fully masking the underlying mineral features. In the visible region (400–700 nm), the reflectance differences between

the samples are more pronounced at lower SOC levels, while the NIR and SWIR regions (700–2500 nm) show greater divergence with increasing SOC content, indicating enhanced absorption by organic matter. Prominent absorption features near 1400 nm and 1900 nm are attributed to water-related vibrational overtones, whereas the dip around 2200 nm is associated with clay minerals, particularly kaolinite. Focusing on a specific soil type with 2.1% SOC (Figure 4b), the full spectral profile highlights regions known to be sensitive to organic carbon, marked as shaded areas. The bands around 600–780 nm, 1400 nm, 1900 nm, and 2200 nm correspond to molecular vibrations of water, organic functional groups (such as C–H and C=O), and mineral-related absorptions. These spectral features are critical for accurate SOC prediction in hyperspectral analysis. Figure 4(b.2) reveals that as the leaf area index (LAI) increases, the simulated photosynthetic vegetation (PV) spectra over bare soil type (1) exhibit enhanced reflectance in the NIR region (750–1100 nm), and reduced reflectance in the visible (400–750 nm) and SWIR (1100–2500 nm) regions—characteristic of healthy canopy growth. Conversely, for non-photosynthetic vegetation (NPV) in Figure 4(b.3), increasing LAI values lead to a general decrease in canopy reflectance across the full wavelength range, indicating higher biomass cover and absorption.

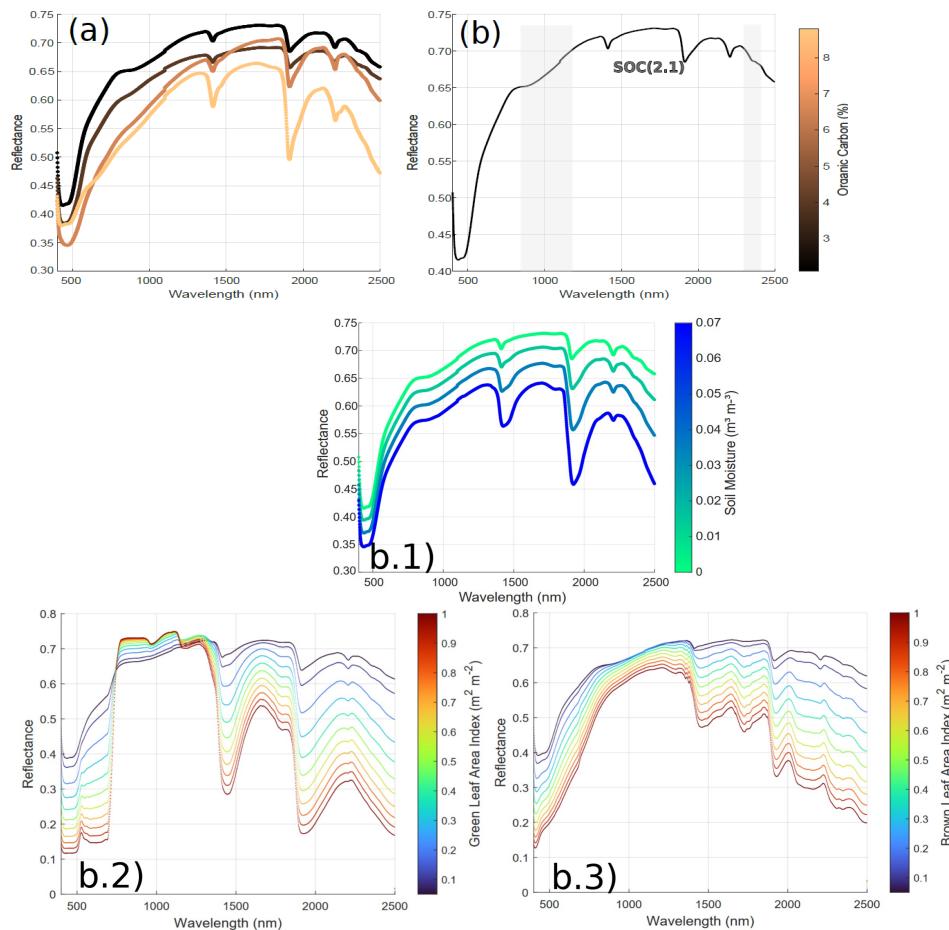


Figure 4. Impact of organic carbon content, soil moisture, and canopy biophysical properties on soil and canopy reflectance. Top row: (a) presents four representative soil spectra, demonstrating the variation in soil types under different organic carbon (OC) concentrations. (b) Shows a selected soil spectrum with a high OC level (8.80%), where the shaded region indicates OC-sensitive wavelengths. Bottom row: Building on the soil type displayed in (b), plot (b.1) illustrates the change in soil reflectance with varying SMC. Plots (b.2,b.3) depict simulated canopy reflectance as a function of increasing green and brown LAI, respectively.

3.3. Model Application

3.3.1. SOC Prediction with Various Associated SMC-LAI PV and NPV

To assess the model's generalizability across green and non-green vegetation, the DSSL pooled database is randomly split into 80% training and 20% validation, ensuring reproducibility with a fixed random seed. A separate dataset of PV and NPV-LAI with SMC is used for testing. For PV, predictive accuracy improves as PV-LAI and SMC decrease, achieving optimal accuracy ($R^2 = 0.8$, RMSE = 4.93 (g/kg)) (Figure 5a,b). In NPV conditions, SOC predictions remain more stable, with consistently high R^2 and low RMSE, peaking at $R^2 = 0.81$ and RMSE = 4.59 (g/kg) (Figure 5c,d). Even at low NPV-LAI, R^2 remains high, indicating a stronger SOC correlation than PV. Under arid conditions, both PV and NPV scenarios show reduced accuracy with higher PV-LAI, highlighting the model's sensitivity to vegetation and moisture variations.

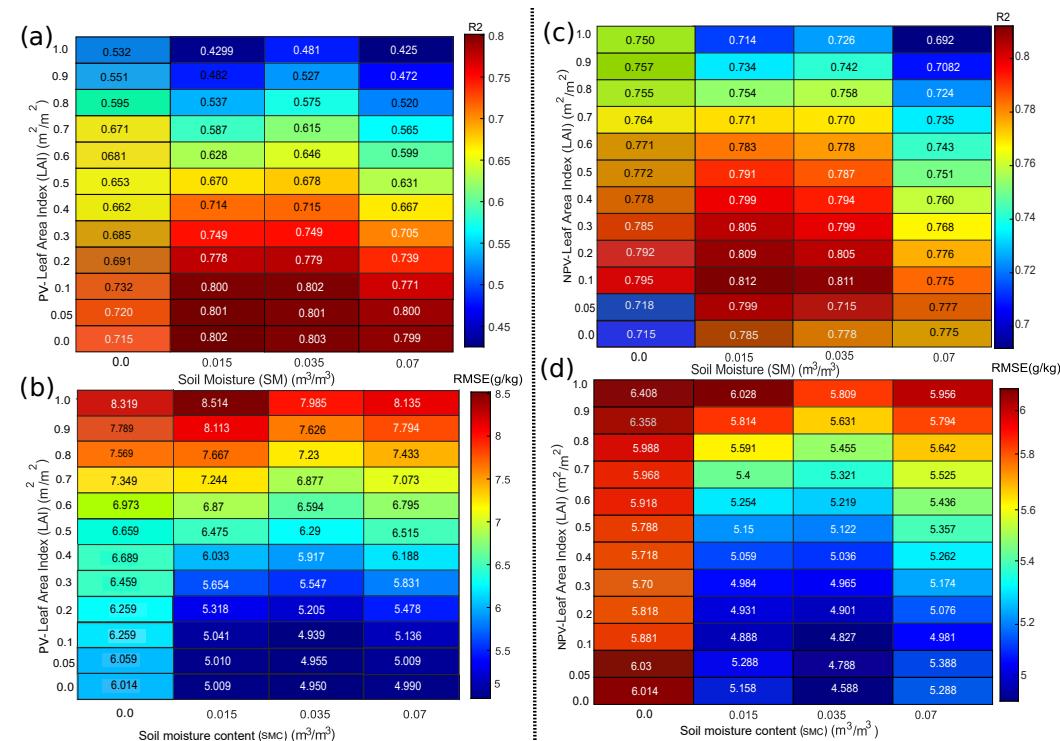


Figure 5. Performance of CNN SOC prediction models represented as heat maps for different mixed scenarios: (a,b) dry and wet soils-PV, and (c,d) dry and wet soils-NPV. Note: PV and NPV content are shown in terms of green and brown LAI values and model performance results are shown in terms of R^2 (top) and RMSE (bottom).

3.3.2. SOC Prediction with Mineral and Organic Soil Data

Figure 6 compares CNN-based SOC prediction under two conditions: bare soil-SSL (baseline) and DSS data, highlighting improved accuracy with the DSSL dataset. The DSSL dataset is split into training (80%), validation (10%), and testing (10%). In the bare soil scenario (Figure 6(a.1,b.1)), the model performs worse than with DSSL data (Figure 6(a.2,b.2)), as most bare soil samples correspond to mineral SOC rather than organic SOC. Consequently, SOC predictions for mineral soils show the highest accuracy, with superior R^2 , lower RMSE, and higher RPD. In contrast, DSSL results show organic SOC (c.2) achieving the highest accuracy ($R^2 = 0.87$, RMSE = 36.62 g/kg), followed by mineral SOC (b.2) ($R^2 = 0.87$, RMSE = 25.6 g/kg), though organic soil samples are limited. Scatterplots confirm this trend, showing better alignment along the 1:1 line in DSSL-based predictions. These findings underscore the importance of diverse spectral data in improving SOC

prediction, particularly for mineral SOC, while bare soil conditions limit accuracy due to reduced spectral information.

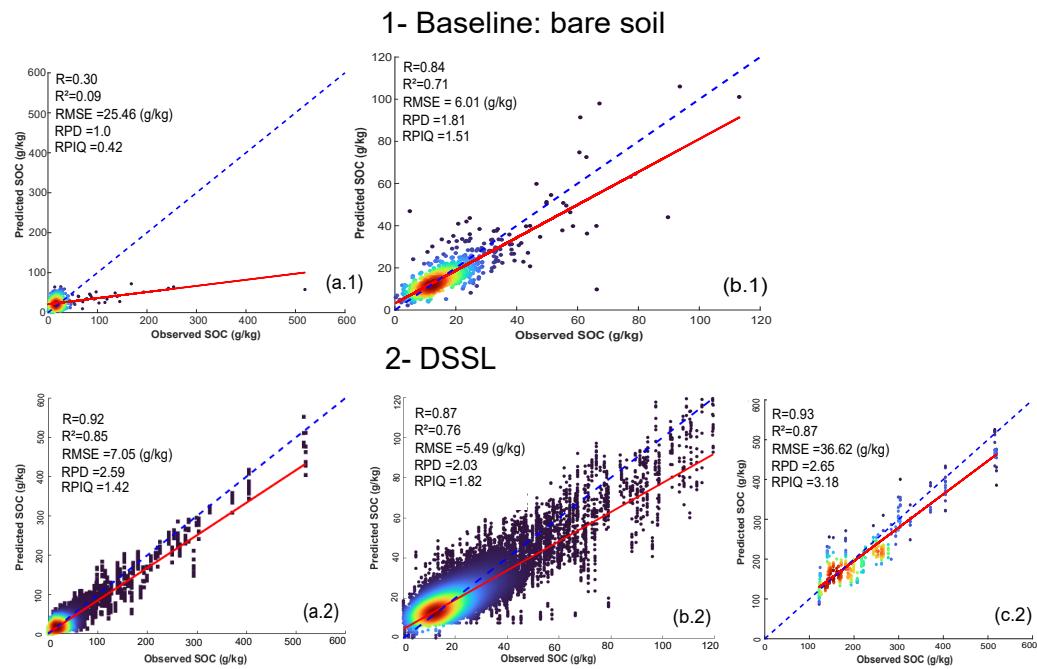


Figure 6. Performance of CNN SOC prediction based on LUCAS (1) and the entire DSSL (2) data: (a.1,a.2) without splitting mineral and organic SOC, (b.1,b.2) mineral SOC, in addition (c.2) for organic SOC using DSSL dataset. However, Figure (c.1) not shown (N/a) due to too few points of LUCAS data.

3.3.3. SOC Prediction for Mixed Dry/Wet Soils and Separated PV or NPV Scenarios

Figure 7 illustrates the model's SOC prediction accuracy. Trained with DSSL and evaluated on a scenario of PV, the model achieves $R^2 = 0.64$, RMSE = 8.14 (g/kg), and RPD = 1.80. Performance improves in scenario of NPV, with $R^2 = 0.72$, RMSE = 5.96 (g/kg), and RPD = 1.90 (Figure 7b). Both datasets show consistent underestimation, more pronounced for green vegetation (Figure 7a).

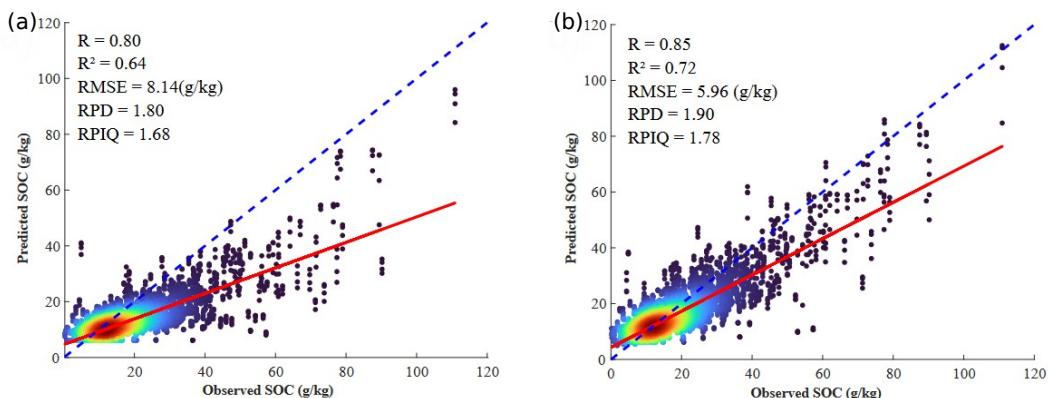


Figure 7. Trained CNN Model Performance: (a) Scenario 2 (Mixed dry/wet soil, PV); (b) Scenario 3 (Mixed dry/wet soil, NPV).

For retraining, Figure 8 compares observed and predicted SOC values under scenarios 2 and 3. Both show strong performance, with R values of 0.93 and 0.92, R^2 of 0.86 and 0.85, low RMSE (4.35 and 0.47 (g/kg)), and favorable RPD (2.257 and 2.50) and RPIQ (2.302 and 2.24) for green and non-green vegetation, respectively. A paired t-test found no significant

difference between datasets. However, underestimation is stronger for green vegetation (Figure 8a), while non-green vegetation predictions (Figure 8b) align more closely with the identity line, indicating better predictive accuracy. The NPV histogram (Figure 8b, 3509 data points) shows a gradual decline and broader predicted value distribution, capturing greater variability. In contrast, the PV histogram (Figure 8a, 2992 data points) has a sharp peak at lower predicted values (10–20), indicating a more constrained range.

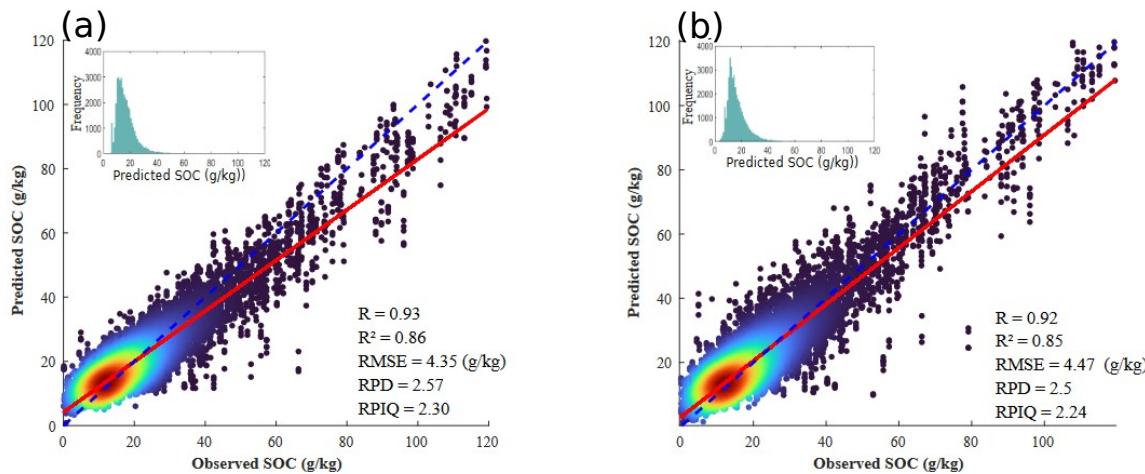


Figure 8. Retrained CNN Model Performance: (a) Scenario 2 (Mixed dry/wet soil, PV); (b) Scenario 3 (Mixed dry/wet soil, NPV).

3.3.4. The Impact of Mixed Scenarios on the Accuracy of SOC Prediction

The impact of different scenarios on SOC prediction accuracy is evaluated, as shown in Table 6. Without vegetation effects, the model performs moderately well ($R^2 = 0.71$, RPIQ = 1.51), but SOC predictions are consistently overestimated (RMSE = 6.01, bias = 0.5). Incorporating soil moisture improves accuracy ($R^2 = 0.86$, RPIQ = 2.44, RMSE = 4.05 (g/kg)) with minimal bias (0.05), emphasizing its importance. Vegetation type also influences accuracy, with NPV yielding better results ($R^2 = 0.74$, RPIQ = 1.71, RMSE = 5.84 (g/kg)) than PV ($R^2 = 0.71$, RPIQ = 1.58, RMSE = 6.31 (g/kg)). While adding NPV reduces overestimation, it still introduces more bias than “bare soil + moisture”. Including all factors (moisture, green, and non-green vegetation) minimizes bias (0.45) while maintaining strong performance ($R^2 = 0.76$, RPIQ = 1.82, RMSE = 5.49 (g/kg)).

Table 6. Overview on the impact of different scenario variations on the accuracy of SOC predictions.

| Scenarios | SOC Prediction Metrics | | | | | |
|---|------------------------|----------------|-------------|-------------|------|-------------|
| | R | R ² | RMSE | RPD | RPIQ | Bias |
| 1—Bare soil | 0.84 | 0.71 | 6.01 | 1.81 | 1.51 | 0.5 |
| 2—Bare soil and moisture effects | 0.93 | 0.86 | 4.05 | 2.69 | 2.44 | 0.05 |
| 3—Bare soil, moisture effects, and PV | 0.85 | 0.71 | 6.31 | 1.77 | 1.58 | 1.27 |
| 4—Bare soil, moisture effects, and NPV | 0.87 | 0.74 | 5.84 | 1.91 | 1.71 | 0.40 |
| 5—DSSL:Bare soil, moisture effects, PV, and NPV | 0.87 | 0.76 | 5.49 | 2.03 | 1.82 | 0.45 |

Note: This table summarizes the impact of various scenario variations on SOC prediction accuracy. Metrics include R (correlation coefficient), R² (coefficient of determination), RMSE (root mean squared error (g/kg)), RPD (residual prediction deviation), RPIQ (residual prediction interquartile range), and bias. Highlighted values indicate improved SOC predictions.

4. Discussion

4.1. RTM-Based Disturbed Soil Spectral Library (DSSL)

This research introduces MLC, a novel RTM framework for improving SOC prediction from Earth Observation data by considering surface disturbances (SMC, PV, and NPV). Traditional vegetation-oriented RTMs (e.g., PROSAIL, SLC, SCOPE, SPART) oversimplify soil reflectance [44,56–58], while soil-specific RTMs (e.g., MARMIT, SOILSPECT, KM, SMART) neglect vegetation [31,39,59,60]. MLC addresses this limitation by combining MARMIT with PROSAIL2 to simulate diverse soil types from the LUCAS laboratory spectra, creating more realistic landscape scenarios for Earth Observation applications. Including surface disturbances in the SOC model calibration improves accuracy using a hybrid CNN-based retrieval approach. Training a 1D-CNN on simulated mixed spectra from DSSL achieved superior accuracy ($R^2 = 0.76$, RMSE = 5.49 (g/kg)) over [50], which used SCOPE with LSTMs ($R^2 = 0.71$, RMSE = 10.60 (g/kg)), highlighting the role of vegetation and soil moisture. Previous studies explored inversion methods, including hybrid ML-based approaches. For instance, Ref. [61] used SVR and PLSR for SOC prediction ($R^2 = 0.55$, 0.69; RMSE = 4.42, 3.68 (g/kg)), while Wu et al. [60,62] applied a modified KM-based RTM with ML retrieval ($R^2 = 0.68$, RMSE = 3.72 (g/kg)). Yuan et al. [63] estimated SOM using spectral indices ($R^2 = 0.73$, RMSE = 4.23 (g/kg)), and Ou et al. [64] used a sensitive band inversion model ($R^2 = 0.42$, RMSE = 5.02 (g/kg)) (Table 7). However, these studies often overlooked soil–vegetation–atmosphere interactions, a gap MLC aims to bridge.

Table 7. Summary of predicted SOC accuracy using RTM-based approaches.

| Method | Input Data | Accuracy Metrics | Reference |
|--|-------------------------------------|---|------------------|
| - MLC with CNN (hybrid model) | Mixed scenarios based on LUCAS | $R^2 = 0.87$, RMSE = 5.49 | Present study |
| - SCOPE with LSTM (hybrid model) | Mixed scenarios based on USDA-SSL | $R^2 = 0.71$, RMSE = 10.60 | Wang et al. [50] |
| - Inversion of KM theory | Local soil spectra (no disturbance) | $R^2 = 0.86$, RMSEP = 0.18% | Yuan et al. [65] |
| - KM theory with wavelength selection | Local soil spectra (no disturbance) | $R^2 = 0.86$, RMSEP = 0.234% | Yuan et al. [63] |
| - SESMRT model with SVR (hybrid model) | ICRAF–ISRIC–SSL datasets | $R^2 = 0.66$, RMSE = 3.923 (GF5); $R^2 = 0.69$, RMSE = 3.54 (HyMap) | Wu et al. [62] |

4.2. Disturbing Factors in SOC Estimation

In addition to these findings, Figure 7 demonstrates improved SOC prediction accuracy under wet conditions, particularly when wet spectra are included in the calibration model, in agreement with previous studies and where vegetation influence is minimal, consistent with prior studies [66,67]. Weak soil moisture data reduced prediction error to 4.05 (g/kg), aligning with [68], who reported accuracy gains within the 25–48% SMC range (Table 6). However, SOC accuracy stabilized around $0.035\text{ m}^3/\text{m}^3$ and declined beyond $0.25\text{ m}^3/\text{m}^3$ [50]. Our refined SMC categories (very low: 0.015, low: 0.035, moderate: $0.07\text{ m}^3/\text{m}^3$) improve accuracy over [50].

Beyond SMC, vegetation type also influenced SOC estimation. Models incorporating PV, NPV, and SMC outperformed PV-LAI or NPV-LAI models by 31% (Figure 7). NPV-LAI performed better ($R^2 = 0.69$ –0.81, RMSE = 4.83–6.02 (g/kg)) than PV-LAI ($R^2 = 0.42$ –0.80, RMSE = 5–8.5 (g/kg)) [50]. However, incorporating all factors did not further enhance accuracy, likely due to dataset size and variability, crucial for CNN model calibration. While adding PV, NPV-LAI, and SMC strengthened SOC prediction, the accuracy remained comparable to using only dry and wet soils with PV-LAI or NPV-LAI (scenarios 3 and 4). This suggests CNN performance depends on dataset size and diversity, with training data quality and reliability being critical for effective calibration and validation.

4.3. SOC Estimation Using CNN for Soil Spectral Library (SSL)

By applying the 1D-CNN model on bare soil of LUCAS data, the results indicate that separate modeling of mineral and organic soils yields superior SOC predictions over a single model for the combined dataset (Figure 6). The best performance for mineral soils achieved an RMSE of 5.49 (g/kg), improving by up to 1.56%, aligning with previous studies [50,69]. Despite these improvements, SOC modeling on the combined dataset also demonstrated acceptable accuracy, with an R^2 of 0.85 and an RMSE of 7.05 g/kg. This suggests that some organic soils remain mixed with mineral soils in the topsoil layer, possibly due to agricultural practices like deep ploughing [70]. Recent studies highlight the superior precision of deep neural networks (DNNs) for SOC prediction, surpassing other machine learning techniques [50,53,54,71]. In particular, Saberioon et al. [54] reported impressive predictive metrics ($R^2 = 0.73$, RMSE = 5.43%, RPD = 3.67) after soil spectral data preprocessing. Aligning with these observations, our findings exhibited comparable accuracy, with an R^2 of 0.71, RMSE of 6.01 (g/kg), and RPD of 1.81.

4.4. Limitations and Future Prospects of SOC Estimation

While the proposed MLC model shows promising results, further research is needed to better understand the interplay between soil properties and spectral characteristics. Factors such as surface roughness, soil texture, particle and pore size distributions, mineralogical composition (e.g., clay content, iron oxides, carbonates), and plant nutrient levels are critical for accurate soil characterization and require further investigation to improve model reliability [72]. These properties significantly influence spectral responses and are essential inputs for process-based models that simulate soil carbon dynamics under various carbon farming practices, including reduced tillage, cover cropping, and organic amendments. To support large-scale assessment of these practices, the current study employs original (unprocessed) SSL spectra within an MLC model grounded in an RTM framework. This approach simulates real-world variability in spectral measurements and enhances the model's generalization across diverse soil conditions. To improve computational efficiency and model interpretability, future work will incorporate feature band selection via global sensitivity analysis to identify the most informative spectral bands. Additionally, active learning strategies may be employed to optimize training with large hyperspectral datasets [73]. Beyond the current architecture, future model development will consider more advanced and hybrid machine learning approaches, such as recurrent neural networks (RNNs), graph neural networks (GNNs), and CNN-ensemble models (e.g., combining CNNs with random forests), to better capture spatiotemporal soil–vegetation interactions and potentially improve prediction accuracy and model adaptability. Moreover, an evaluation of the model's real-time processing capabilities will be important for developing lightweight, deployable solutions suited for operational EO-based soil monitoring. These enhancements will be addressed in future research or incorporated into the broader methodological roadmap to ensure scalability, robustness, and practical applicability. Although in situ validation remains limited, the current MLC model demonstrates strong potential for advancing Earth observation (EO)-based soil monitoring by upscaling soil spectral libraries and enabling large-scale SOC estimation, critical for carbon sequestration tracking and sustainable land management assessment.

5. Conclusions

The MARMIT–Leaf–Canopy (MLC) model, integrating MARMIT and PRO4SAIL2, delivers a novel Disturbed Soil Spectral Library (DSSL) that comprehensively accounts for key surface disturbances: soil moisture content (SMC), photosynthetic vegetation (PV), and non-

photosynthetic vegetation (NPV). This study conclusively demonstrates that incorporating these factors dramatically improves the accuracy of soil organic carbon (SOC) prediction.

Our findings validate the hypothesis that a holistic approach, encompassing SMC, PV, and NPV (scenario 5), significantly enhances SOC prediction. While including SMC alone (scenario 1) already outperformed the bare soil (SSL) model (scenario 2), the presence of vegetation initially reduced accuracy, particularly for PV (scenario 3). However, crucially, the model showed improved accuracy when NPV was combined with bare soil and soil moisture (scenario 4). This adaptability confirms the MLC model's capacity to accurately represent real-world conditions, providing a strong basis for advanced Earth Observation SOC retrieval and robust soil property mapping.

The MLC model's hybrid methodology offers substantial promise for remote sensing applications, especially for upcoming satellite missions. Its proven ability to precisely map SOC in diverse and challenging surface environments is vital for informing critical policies, such as the EU soil monitoring law, thereby advancing sustainable land management, mitigating climate change, and strengthening global food security.

Author Contributions: Conceptualization, S.C. and R.M. and A.A.; formal analysis, A.A.; investigation, M.S. and A.A.; resources, A.A.; data curation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, K.B., M.S., J.A.M.D. and S.C.; visualization, S.C., M.S. and R.M.; supervision, S.C.; project administration, R.M. and S.C.; funding acquisition, S.C. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The database is available upon request from users.

Acknowledgments: We acknowledge financial support from the Worldsoils project (ESA contract No. 400131273/20/I-NB). Additional financial support is provided by the MRV4SOC project, funded under the Horizon programme (Horizon Europe, contract No. 101112754), and the EnMAP science program (grants Nos. 50EE1923 and 50EE2401) under the DLR Space Agency, with resources from the German Federal Ministry of Economic Affairs and Climate Action (BMWK).

Conflicts of Interest: The authors declare no conflicts of interest.

Declaration of Generative AI in the Writing Process: The authors used AI tools to enhance readability and edited the content, taking full responsibility for the published article.

Abbreviations

The following abbreviations are used in this manuscript:

| Acronym | Definition |
|----------|--|
| RTM | Radiative Transfer Model |
| SOC | Soil Organic Carbon |
| CNN | Convolutional Neural Network |
| EnMAP | Environmental Mapping and Analysis Program |
| PRISMA | PRecursore IperSpettrale della Missione Applicativa |
| GaoFen-5 | High-Resolution Earth Observation System |
| CHIME | Copernicus Hyperspectral Imaging Mission for the Environment |
| SBG | Surface Biology and Geology |
| VIS | Visible Spectrum Range |
| NIR | Near-Infrared Spectrum Range |
| VNIR | Visible and Near-Infrared Spectrum Ranges |
| PV | Photosynthetic Vegetation |
| NPV | Non-Photosynthetic Vegetation |
| PROSPECT | Leaf Optical Properties SPECTra Model |

| | |
|----------------------------------|---|
| SAIL | Scattering by Arbitrarily Inclined Leaves |
| PROSAIL | Combined Model: PROSPECT and 4SAIL2 |
| SLC | Soil-Leaf-Canopy |
| INFORM | INvertible FOrest Reflectance Model |
| SCOPE | Soil Canopy Observation, Photochemistry, and Energy Fluxes |
| DART | Discrete Anisotropic Radiative Transfer |
| PRO4SAIL2 | Combined Model: PROSPECT and 4SAIL2 |
| EU-LUCAS | European Union Land Use and Cover Area Frame Survey |
| SOILSPECT | Soil Property Estimation Using Spectral Data |
| MARMIT | Multilayer Radiative Transfer Model of Soil Reflectance |
| MLC | Marmit-Leaf-Canopy Model |
| DSSL | Disturbed Soil Spectral Library |
| LSTM | Long Short-Term Memory Model |
| SVR | Support Vector Regression |
| KM | Kubelka-Munk Model |
| SESMRT | Semi-Empirical Soil Radiative Transfer Model |
| Leaf or Canopy Parameters (g, b) | g = green (photosynthetic) vegetation; b = brown (non-photosynthetic) vegetation |

References

1. van Wesemael, B.; Abdelbaki, A.; Ben-Dor, E.; Chabriat, S.; d’Angelo, P.; Dematte, J.A.M.; Genova, G.; Gholizadeh, A.; Heiden, U.; Karlshofer, P.; et al. A European Soil Organic Carbon Monitoring System Leveraging Sentinel 2 Imagery and the Lucas Soil Data Base. *Geoderma* **2024**, *452*, 117113. [[CrossRef](#)]
2. Baveye, P.C.; Schnee, L.S.; Boivin, P.; Laba, M.; Radulovich, R. Soil organic matter research and climate change: Merely re-storing carbon versus restoring soil functions. *Front. Environ. Sci.* **2020**, *8*, 579904. [[CrossRef](#)]
3. Prudnikova, E.; Savin, I.; Vindeker, G.; Grubina, P.; Shishkonakova, E.; Sharychev, D. Influence of soil background on spectral reflectance of winter wheat crop canopy. *Remote Sens.* **2019**, *11*, 1932. [[CrossRef](#)]
4. Demattè, J.A.; Minasny, B.; Hartemink, A.E. Vital for Sustainable Agriculture: Pedological Knowledge and Mapping. *Eur. J. Soil Sci.* **2025**, *76*, e70040. [[CrossRef](#)]
5. Tahat, M.M.; Alananbeh, K.M.; Othman, Y.A.; Leskovar, D.I. Soil health and sustainable agriculture. *Sustainability* **2020**, *12*, 4859. [[CrossRef](#)]
6. Verrelst, J.; Halabuk, A.; Atzberger, C.; Hank, T.; Steinhauser, S.; Berger, K. A comprehensive survey on quantifying non-photosynthetic vegetation cover and biomass from imaging spectroscopy. *Ecol. Indic.* **2023**, *155*, 110911. [[CrossRef](#)]
7. Fernández-Guisuraga, J.M.; Calvo, L.; Quintano, C.; Fernández-Manso, A.; Fernandes, P.M. Fractional vegetation cover ratio estimated from radiative transfer modeling outperforms spectral indices to assess fire severity in several Mediterranean plant communities. *Remote Sens. Environ.* **2023**, *290*, 113542. [[CrossRef](#)]
8. Bartholomeus, H.; Kooistra, L.; Stevens, A.; van Leeuwen, M.; van Wesemael, B.; Ben-Dor, E.; Tychon, B. Soil organic carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 81–88. [[CrossRef](#)]
9. Verrelst, J.; Camps-Valls, G.; Muñoz-Marí, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.; Moreno, J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 273–290. [[CrossRef](#)]
10. Sunantha, O.; Shao, Z.; Pattama, P.; Potchara, A.; Huang, X.; Zeeshan, A. Machine learning-based estimation of soil organic carbon in Thailand’s cash crops using multispectral and SAR data fusion combined with environmental variables. *Geo-Spat. Inf. Sci.* **2025**, *2025*, 1–23. [[CrossRef](#)]
11. Parvizi, Y.; Fatehi, S. Geospatial digital mapping of soil organic carbon using machine learning and geostatistical methods in different land uses. *Sci. Rep.* **2025**, *15*, 4449. [[CrossRef](#)] [[PubMed](#)]
12. Abdelbaki, A.; Udelhoven, T. A review of hybrid approaches for quantitative assessment of crop traits using optical remote sensing: Research trends and future directions. *Remote Sens.* **2022**, *14*, 3515. [[CrossRef](#)]
13. Berger, K.; Verrelst, J.; Féret, J.B.; Hank, T.; Wocher, M.; Mauser, W.; Camps-Valls, G. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102174. [[CrossRef](#)]
14. Danner, M.; Berger, K.; Wocher, M.; Mauser, W.; Hank, T. Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 278–296.

15. Tagliabue, G.; Boschetti, M.; Bramati, G.; Candiani, G.; Colombo, R.; Nutini, F.; Pompilio, L.; Rivera-Caicedo, J.P.; Rossi, M.; Rossini, M.; et al. Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 362–377. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [\[CrossRef\]](#)
17. Zhang, J.; Xie, T.; Yang, C.; Song, H.; Jiang, Z.; Zhou, G.; Zhang, D.; Feng, H.; Xie, J. Segmenting purple rapeseed leaves in the field from UAV RGB imagery using deep learning as an auxiliary means for nitrogen stress detection. *Remote Sens.* **2020**, *12*, 1403. [\[CrossRef\]](#)
18. Asner, G.P.; Martin, R.E. Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. *Front. Ecol. Environ.* **2009**, *7*, 269–276. [\[CrossRef\]](#)
19. Wang, L.; Qu, Y.; Li, W.; Zhou, G. Deep learning for remote sensing image classification: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2020**, *10*, e1346. [\[CrossRef\]](#)
20. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sens.* **2015**, *7*, 8830–8857. [\[CrossRef\]](#)
21. Chabrillat, S.; Foerster, S.; Segl, K.; Beamish, A.; Brell, M.; Asadzadeh, S.; Milewski, R.; Ward, K.J.; Brosinsky, A.; Koch, K.; et al. The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch. *Remote Sens. Environ.* **2024**, *315*, 114379. [\[CrossRef\]](#)
22. Labate, D.; Ceccherini, M.; Cisbani, A.; De Cosmo, V.; Galeazzi, C.; Giunti, L.; Melozzi, M.; Pieraccini, S.; Stagi, M. The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronaut.* **2009**, *65*, 1429–1436. [\[CrossRef\]](#)
23. Nieke, J.; Despoisse, L.; Gabriele, A.; Weber, H.; Strese, H.; Ghasemi, N.; Gascon, F.; Alonso, K.; Boccia, V.; Tsonevska, B.; et al. The copernicus hyperspectral imaging mission for the environment (CHIME): An overview of its mission, system and planning status. In Proceedings of the SPIE Remote Sensing: Sensors, Systems, and Next-Generation Satellites XXVII, Amsterdam, The Netherlands, 3–7 September 2023; SPIE Publications: Bellingham, WA, USA, 2023; Volume 12729, pp. 21–40.
24. Cawse-Nicholson, K.; Townsend, P.A.; Schimel, D.; Assiri, A.M.; Blake, P.L.; Buongiorno, M.F.; Campbell, P.; Carmon, N.; Casey, K.A.; Correa-Pabón, R.E.; et al. NASA’s surface biology and geology designated observable: A perspective on surface imaging algorithms. *Remote Sens. Environ.* **2021**, *257*, 112349. [\[CrossRef\]](#)
25. Wilson, A.M.; Jetz, W. Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biol.* **2016**, *14*, e1002415. [\[CrossRef\]](#)
26. Proença, V.; Martin, L.J.; Pereira, H.M.; Fernandez, M.; McRae, L.; Belnap, J.; Böhm, M.; Brummitt, N.; García-Moreno, J.; Gregory, R.D.; et al. Global biodiversity monitoring: From data sources to essential biodiversity variables. *Biol. Conserv.* **2017**, *213*, 256–263. [\[CrossRef\]](#)
27. Nocita, M.; Stevens, A.; van Wesemael, B.; Aitkenhead, M.; Bachmann, M.; Barthès, B.; Dor, E.B.; Brown, D.J.; Clairotte, M.; Csorba, A.; et al. Soil spectroscopy: An alternative to wet chemistry for soil monitoring. *Adv. Agron.* **2015**, *132*, 139–159.
28. Vohland, M.; Ludwig, M.; Thiele-Bruhn, S.; Ludwig, B. Quantification of soil properties with hyperspectral data: Selecting spectral variables with different methods to improve accuracies and analyze prediction mechanisms. *Remote Sens.* **2017**, *9*, 1103. [\[CrossRef\]](#)
29. Chabrillat, S.; Eisele, A.; Guillaso, S.; Rogäß, C.; Ben-Dor, E.; Kaufmann, H. HYSOMA: An easy-to-use software interface for soil mapping applications of hyperspectral imagery. In Proceedings of the 7th EARSeL SIG Imaging Spectroscopy Workshop, Edinburgh, UK, 11–13 April 2011; pp. 11–13.
30. Mielke, C.; Chabrillat, S.; Rogass, C.; Boesche, N.K.; Guillaso, S.; Foerster, S.; Segl, K.; Guanter, L. Engeomap and ensomap: Software interfaces for mineral and soil mapping under development in the frame of the enmap mission. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 8369–8372.
31. Jacquemoud, S.; Baret, F.; Hanocq, J. Modeling spectral and bidirectional soil reflectance. *Remote Sens. Environ.* **1992**, *41*, 123–132. [\[CrossRef\]](#)
32. Verhoef, W.; Bach, H. Coupled soil–leaf–canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. *Remote Sens. Environ.* **2007**, *109*, 166–182. [\[CrossRef\]](#)
33. van der Tol, C.; Rossini, M.; Cogliati, S.; Verhoef, W.; Colombo, R.; Rascher, U.; Mohammed, G. A model and measurement comparison of diurnal cycles of sun-induced chlorophyll fluorescence of crops. *Remote Sens. Environ.* **2016**, *186*, 663–677. [\[CrossRef\]](#)
34. North, P.R. Three-dimensional forest light interaction model using a Monte Carlo method. *IEEE Trans. Geosci. Remote Sens.* **2002**, *34*, 946–956. [\[CrossRef\]](#)
35. Gastellu-Etchegorry, J.P.; Demarez, V.; Pinel, V.; Zagolski, F. Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote Sens. Environ.* **1996**, *58*, 131–156. [\[CrossRef\]](#)

36. Labarre, S.; Ferrari, C.; Jacquemoud, S. Surface roughness retrieval by inversion of the Hapke model: A multiscale approach. *Icarus* **2017**, *290*, 63–80. [\[CrossRef\]](#)

37. Verhoef, W.; Van Der Tol, C.; Middleton, E.M. Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX–Sentinel-3 tandem mission multi-sensor data. *Remote Sens. Environ.* **2018**, *204*, 942–963. [\[CrossRef\]](#)

38. Sadeghi, M.; Jones, S.B.; Philpot, W.D. A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. *Remote Sens. Environ.* **2015**, *164*, 66–76. [\[CrossRef\]](#)

39. Bablet, A.; Vu, P.; Jacquemoud, S.; Viallefont-Robinet, F.; Fabre, S.; Briottet, X.; Sadeghi, M.; Whiting, M.L.; Baret, F.; Tian, J. MARMIT: A multilayer radiative transfer model of soil reflectance to estimate surface soil moisture content in the solar domain (400–2500 nm). *Remote Sens. Environ.* **2018**, *217*, 1–17. [\[CrossRef\]](#)

40. Hapke, B. Bidirectional reflectance spectroscopy: 1. Theory. *J. Geophys. Res. Solid Earth* **1981**, *86*, 3039–3054. [\[CrossRef\]](#)

41. Ångström, A. The albedo of various surfaces of ground. *Geogr. Ann.* **1925**, *7*, 323–342.

42. Lekner, J.; Dorf, M.C. Why some things are darker when wet. *Appl. Opt.* **1988**, *27*, 1278–1280. [\[CrossRef\]](#)

43. Bach, H.; Mauser, W. Modelling and model verification of the spectral reflectance of soils under varying moisture conditions. In Proceedings of the IGARSS’94–1994 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 8–12 August 1994; Volume 4, pp. 2354–2356.

44. Jacquemoud, S.; Baret, F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* **1990**, *34*, 75–91. [\[CrossRef\]](#)

45. Feret, J.B.; François, C.; Asner, G.P.; Gitelson, A.A.; Martin, R.E.; Bidel, L.P.; Ustin, S.L.; Le Maire, G.; Jacquemoud, S. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens. Environ.* **2008**, *112*, 3030–3043. [\[CrossRef\]](#)

46. Verhoef, W. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sens. Environ.* **1984**, *16*, 125–141. [\[CrossRef\]](#)

47. Jones, A.; Fernandez-Ugalde, O.; Scarpa, S. LUCAS 2015 Topsoil Survey. *Present. Dataset Results EUR* **2020**, *30332*, 616084.

48. Kooistra, L.; Clevers, J.G. Estimating potato leaf chlorophyll content using ratio vegetation indices. *Remote Sens. Lett.* **2016**, *7*, 611–620. [\[CrossRef\]](#)

49. Botha, E.J.; Leblon, B.; ZebARTH, B.; Watmough, J. Non-destructive estimation of potato leaf chlorophyll from canopy hyperspectral reflectance using the inverted PROSAIL model. *Int. J. Appl. Earth Obs. Geoinf.* **2007**, *9*, 360–374. [\[CrossRef\]](#)

50. Wang, S.; Guan, K.; Zhang, C.; Lee, D.; Margenot, A.J.; Ge, Y.; Peng, J.; Zhou, W.; Zhou, Q.; Huang, Y. Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: Assessing potential of airborne and spaceborne optical soil sensing. *Remote Sens. Environ.* **2022**, *271*, 112914. [\[CrossRef\]](#)

51. Casa, R.; Jones, H. Retrieval of crop canopy properties: A comparison between model inversion from hyperspectral data and image classification. *Int. J. Remote Sens.* **2004**, *25*, 1119–1130. [\[CrossRef\]](#)

52. Abdelbaki, A.; Schlerf, M.; Verhoef, W.; Udelhoven, T. Introduction of variable correlation for the improved retrieval of crop traits using canopy reflectance model inversion. *Remote Sens.* **2019**, *11*, 2681. [\[CrossRef\]](#)

53. Tsakiridis, N.L.; Keramaris, K.D.; Theocharis, J.B.; Zalidis, G.C. Simultaneous prediction of soil properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network. *Geoderma* **2020**, *367*, 114208. [\[CrossRef\]](#)

54. Saberioon, M.; Gholizadeh, A.; Ghaznavi, A.; Chabrilat, S.; Khosravi, V. Enhancing soil organic carbon prediction of LUCAS soil database using deep learning and deep feature selection. *Comput. Electron. Agric.* **2024**, *227*, 109494. [\[CrossRef\]](#)

55. Stolt, M.H.; Bakken, J. Inconsistencies in terminology and definitions of organic soil materials. *Soil Sci. Soc. Am. J.* **2014**, *78*, 1332–1337. [\[CrossRef\]](#)

56. Verhoef, W.; Bach, H. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. *Remote Sens. Environ.* **2003**, *87*, 23–41. [\[CrossRef\]](#)

57. Van der Tol, C.; Verhoef, W.; Timmermans, J.; Verhoef, A.; Su, Z. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences* **2009**, *6*, 3109–3129. [\[CrossRef\]](#)

58. Yang, P.; van der Tol, C.; Yin, T.; Verhoef, W. The SPART model: A soil-plant-atmosphere radiative transfer model for satellite measurements in the solar spectrum. *Remote Sens. Environ.* **2020**, *247*, 111870. [\[CrossRef\]](#)

59. Kubelka, P. Ein Beitrag zur Optik der Farbanstriche. *Z. Tech. Phys.* **1931**, *12*, 593–601.

60. Wu, F.; Tan, K.; Wang, X.; Ding, J.; Liu, Z.; Han, B. A semi-analytical radiative transfer model for explaining soil spectral features. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103250. [\[CrossRef\]](#)

61. Ou, D.; Tan, K.; Li, J.; Wu, Z.; Zhao, L.; Ding, J.; Wang, X.; Zou, B. Prediction of soil organic matter by Kubelka-Munk based airborne hyperspectral moisture removal model. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103493. [\[CrossRef\]](#)

62. Wu, F.; Tan, K.; Wang, X.; Ding, J.; Liu, Z. A novel semi-empirical soil multi-factor radiative transfer model for soil organic matter estimation based on hyperspectral imagery. *Geoderma* **2023**, *437*, 116605. [\[CrossRef\]](#)

63. Yuan, J.; Wang, X.; Yan, C.; Chen, S.; Wang, S.; Zhang, J.; Xu, Z.; Ju, X.; Ding, N.; Dong, Y.; et al. Wavelength selection for estimating soil organic matter contents through the radiative transfer model. *IEEE Access* **2020**, *8*, 176286–176293. [\[CrossRef\]](#)

64. Ou, D.; Tan, K.; Wang, X.; Wu, Z.; Li, J.; Ding, J. Modified soil scattering coefficients for organic matter inversion based on Kubelka-Munk theory. *Geoderma* **2022**, *418*, 115845. [[CrossRef](#)]
65. Yuan, J.; Hu, C.; Yan, C.; Li, Z.; Chen, S.; Wang, S.; Wang, X.; Xu, Z.; Ju, X. Semi-empirical soil organic matter retrieval model with spectral reflectance. *IEEE Access* **2019**, *7*, 134164–134172. [[CrossRef](#)]
66. Seidel, M.; Vohland, M.; Greenberg, I.; Ludwig, B.; Ortner, M.; Thiele-Bruhn, S.; Hutzengs, C. Soil moisture effects on predictive VNIR and MIR modeling of soil organic carbon and clay content. *Geoderma* **2022**, *427*, 116103. [[CrossRef](#)]
67. Jiang, Q.; Chen, Y.; Guo, L.; Fei, T.; Qi, K. Estimating soil organic carbon of cropland soil at different levels of soil moisture using VIS-NIR spectroscopy. *Remote Sens.* **2016**, *8*, 755. [[CrossRef](#)]
68. Yang, P.; Wang, Y.; Hu, B.; Li, S.; Chen, S.; Luo, D.; Peng, J. Predicting soil organic carbon content using simulated insitu spectra and moisture correction algorithms in southern Xinjiang, China. *Geoderma Reg.* **2024**, *37*, e00783. [[CrossRef](#)]
69. Sakhaei, A.; Gebauer, A.; Ließ, M.; Don, A. Spatial prediction of organic carbon in German agricultural topsoil using machine learning algorithms. *Soil* **2022**, *8*, 587–604. [[CrossRef](#)]
70. Alcántara, V.; Don, A.; Well, R.; Nieder, R. Deep ploughing increases agricultural soil organic matter stocks. *Glob. Chang. Biol.* **2016**, *22*, 2939–2956. [[CrossRef](#)]
71. Zhong, L.; Guo, X.; Xu, Z.; Ding, M. Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. *Geoderma* **2021**, *402*, 115366. [[CrossRef](#)]
72. Chabrillat, S.; Ben-Dor, E.; Cierniewski, J.; Gomez, C.; Schmid, T.; van Wesemael, B. Imaging spectroscopy for soil mapping and monitoring. *Surv. Geophys.* **2019**, *40*, 361–399. [[CrossRef](#)]
73. Berger, K.; Rivera Caicedo, J.P.; Martino, L.; Woher, M.; Hank, T.; Verrelst, J. A survey of active learning for quantifying vegetation traits from terrestrial earth observation data. *Remote Sens.* **2021**, *13*, 287. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.