Journal of Plant Nutrition and Soil Science

WILEY-VCH

RESEARCH ARTICLE OPEN ACCESS

A Stochastic Approach to Quantifying the Propagation of Uncertainty in Soil Organic Carbon Content

Leonardo Inforsato[1] | Pablo Rosso[1] | Magdalena Main-Knorn[1] | Ahsan Raza[1] | Siyu Huang[1] | Gohar Ghazaryan[1,2] | Claas Nendel[1,3,4]

[1]Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany | [2]Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany | [3]Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany | [4]Global Change Research Institute of the Czech Academy of Sciences, Brno, Czech Republic

Correspondence: Leonardo Inforsato (Leonardo.Inforsato@zalf.de)

Received: 17 August 2024 | **Revised:** 3 July 2025 | **Accepted:** 7 November 2025

Academic Editor: Astley Hastings

Keywords: Monte Carlo simulations | soil organic carbon content | stochastic imaging | uncertainty analysis

ABSTRACT

Background: Precision agriculture (PA) is a site-specific management approach that utilises spatiotemporal information to improve productivity while also promoting sustainability. Accurate estimates of soil properties, along with the uncertainty of these estimates, are necessary for decision-making in PA. An essential soil quantity required to accurately predict crop yield is the soil organic carbon (SOC) content. To obtain the large amount of information necessary for PA implementation, the use of satellite images has become a common practice. This allows the spatial interpolation of soil properties. However, this type of indirect approach carries higher relative uncertainties than direct measurements (e.g., laboratory experiments). Although error evaluations of soil properties resulting from indirect approaches are constantly considered, the consequences of error are not.

Aim: This work introduces a methodology to analyse the error propagation from predictions of SOC digital maps, using the Monte Carlo (MC) method.

Method: We stochastically generated an error range for SOC maps, using one original map of SOC, and used these maps as inputs for a process-based model that simulated crop yields. Our approach evaluates how the error inherent in SOC observations and the subsequent spatial interpolation impacts crop yield forecasting, providing insights for decision-making and further PA implementation.

Results: Our results show promise in the proposed method, delivering results that are difficult to obtain. The MC method was able to handle complex, non-linear error distributions and provide a comprehensive probabilistic assessment of uncertainty, which is important for accurately predicting the impact of SOC variability on crop yield.

Conclusion: This method offers a degree of flexibility and robustness that is not achievable with deterministic or simpler analytical approaches, ensuring more reliable and informative insights for PA.

1 | Introduction

Food production increasingly relies on technology-intensive farming systems to meet global demand while maintaining sustainability. One recent innovation is precision agriculture (PA), which has grown as a site-specific management approach, integrating spatial and temporal data to optimise decisions at the sub-field level (Bongiovanni and Lowenberg-DeBoer 2004;

[Correction added on 7 January 2026, after first online publication: Funding and Acknowledgement sections have been updated.]

Taylor 2023). One requirement for the implementation of PA is the accurate spatial mapping of the soil properties across fields. Soil texture and soil organic carbon (SOC) are key variables, as they influence water availability (WA) for plants (Ankenbauer and Loheide 2017; Minasny and McBratney 2018), root development, greenhouse gas emissions (Lehtinen et al. 2014) and nutrient cycling (Plante and Parton 2007). As SOC is highly variable at the field scale, inaccurate estimates can lead to suboptimal management and yield variability (Groß et al., 2023). Therefore, improving SOC mapping is essential for effective PA implementation.

Multiple approaches to estimate WA for agronomical purposes are available (Asgarzadeh et al. 2014; Brisson 1998; De Melo et al. 2023; Inforsato and de Jong van Lier 2021); nevertheless, the most common method is based on two soil hydraulic parameters (SHPs): field capacity (θ_{fc}) and permanent wilting point (θ_{wp}) using the equation $WA = \theta_{fc} - \theta_{wp}$ (Gómez et al. 2023; Rai et al. 2017). It is difficult to determine θ_{fc} and θ_{wp} directly through experimentation, leading to the use of pedotransfer functions (PTFs). SOC and soil texture data (i.e., sand, clay and silt content) are common inputs for PTFs, which provide the SHPs (Botula et al. 2012; Givi et al. 2004; Qiao et al. 2019). One modern PTF that has stood out for European soils was proposed by Tóth et al. (2015), who used 18,537 individual European soil samples to develop it.

One challenge in determining SHPs across arable fields is the fine-scale spatial variability of both SOC and texture. Providing SOC and texture inputs to PTFs in a way that represents all within-field zones requires extensive soil sampling and analysis. A more practical and less labour-intensive alternative involves using processed satellite imagery combined with machine learning to predict SHPs across the field. This approach falls under the concept of digital soil mapping (DSM), defined as the process of inferring spatially explicit soil information by integrating measured observations with spatial and non-spatial predictors, typically through data-driven models (Lagacherie and McBratney 2006; Wadoux et al. 2020). The core components of DSM include input data, predictive modelling methods (e.g., machine learning) and the resulting spatial soil information, such as maps and the uncertainties in them (Chen et al. 2022; Minasny and McBratney 2016; Söderström et al. 2016). DSM follows the SCORPAN framework, $S = f(s, c, o, r, p, a, n)$, which expresses soil attributes or classes (S) as a function of other soil properties (s), climate (c), organisms (o), relief or topography (r), parent material (p), age (a) and spatial position (n) (McBratney et al. 2003). Although horizontal soil heterogeneity may indeed require resolutions finer than 10 per 10 m, information from satellites such as Sentinel-2 L2A are readily accessible and provide multiple spectral bands at this resolution. Moreover, even field data collected by automated machinery, such as harvesters, are rarely acquired at higher resolutions (Kumari and Karthikeyan 2023; Tran et al. 2022).

Dynamic process-based, or mechanistic, crop models provide a comprehensive understanding of plant development and estimates of crop yield potential. If accurate plant and soil parameters are available, the simulations offer valuable knowledge of the various stages of crop development. Agroecosystem models (AEMs) may also provide insights into soil–plant interaction processes, and they can link the response of crops to environmental factors. Various validated AEMs are available, including APSIM (Holzworth et al. 2014), DSSAT (Jones et al. 2003), MONICA (Nendel et al. 2011, 2014) and Hermes (Graß et al. 2015; Kersebaum 2007). Using an AEM for explanatory or advisory purposes at the field scale, aiding decision-making strategies in PA, requires information on soil properties (such as texture and SOC) at the sub-field scale, so the model can reproduce crop growth that typically occurs across heterogeneous fields (Ewert et al. 2015; Wallor et al. 2018).

Quantifying soil properties inherently involves measurement errors, which tend to be higher when estimations are made indirectly, such as predicting SOC from satellite imagery using machine learning models (Yuzugullu et al. 2024; Zhou et al. 2021). Although such errors can often be quantified, their impact on downstream modelling, for example, yield forecasting, remains largely unassessed (Angelopoulou et al. 2019; Ellinger et al. 2019; Vaudour et al. 2022). Even moderate errors in soil property estimates can propagate through process-based models and significantly compromise the accuracy of decisions derived from them (Vågen et al. 2018).

A widely used method to evaluate error and uncertainty propagation is the Monte Carlo (MC) approach, which generates repeated random samples based on a known probability distribution (Heuvelink et al. 1989; Papoulis 1991). MC has proven particularly effective for complex, nonlinear models such as AEMs. For instance, Pinheiro and de Jong van Lier (2021) applied MC simulations to evaluate how soil parameter uncertainty propagates into yield predictions. In a related context, Wilks (1997) used a moving-block bootstrap to resample geophysical time series while preserving autocorrelation by controlling the block sizes. These studies emphasise the importance of stochastic modelling in environmental simulation.

To address spatial uncertainty in SOC predictions, we propose a stochastic imaging protocol that explicitly incorporates spatial autocorrelation into the MC framework. Our method generates multiple spatially autocorrelated realisations of SOC maps by combining pixel-level prediction uncertainty and local spatial correlation. These maps are then used as inputs for the Hermes AEM to simulate crop yield distributions. This allows for a direct assessment of how uncertainty in SOC maps propagates through process-based modelling—an aspect often overlooked in PA workflows.

The methodological foundation of this study lies in pedometrics—the application of quantitative methods to analyse and predict the spatial distribution of soil properties while explicitly accounting for uncertainties (McBratney et al. 2003; McBratney and Minasny 2018). Our framework combines pedometric modelling, spatial statistics and mechanistic crop modelling to assess uncertainty propagation from SOC predictions to simulated yields. This work therefore represents a novel pedometric contribution to uncertainty propagation in DSM, with direct relevance for decision-making in PA.

2 | Materials and Methods

2.1 | Booßen Study Site

The study was conducted on an agricultural field of 70 ha, located 10 km from the town of Booßen (52.394° N, 14.463° E, 77 m asl), in the Federal State of Brandenburg, Germany. The site, further referred to as 'Boo', has sandy loamy soil, resulting from repeated Pleistocene glaciations by the continental Scandinavian ice sheet as well as by subsequent periglacial and interglacial Holocene geomorphic processes. The climate is characterised by an average annual precipitation of 544 mm and an average temperature of 9.7°C. The study started on 17 September 2020, with the sowing of winter rye, variety KWS Tayo. An amount of 6.5×10^3 kg m^{-2} of P and K was applied to the soil a few days later. In March 2021, another 4.0×10^3 kg m^{-2} of nitrogen was applied with other nutrients as a combination of solid and liquid fertilisers. The crop was harvested on 23 October 2021; the harvester was equipped with a digital yield monitor and a navigation system based on the global positioning system (GPS), so the location and yield were measured throughout the field.

Germany's National Meteorological Service (*Deutscher Wetterdienst*—DWD) provided the data for daily temperature, global radiation, wind speed and air humidity. The data were measured at four weather stations in the vicinity of Boo—in Müncheberg, Manschnow, Lindenberg and Frankfurt an der Oder. Daily precipitation data were provided by a private weather station located within the town of Booßen.

2.2 | Soil Sampling

Soil samples were collected in two surveys, one in April 2020 and the other in March 2021, providing two soil datasets: (1) 250 soil samples from the topsoil layer (0.0 to 0.3 m) and (2) 50 samples from three different layers each, 0.0–0.3, 0.3–0.6 and 0.6–0.9 m. All samples were analysed for SOC content, defined as the percentage of the weight of the organic carbon per weight of soil sample (mass %), and texture, defined as the percentage of sand, silt and clay by weight in the soil sample (mass %), in a certified laboratory. The locations within the field are shown in Figure 1. Although the datasets were collected 1 year apart, temporal changes in SOC and texture were assumed negligible, as no significant land use or management changes occurred between samplings. This assumption is supported by Wuest and Durfee (2024), who observed minimal short-term variability in SOC under stable field conditions.

The statistical description of the laboratory analysis results is provided in Table 1. The table summarises the mean (μ) and standard deviation (σ) for SOC and soil texture (sand and clay content) across different soil layers. These statistics provide a comprehensive overview of the variability and distribution of soil properties within the study site, which are crucial for understanding the input data used in the subsequent error propagation analysis and crop yield simulations.

TABLE 1 | The mean μ and the standard deviation σ of the sampled soil organic carbon (SOC) in %, and texture as sand and clay contents (%).

	n (−)	Layer depth (m)	μ (%)	σ (%)
SOC	300	0.0–0.3	1.022	0.489
SOC	50	0.3–0–6	0.436	0.479
SOC	50	0.6–0.9	0.207	0.253
Sand	300	0.0–0.3	76.996	8.098
Sand	50	0.3–0–6	75.651	8.350
Sand	50	0.6–0.9	74.493	9.700
Clay	300	0.0–0.3	5.502	2.812
Clay	50	0.3–0–6	6.987	3.836
Clay	50	0.6–0.9	8.292	5.083

Note: The results are given per soil layer. n represents the number of locations from which the soil samples were taken.

2.3 | SOC Content Data

2.3.1 | Sentinel-2 Images

'Sentinel' is a European Earth observation mission developed by the European Space Agency (ESA) that features optical imaging for land services. The mission consists of satellites that provide high-resolution (10–60 m) multispectral images with a revisit time of 5 days at the equator. One of the satellites is the Sentinel-2, which provides Level 2A (S2A) processed images. The S2A is composed of 13 spectral bands that span from the visible and near-infrared (VNIR) to the short-wave infrared (SWIR) regions of the electromagnetic spectrum. This spectral diversity makes S2A, particularly suitable for soil and vegetation monitoring (Dedeoğlu et al. 2020).

Three S2A images from 8 April 2020, 20 September 2020 and 10 September 2021 were selected to facilitate the SOC digital map estimates. These images were selected on the basis of their temporal characteristics (outside of the growing season, which ensured bare soil conditions) and no cloud cover over the area. The selection was based on the available L2A SCL scene classification, which classified the whole Boo field as 'bare soil' using the reflectance ratio of Band 2/Band 11 (ESA 2015), where Band 2 captures the wavelengths in the blue region, whereas Band 11 covers the SWIR region. Preprocessing of the satellite images included clipping to the Boo field extent (6169 pixels per image), masking and resampling all selected bands to 10 m spatial resolution using nearest neighbour interpolation.

Although every S2A image for the years 2020 and 2021 has 13 bands, only 10 bands were used to predict the SOC of the topsoil layer (specifically: 2, 3, 4, 5, 6, 7, 8, 8A, 11 and 12). Some bands have different resolutions (with squared pixels representing 10, 20 and 60 m edges on the soil). To homogenise the images to the same resolution, the nearest neighbour interpolation method was implemented, resulting in a resolution representing 10 m for all bands.

To ensure that the soil was bare, we calculated the Normalised Difference Vegetation Index (NDVI) for every pixel using Equa-

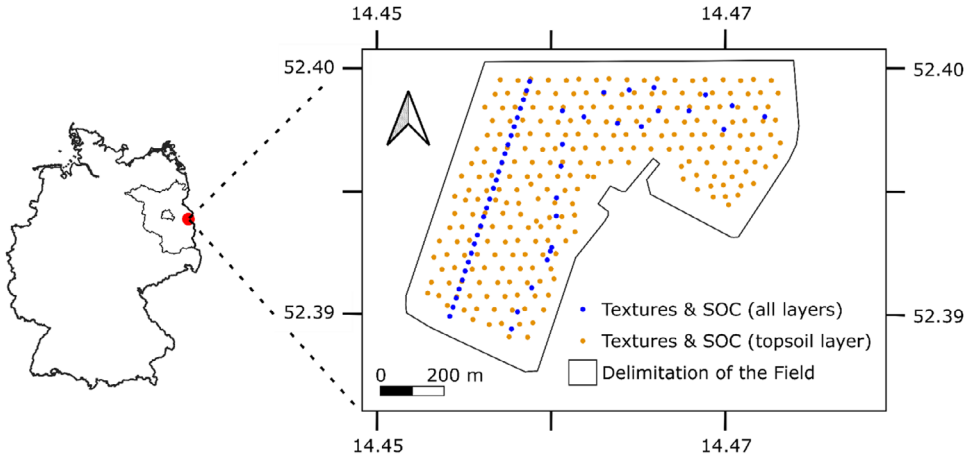

FIGURE 1 | Left: Location of the Boßen field with Germany. Right: Locations of soil samples collected within the Booßen field for soil organic carbon content and soil texture determination. Two soil datasets are displayed: topsoil samples in orange (0.0–0.3 m depth), taken at 250 locations, and three-layer samples in blue (0.0–0.3, 0.3–0.6 and 0.6–0.9 m), taken at 50 locations. SOC, soil organic carbon.

tion (1). This calculation was part of the data preprocessing to confirm the suitability of the images for SOC estimation. For each of the three images, the mean and deviation of the NDVI were 0.177 (±0.03), 0.125 (±0.02) and 0.258 (±0.06), respectively. These values indicated minimal vegetation cover, providing more evidence that the soil was bare for accurate SOC prediction:

$$\mathrm{NDVI} = \frac{\mathrm{NIRn} - \mathrm{red}}{\mathrm{NIRn} + \mathrm{red}}, \quad (1)$$

where *NIRn* is the narrow infra-red band.

2.3.2 | SOC Digital Mapping

The SOC content profile was determined in two steps: First, we produced estimates of SOC for the topsoil layer (0.0–0.3 m) of the Boo field, using topsoil data from both datasets. Second, we produced an estimate of SOC for the deeper layers (0.3–0.6 and 0.6–0.9 m) on the basis of the topsoil layer (0.0–0.3 m) content. For these steps, all SOC values were log-transformed using the natural logarithm to reduce the bias of the data distribution. The SOC bias was quantified using the Fisher–Pearson coefficient of skewness. The logarithm transformation successfully reduced the bias of measured SOC values: from 1.57, 2.35 and 2.78 to 0.10, 0.63 and 0.65 for layers 0.0–0.3, 0.3–0.6 and 0.6–0.9 m, respectively. These transformed SOC values were then used in successive analyses to ensure more accurate and unbiased results.

Subsequently, we employed a Random Forest (RF) algorithm to estimate the SOC for the topsoil layer (0.0–0.3 m) across the entire Boo (Breiman 2001; Ho 1995). The model was trained using the dataset containing the 300 topsoil samples (Table 1) as the target variable. From the three satellite images selected, pixels corresponding to the location of each soil sample were extracted, with the respective spectral data, to serve as the independent variables in the model. In total, we used 10 bands from the three images to predict the topsoil SOC. To train the RF model, we used the Scikit-learn package version 1.3.2 in Python version 3.12.1. The RF parameters were 10 estimators (trees), the maximum number of features for best split as the square root of the number of features (i.e., max_features = 'sqrt'); the other RF parameters were set as the default. After the model was trained, the topsoil SOC was predicted using the independent variables (image bands).

We also used RF to estimate the SOC for deeper layers (0.3–0.6 and 0.6–0.9 m), but instead of using the image bands, the SOC contents of the topsoil layer (0.0–0.3 m) served as the independent variables. The RF parameters and cross-validation parameters were the same as for the topsoil layer 0.0–0.3 m.

For the validation, k-fold cross-validation was used with $k = 10$. The data were split into 10 equal folds; each fold was used once for validation and nine times for training, and the results were averaged (Kohavi 1995) considering the three layers.

For this article, the SOC digital map generated using this process is designated the 'Original' prediction.

2.3.3 | Benchmark Mapping With Block Ordinary Kriging (BOK) for the Topsoil Layer

The topsoil dataset contains 300 SOC observations for the 0.0–0.30 m layer (≈3 samples × 100 m^{-2}). To benchmark the satellite-based RF described in Section 2.3.2, we produced an independent SOC map using BOK (Mondal et al. 2017; Parvizi and Fatehi 2025). The interpolation was carried out using SAGA-GIS 7.8.2 software (Conrad et al. 2015). The BOK and the performance comparison took place in the following steps:

Variogram modelling: SOC values were log-transformed in the same way as for the other SOC observed data used for the SOC mapping. A linear semi-variogram was fitted by least squares.

Block dimension: Kriging estimates were computed for 10 m × 10 m blocks, identical to the Sentinel-2 pixel grid structure, so that both approaches shared the same spatial support.

TABLE 2 | Statistical overview of the predicted soil texture digital map of Booßen, with mean (μ), and standard deviation (σ) of the texture content for each of the three different layer depths, predicted for each pixel image of the field.

Texture	Layer depth (m)	μ (%)	σ (%)
Sand	0.0–0.3	77.32	6.93
Clay	0.0–0.3	5.40	2.24
Sand	0.3–0.6	76.32	6.38
Clay	0.3–0.6	5.95	2.51
Sand	0.6–0.9	76.26	7.14
Clay	0.6–0.9	7.23	3.38

Cross-validation: Spatially stratified 10-fold cross-validation (Kohavi 1995) was applied.

Performance metrics: Accuracy was quantified by the cross-validated root mean square error ($RMSE_{CV}$) and the coefficient of determination (R^2), allowing direct comparison with the satellite-based RF model.

2.4 | Digital Map of Soil Texture for Booßen

The textures of the soil samples were combined with georeferenced electrical resistivity data and collected with proximal sensing throughout the field; this resulted in digital soil texture maps at three layer depths (0.0–0.3, 0.3–0.6 and 0.6–0.9 m). The three images were combined to form the digital soil texture profile, composed of 6169 pixels carrying texture data for the three layers of each image pixel.

The pixels of the image are displayed as a regular square grid, with each pixel representing an area of 10 m × 10 m of the field. Table 2 provides the soil texture statistics for all pixels across the Boo.

The soil texture errors (Table 2) must be taken into account for further comparison; the $RMSE_{CV}$ for the prediction of texture classes is 4.22% for sand and 2.17% for clay for all three layers.

The methodology used for the determination of texture maps and thus the $RMSE_{CV}$ is available in Sections S1, S1.1, and S1.2.

2.5 | Sensitivity of Crop Yields to SOC and Texture

We then conducted a sensitivity analysis based on the work of Turanyi (1990) and Saltelli and Bolado (1998) to make the uncertainty comparison of different soil properties possible. This in turn makes it possible to assess the relative impact of SOC and texture content on hydraulic properties and their effect on simulated crop yield.

Local sensitivity (Se) is defined here as the partial derivative of the model output variable y with respect to one specific input variable of the model x, at a specific parameter value, $x = x_0$. This method (Equation 2) enables to quantify the immediate response of y to x:

$$\mathrm{Se} = \left.\frac{\delta y}{\delta x}\right|_{x=x_0}. \tag{2}$$

However, for the same model output (y), it is cumbersome to use Equation (2) to directly compare the different input variables. To directly compare the sensitivity between two input parameters, we therefore weighted Se by the respective $RMSE_{CV}$ of the variable. This results in the weighted sensitivity, wS, defined as

$$\mathrm{wS} = \left.\frac{\delta y}{\delta x}\right|_{x=x_0} \times \mathrm{RMSE}_{\mathrm{CV}}(x) = \mathrm{Se} \times \mathrm{RMSE}_{\mathrm{CV}}(x). \tag{3}$$

The first desired characteristic of Equation (3) was the dimension of wS. Regardless of the unit of the variable x, wS maintains the same unit of the variable y. The characteristic of having the same physical dimension (unit) allows the direct comparison of wS considering two different input variables, x_1 and x_2, as shown in the following equation:

$$\dim\left\{\frac{\delta y}{\delta x_1} \times \mathrm{RMSE}_{\mathrm{CV}}(x_1)\right\} = \dim\left\{\frac{\delta y}{\delta x_2} \times \mathrm{RMSE}_{\mathrm{CV}}(x_2)\right\}. \tag{4}$$

The term dim in Equation (4) denotes the physical dimension of the expression inside the curly brackets.

The second desired characteristic of wS is the weighting of the error itself, which means that wS quantifies the influence of uncertainty in x on the model output y, incorporating both the model sensitivity and the expected input error. This formulation is conceptually aligned with approaches in traditional risk assessment, where the impact of uncertain variables is evaluated by weighting sensitivity (or hazard) by variability. A similar principle also underlies the first-order error propagation framework described by Heuvelink et al. (1989), chapter 4.1, where the model sensitivity and input uncertainty together determine the output variance.

For our analysis, the SOC content and the sand and clay fractions were analysed in relation to field capacity, θ_{fc}, ($m^3\ m^{-3}$), wilting point, θ_{wp}, ($m^3\ m^{-3}$) and crop yield (kg m^{-2}), respectively. For the field capacity and wilting point estimate, the modern PTFs proposed by Tóth et al. (2015) were used, and Hermes was applied to predict yields. As the sensitivity analysis is limited to a fixed point in space, we randomly selected 100 profile points from Boo to calculate all described wS, followed by taking the average for the entire sensitivity analysis.

2.6 | Stochastic Imaging Protocol

The stochastic process framework is based on a random sampling of a multivariate normal distribution with correlated parameters. The process requires a vector containing the mean values of a parameter set, and the covariance matrix containing covariances between each parameter pair. We have adapted the covariance matrix to make it possible to implement it in SOC digital maps. The method uses the mean SOC values as the mean values,

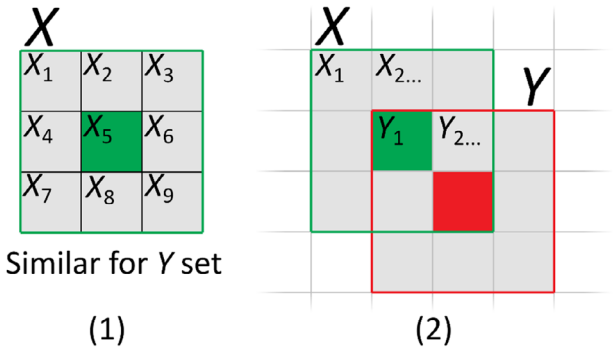

FIGURE 2 | Schematic of the pixel region selection to calculate the correlation between each pair of pixels in an image. On the left (1), the region of 3 × 3 pixels is selected as variable *X*. On the right, (2), the schematic of the correlation between the red and green pixels in the centre of regions *X* and *Y*.

and for the covariance matrix, we propose the use of spatial autocorrelation and the $RMSE_{CV}$ of the SOC values.

2.6.1 | Covariance and Autocorrelation Matrix

The autocorrelation matrix of the SOC values is based on the Pearson correlation. The procedure is a particular case of the normalised cross-correlation presented in Lewis (1995). To calculate the autocorrelation matrix, we first use a method to calculate the correlation between two pixels. First, a 3 × 3 pixel area is defined, with the pixels inside the area numbered X_1 to X_9, as shown in the image on the left in Figure 2; then a second area is delimited, represented as the square outlined in red, on the right side of Figure 2, with the pixels designated as Y_1 to Y_9. Then the pixels are arranged as (X, Y) ordered pairs: $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots, (X_9, Y_9)$. After linking the nine-pixel pairs, the autocorrelation is calculated on the basis of

$$\mathbf{R}_{i,j} = \frac{\mathrm{cov}\,(X, Y)}{\sigma_X \sigma_Y}, \tag{5}$$

where $\mathrm{cov}(X, Y)$ is the covariance calculated from the pixel pairs, σ is the standard deviation of the subscripted pixel set X or Y, and the subscripts i and j represent the pixel coordinates. The calculated correlation is established as the correlation between pixels X_5 and Y_5 (the central pixels). The process is repeated for all possible pixel pairs in the image. For the pixel pairs that are not neighbours, the autocorrelation is considered zero. The correlation matrix **R** is constructed with the $\mathbf{R}_{i,j}$ elements.

The covariance matrix **M** is generated using the correlation matrix, similar to that of De Jong Van Lier (2023). In this step, the error of each pixel value is used to build the covariance matrix. Each element of the correlation matrix is multiplied by both respective error values, expressed as the $RMSE_{CV}$,

$$\mathbf{M}_{i,j} = \mathbf{R}_{i,j} \times \left(\mathrm{RMSE}_{\mathrm{CV},i} \times \mathrm{RMSE}_{\mathrm{CV},j}\right), \tag{6}$$

where $\mathbf{M}_{i,j}$ is the covariance matrix element with coordinates i and j. In Equation (6), the $RMSE_{CV,i}$ and $RMSE_{CV,j}$ substitute the deviations σ_X and σ_Y. To ensure that the covariance matrix is positive semidefinite, we use a Python adaptation of D'Errico' (2024) algorithm for MATLAB, based on Higham (1988).

For Boo specifically, as the images are composed of 6961 pixels each, the total elements in the correlation matrix are $6961^2 = 48{,}455{,}521$. As the input of the correlation matrix was deliberately limited to the adjacent pixels, most elements in the matrix are zeros.

2.6.2 | Decomposition of the Covariance Matrix and Random Samplings

Having generated the Original SOC map and the covariance matrix, the next step was to decompose the covariance matrix, **M**, into lower triangular matrices **L**:

$$\mathbf{M} = \mathbf{L}\mathbf{L}^{\mathrm{T}}, \tag{7}$$

where the superscript T indicates the transposition of a matrix. Cholesky decomposition is computationally efficient and was used for this step.

After the decomposition, the random autocorrelated multivariate maps were generated using the equation:

$$\mathbf{n} = \mathbf{L}\mathbf{z} + \boldsymbol{\mu}, \tag{8}$$

where the vector **n** contains the newly sampled SOC values, **z** is a univariate random variable obeying normal distribution with zero as mean and variance one, and **μ** is the vector containing all Original SOC image pixel values. The procedure can be performed an arbitrary number of times, generating one new random SOC digital map each time. For Boo, a total of 5000 SOC images for layer 0.0–0.3 m were generated.

2.7 | Subsoil Layers Random Sampling

The SOC subsoil layers were estimated on the basis of the topsoil layer. For each of the 5000 topsoil (0.0–0.3 m) SOC maps generated, two additional subsoil SOC maps were created: one for layer 0.3–0.6 m and the other for layer 0.6–0.9 m. The process is defined in two stages:

i. From one sampled SOC topsoil map (0.0–0.3 m), an estimate for both layers 0.3–0.6 and 0.6–0.9 m was generated using the respective calibrated RF model.

ii. For each of the generated subsoil layers described in Stage (i), a new random sampling process was done for the same layer. For this step, we applied the stochastic imaging protocol described before. The randomly generated SOC map uses the mean values from the SOC map predicted by the model to generate the covariance matrix; the $RMSE_{CV}$ used is the $RMSE_{CV}$ of the model.

By repeating the process for every sampled SOC map from the topsoil layer (0.0–0.3 m), we obtained 5000 SOC image profiles for all three layers. The digital SOC profile images provided by this method will be referred to as 'Samples' in this text.

2.8 | Crop Yield Simulation

In this step, we first simulated crop yields of Boo using the Hermes model. Weather and crop parameters for the model simulations were considered as described in the methods section. A Hermes simulation was carried out for each of the pixels of the soil digital maps, that is, one for each 10 m × 10 m field region. The Hermes model requires soil data below a depth of 0.9 m, so we expanded the information of the 0.6–0.9 layer down to 2.0 m.

As output, in addition to crop yield, Hermes provides information on plant growth stress due to the absence of water and nitrogen (N). If the plant lacks water, the water growth-limiting factor is less than 1, indicating non-optimal growth. Similarly, if there is a lack of N, the N growth-limiting factor is also less than 1, also resulting in non-optimal growth. Both growth-limiting factors are multiplied together to produce a single growth-limiting factor. The supply of N from soil organic matter turnover is a function of the C/N ratio of the soil, which in Hermes was kept constant at 11 throughout the simulation. Mean values of water and N stress were calculated for the entire field.

The first simulation batch used the Original SOC map. After that, all the 5000 SOC samples were simulated. All other model variables besides the SOC were kept constant through all simulations.

After the simulations, all the outputs were organised, resulting in a single Original Boo SOC prediction, and 5000 additional Boo sampled scenarios, with each scenario containing 6961 crop yield values (one for each image pixel).

The process flow with the stochastic sampling to generate the SOC digital maps for the three layers is shown in Flowchart I of Figure 3. The simulation process using Hermes is also shown in Figure 3, as Flowchart II.

2.9 | SOC Random Sampling and Simulation Output Analysis

Next, the SOC random sampling outputs from the entire Boo site were analysed. We used Moran's I (Li et al. 2007) to calculate all three different layers for the spatial autocorrelation. In addition, the mean values and the standard deviation of the SOC maps' random sampling were analysed.

To understand how the SOC sampling process behaved towards a particular pixel through the different realisations, we also conducted a single-pixel analysis for the SOC Sampled maps. Pixel choice was made on the basis of skewness, the highest values and the lowest values.

Multiple statistical analyses were conducted for crop yield. The RMSE between the measured crop yield throughout the field was compared with the crop yield provided from the Original and the Sampled scenarios. The global autocorrelation for the measured data, the Original and the Sampled simulated crop yields were also calculated. Finally, we calculated R^2 and Willmott statistics (Willmott and Matsuura 2005) to compare performance.

3 | Results and Discussion

3.1 | The Accuracy of SOC Measurements in Topsoil for Interpolation Based on Satellite Imaging as Well as Block Kriging

The method based on satellite imaging with RF achieved an $RMSE_{CV}$ = 0.14 mass % and R^2 = 0.90 for the topsoil layer. BOK yielded an $RMSE_{CV}$ = 0.15 mass % and R^2 = 0.84. Thus, incorporating multispectral information and machine learning provided a small improvement over purely geostatistical interpolation (ΔRMSE = 0.01 mass %; ΔR^2 = 0.06). The difference is smaller than the analytical uncertainty of laboratory SOC measurement, according to De Vos et al. (2007), which established a limit of 0.04 mass % for SOC determination, indicating that the two mapping approaches are statistically equivalent for this dataset.

3.2 | Weighted Sensitivity

The weighted sensitivities (wS) for soil texture and SOC profiles, along with the $RMSE_{CV}$ between the measured and predicted values, are summarised in Table 3. The analysis reveals that the sand fraction has the highest absolute weighted sensitivity concerning field capacity (θ_{fc}), indicating its significant influence. The sign of the sensitivities was retained to preserve directional information. SOC follows the sand fraction closely, with a difference of 0.14 in the magnitude of absolute values. For the wilting point (θ_{wp}), sand also exhibits the most substantial impact on average, followed by clay and SOC.

Despite the high sensitivity of θ_{fc} and θ_{wp} to the presence of sand, crop yield is most sensitive to SOC according to the weighted sensitivity values. This can be attributed to two main factors: the inherently high sensitivity (Se) of crop yield to SOC (as shown in Table 4) and the role of SOC in the Hermes model for calculating the available nitrogen to plants. An in-depth examination of the Hermes daily output files revealed that nitrogen stress occurred on 62% of the days on average, with an average N growth-limiting factor of 0.79. In comparison, water stress was observed on only 11% of the days, with an average water growth-limiting factor of 0.93, evidencing the importance of N, which is driven by SOC levels.

3.3 | SOC Analysis of the Booßen Field

The Sampled images exhibit a lower global autocorrelation than the Original image, as illustrated in Table 5. Specifically, the mean value of Moran's I for the Sampled SOC maps is consistently lower than that of the Original SOC map across all layers. The standard deviation of Moran's I values is also the lowest for the 0.0–0.3 m layer among the sampled maps.

This decrease in Moran's I values for the Sampled SOC maps indicates that the random sampling process inevitably introduces noise into the data, which was expected. Despite this, it preserves the SOC pattern throughout the field, as can be seen through visual analysis (Figure 4). The Sampled SOC maps appear noisier

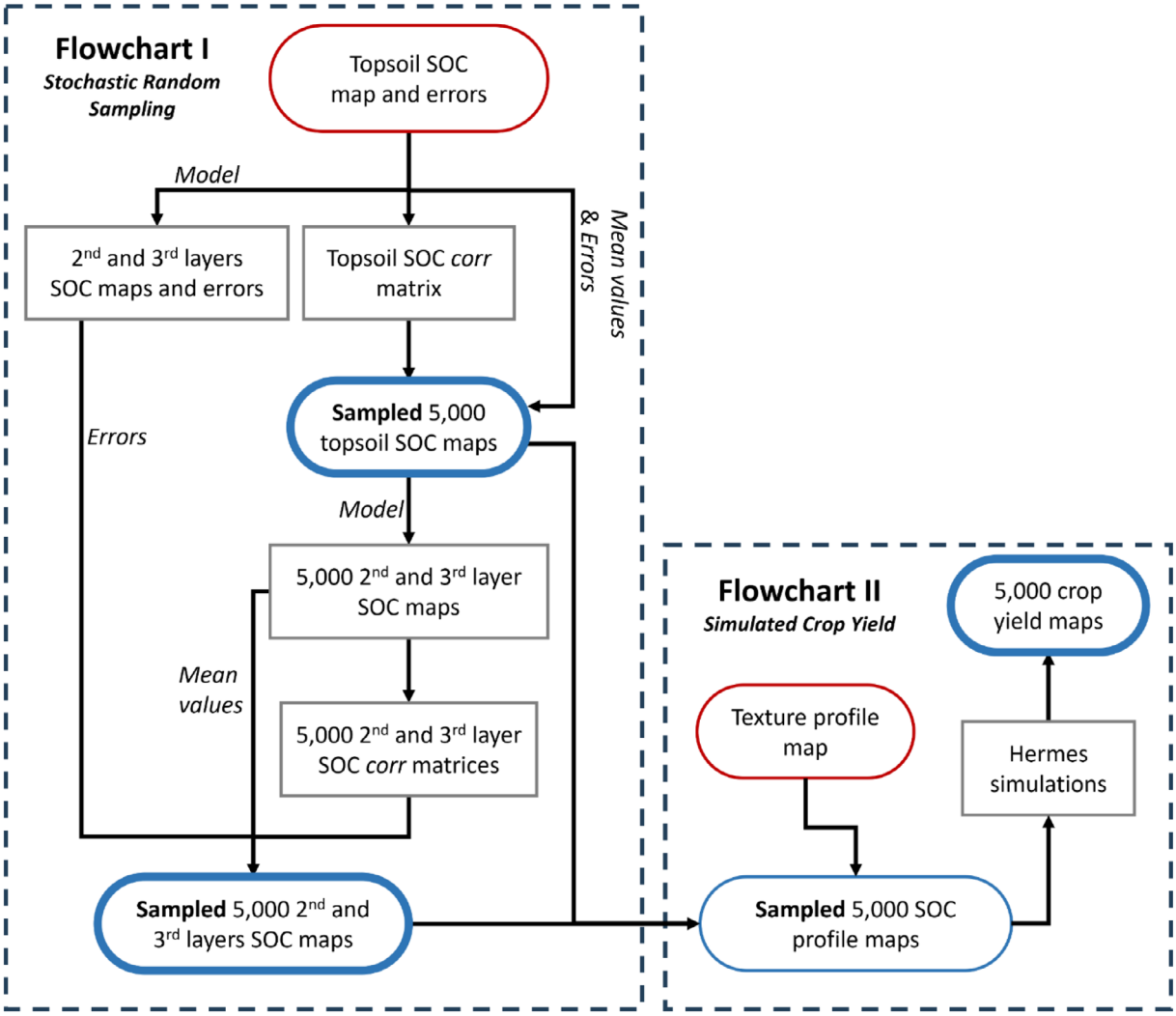

FIGURE 3 | Process flows of the stochastic random sampling (Flowchart I) and simulated crop yield (Flowchart II). Flowchart I shows the process to generate 5000 soil organic carbon maps (SOC) for three layers: topsoil (0.0–0.3 m) layer, 2nd (0.3–0.6 m) and 3rd (0.6–0.9 m) layers, inside the blue rounded boxes. The abbreviation corr refers to the spatial autocorrelation matrix. Flowchart II shows the process used to simulate the crop yield using the randomly sampled SOC digital maps. SOC, soil organic carbon.

TABLE 3 | Mean weighted sensitivity tests for field capacity, θ_{fc}, wilting point, θ_{wp}, and crop yield prediction for the variables sand, clay and organic carbon (SOC) content in the soil profiles.

		Weighted sensitivity (wS)		
	RMSEcv (%)	θ_{fc} ($m^3\ m^{-3}$)	θ_{wp} ($m^3\ m^{-3}$)	Crop yield ($kg\ m^{-2}$)
Sand	4.22	−1.43	−0.96	-2.77×10^{-2}
Clay	2.17	0.87	0.83	1.23×10^{-2}
SOC	0.18	1.29	0.28	9.38×10^{-2}

Note: The $RMSE_{CV}$ is the cross-validated root mean square error of the mass %.

than the Original SOC map; however, the overall spatial pattern of SOC distribution has remained.

Figure 4 presents a digital SOC map for the soil layer depth of 0.0−0.3 m. On the left is the Original estimated SOC map, and on the right is an example of the 5000 Sampled SOC maps.

The analysis confirms that although the introduction of stochastic variations reduces the global autocorrelation, the essential spatial trends of SOC across the Boo are still recognisable. This indicates that the proposed stochastic sampling method is effective in maintaining the overall structure of the original SOC map while allowing for the quantification of uncertainty.

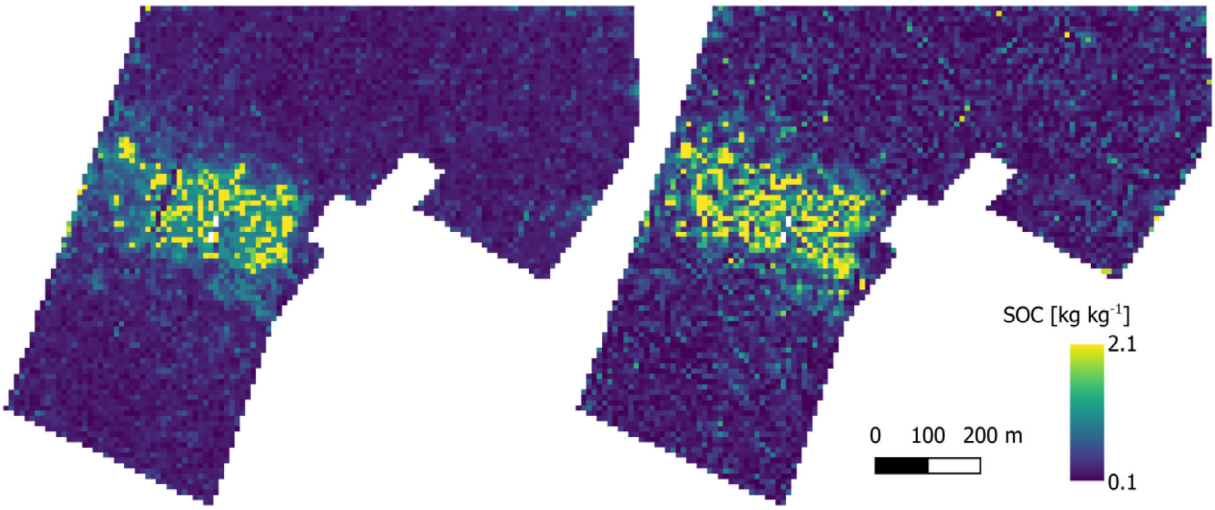

FIGURE 4 | Digital soil organic carbon map (SOC), of the soil layer depth 0.0–0.3 m. On the left, is the originally estimated SOC, and on the right, one example out of 5000 sampled SOC maps.

TABLE 4 | Mean sensitivity (Se) of the crop yield to the sand, clay and organic carbon (SOC) contents in soil.

	Se to crop yield (kg m^{-2})
Sand	-0.657×10^{-2}
Clay	0.568×10^{-2}
SOC	52.104×10^{-2}

TABLE 5 | Moran's test values of the soil organic carbon digital maps (SOC).

(m)	Moran's original	Moran's μ	Moran's σ
0.0–0.3	0.969	0.869	0.005
0.3–0.6	0.703	0.458	0.023
0.6–0.9	0.485	0.226	0.022

Note: The Original SOC estimate for Booßen field, the mean Moran's values (μ) for the multiple randomly Sampled SOC maps, and the standard deviation (σ) of Moran's test for the Sampled SOC maps.

Figure 5 illustrates the frequency distributions of the SOC values for Boo across three different soil layer depths: 0.0–0.3, 0.3–0.6 and 0.6–0.9 m. The left column displays the distributions for the Original SOC estimates, whereas the right column presents distributions for a randomly sampled SOC digital map. In the Original SOC maps, the frequency distributions show distinct peaks that become less pronounced with increasing soil depth. The Sampled SOC maps, in contrast, exhibit smoother frequency distributions, reflecting the introduction of stochastic variability. Despite this added noise, the overall trends are consistent across layers, with SOC values generally decreasing as soil depth increases. The selected Sampled scenario demonstrates a smoother distribution compared to the Original SOC data, which indicates the efficacy of the stochastic sampling method in preserving the general SOC distribution while introducing the necessary variability. Although only one Sampled SOC map is shown, the visual characteristics of the others are similar, confirming the robustness of the sampling process. These frequency distributions emphasise the reliability of the stochastic approach in maintaining the underlying patterns of SOC across the field while accounting for uncertainty, thereby enhancing the robustness of subsequent analyses and decision-making processes.

Figure 6 presents the distribution of SOC values for two selected pixels, px_1 and px_2, across the 5000 randomly generated SOC digital maps. Each histogram represents the distribution of SOC values for one of these pixels at the different soil depths 0.0–0.3, 0.3–0.6 and 0.6–0.9 m. The pixels were chosen on the basis of the skewness of their SOC values in the 0.0–0.3 m layer, with px_1 exhibiting the lowest skewness (left column) and px_2 the highest skewness (right column). The skewness values for each depth layer are indicated in the figure. All skewness values are positive due to the logarithmic transformation applied to the SOC data. This transformation ensures that the distribution frequencies maintain a well-behaved pattern, despite the spatial correlation introduced by the random sampling. For px_1, the SOC distributions display lower skewness across all layers, indicating a more symmetric distribution of values around the mean. Conversely, px_2 shows higher skewness, reflecting a more asymmetric distribution with a longer tail on the right side. The distributions highlight the impact of stochastic sampling on individual pixel SOC values, preserving the inherent variability and ensuring realistic spatial correlations. This approach allows for a comprehensive analysis of uncertainty propagation through the SOC predictions, enhancing the robustness and reliability of the findings.

3.4 | Results of Simulations

Figure 7 presents the frequency distributions of crop yield values for Boo, comparing the measured data, the Original predicted data, and one case from the Sampled scenario. The histograms reveal notable differences between the measured and simulated data. The measured data exhibit a bimodal distribution, indicating the presence of two predominant yield values. In contrast, both the Original predicted data and the Sampled scenario display

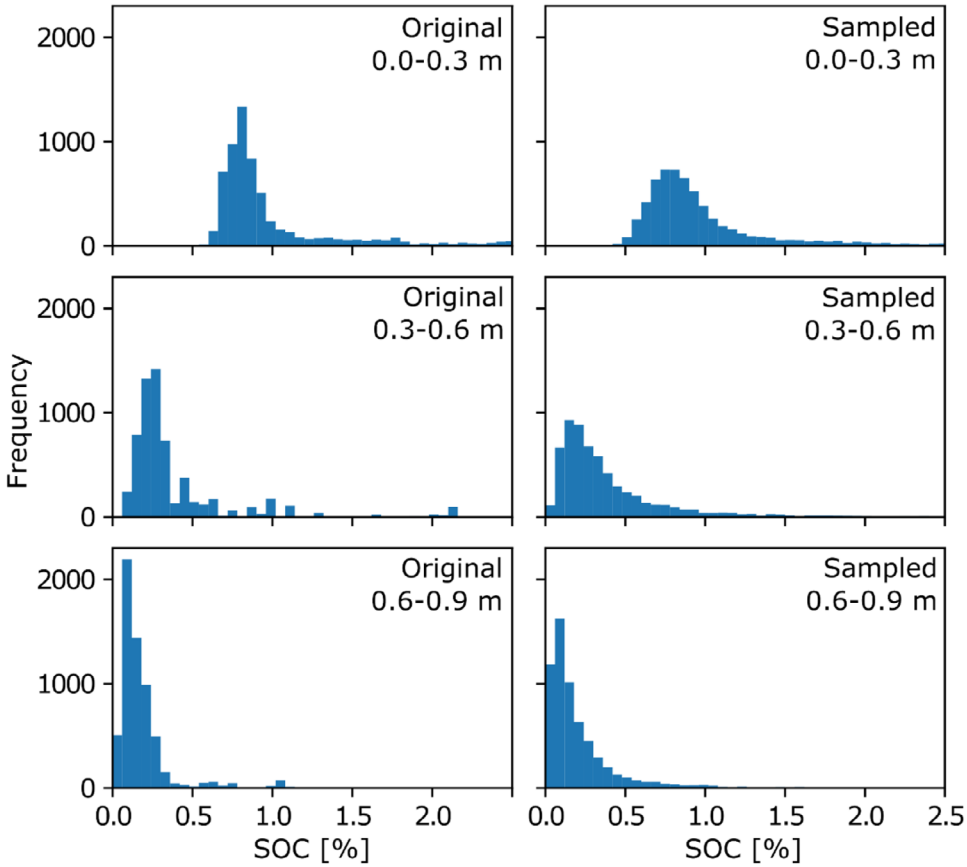

FIGURE 5 | Frequencies of soil organic carbon (SOC, in mass %) map values on the Booßen field map for three different layer depths: 0.0–0.3, 0.3–0.6 and 0.6–0.9 m. Left: the original estimate for the SOC map values of the Booßen field. Right: the frequencies of one exemplary sampled SOC map.

a multimodal distribution, with three distinct peaks. These peaks correspond to the regions observed in the measured data, albeit with reduced intensity. In addition, the measured crop yield data showed a broader distribution range (0.3–1.0 kg m^{-2}) compared to the Original and Sampled scenarios (0.4–0.9 kg m^{-2}). This suggests that the simulated results fail to capture the full variance observed in the measured crop yields. Despite these differences, the general pattern of crop yield distribution remains across the simulated scenarios. This consistency indicates that the stochastic sampling method effectively incorporates the underlying variability of the Original SOC data into the randomly sampled crop yield predictions. The R^2 and Willmott statistics further quantify the performance of the simulations. For the Original prediction compared to the measured data, the R^2 and Willmott statistics are 0.56 and 0.68, respectively. For the Sampled scenarios, these values are 0.43 and 0.65, respectively. Although the Sampled scenarios show slightly lower performance metrics, they still provide a reasonable approximation of the measured crop yields. Although the simulated crop yield distributions do not fully replicate the measured data's variance, they successfully capture the key patterns and variability, demonstrating the utility of the stochastic sampling approach in providing robust crop yield predictions that account for the inherent uncertainties in SOC estimates.

The simulations using the Original SOC digital map produced crop yield estimates that significantly differed from those derived from the Sampled SOC maps. Table 6 presents the RMSE values between the measured crop yield data and the simulated scenarios. The RMSE for the Original SOC map is provided, along with the mean, maximum and minimum RMSE values for the simulations using the Sampled SOC maps. The Original SOC map resulted in an RMSE of 0.1027 kg m^{-2}, which indicates the error magnitude between the predicted and measured crop yields. In contrast, the Sampled SOC scenarios produced a slightly higher mean RMSE of 0.1192 kg m^{-2}, with minimum and maximum RMSE values of 0.1147 and 0.1234 kg m^{-2}, respectively. The increase in RMSE for the Sampled scenarios can be expected, as

TABLE 6 | Comparative analysis of RMSE values for crop yield predictions in the Booßen field.

Simulation	RMSE (kg m^{-2})
Original	0.1027
Sampled mean	0.1192
Sampled minimum	0.1147
Sampled maximum	0.1234

Note: The Original simulation refers to the RMSE of the yield prediction using the originally estimated soil organic carbon map (SOC). The mean, the maximum and the minimum, refer to the RMSE of the simulations realized with the sampled SOC.
Abbreviation: RMSE, root mean square error.

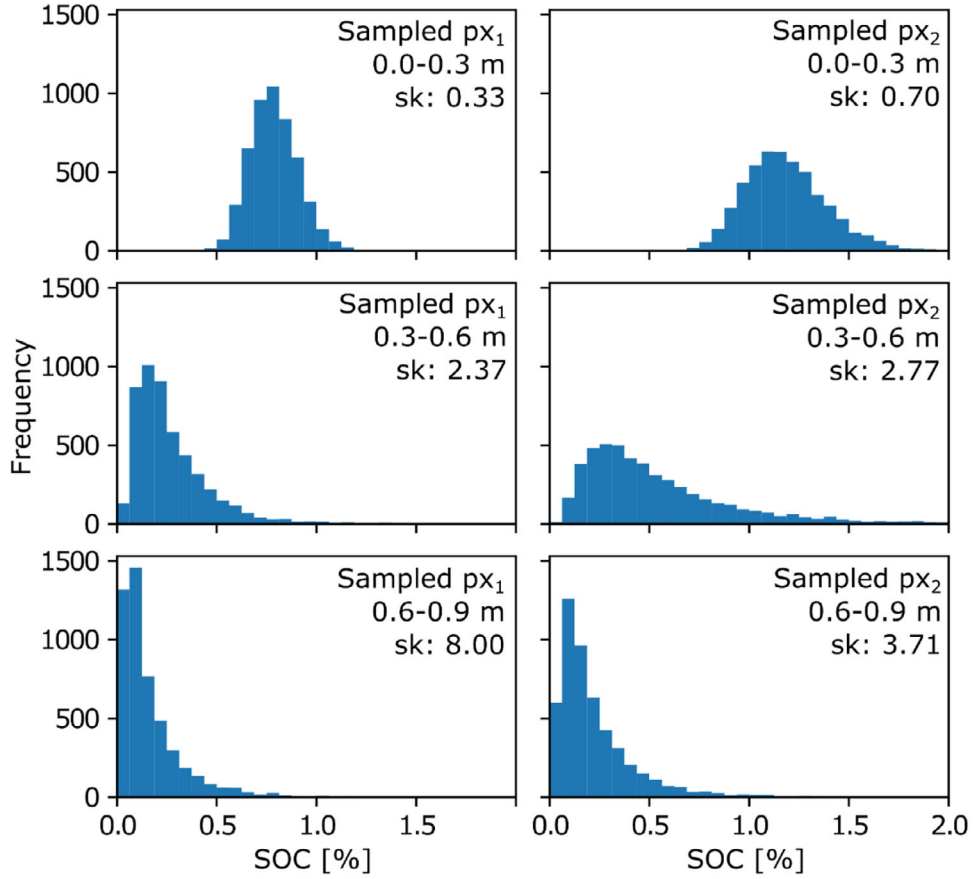

FIGURE 6 | Frequencies of the soil organic carbon (SOC) values of two pixels px_1 (left column) and px_2 (right column) across randomly generated SOC digital maps. The two pixels, px_1 and px_2, correspond to the minimum and maximum skewness of the SOC values of the layer 0.0–0.3 m, respectively.

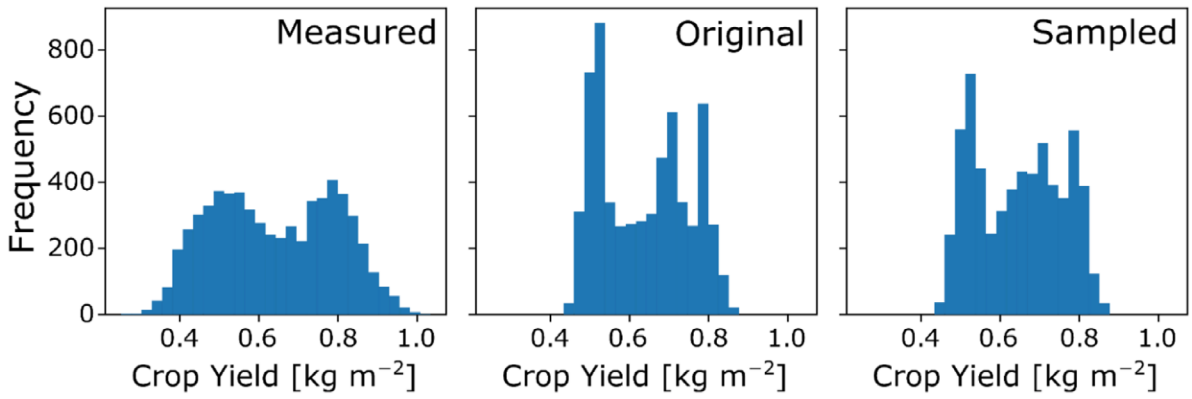

FIGURE 7 | Frequency distributions of the crop yield (kg m^{-2}) the measured, Original and one exemplary sampled scenario.

the stochastic sampling method introduces additional noise into the data. Despite this increase, the RMSE values remain within a reasonable range, demonstrating the robustness of the sampling approach in maintaining prediction accuracy. The slightly higher RMSE values for the Sampled scenarios highlight the trade-off between incorporating uncertainty and maintaining prediction accuracy. Although the introduction of noise through stochastic sampling can increase the error of an individual random sample, it also provides a more comprehensive understanding of the potential variability in crop yield predictions caused by errors in input variables, thereby enhancing the reliability of the results for decision-making processes.

Figure 8 provides a comparison between the measured crop yields and the simulated outputs. The left panel shows simulations based on the initially estimated SOC values, whereas the right panel displays simulations using one of the randomly sampled SOC digital maps. The colour coding indicates the simulations' deviation from the 1:1 axis, green for less than one standard deviation, blue for less than two deviations and red for less

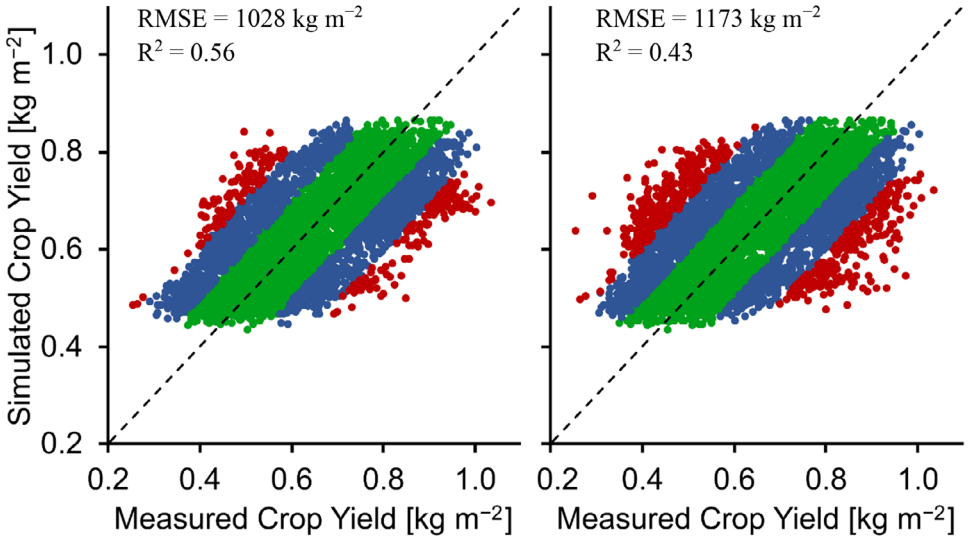

FIGURE 8 | Comparison between simulated and measured crop yield (kg m^{-2}). On the left, the simulations were realised with the firstly estimated soil organic carbon (SOC) values (%). On the right, one of the randomly sampled digital soil maps was used as SOC for the simulations. In green, the simulations with less then one standard deviation difference between the 1:1 axis, In blue and red similar, but for less then 2 and 3 deviations, respectively. RMSE, root mean square error.

than three deviations. The comparison reveals an increase in the amplitude of data distribution around the 1:1 axis, resulting in a slightly rounder shape. This suggests that the proposed stochastic sampling method introduces variability that affects the crop yield predictions, expanding the range of simulated outputs. The comparison demonstrates that although the original simulations closely align with the measured data, the incorporation of stochastic variations provides a more comprehensive view of potential yield variations. This approach allows for a more robust analysis of error impact and enhances the reliability of crop yield predictions under varying SOC conditions.

4 | Conclusions

This study introduces a novel stochastic approach to quantify the propagation of uncertainty in SOC content predictions derived from satellite imagery, utilising the MC method. Using a process-based crop model with stochastically generated SOC maps provide valuable insights into how SOC estimation errors impact crop yield forecasts. The methodology proved successful in capturing the inherent variability and uncertainty in SOC predictions, thereby offering a more reliable tool for PA.

The results demonstrated that accounting for SOC uncertainty can significantly influence crop yield predictions, highlighting the necessity of incorporating uncertainty analysis in PA practices. This approach enables better risk assessment and decision-making, ultimately contributing to more sustainable and efficient crop management strategies.

Although the study focused on a specific field in Booßen, Germany, the methodology has the potential to be applied to various agricultural settings, enhancing the generalisability and robustness of the findings. Future research should explore the incorporation of additional variables or a combination of them. In addition, expanding the application of this method to different crop types and diverse environmental conditions will help to validate its broader applicability. We have provided a code example to generate the SOC map samples at https://github.com/infoleon/StochImag.

Funding

This work was supported by the Federal Ministry of Education and Research (BMBF – Bundesministerium für Bildung und Forschung) through the project Multi-Modale Datenintegration, Domänenspezifische Methoden und KI zur Stärkung der Datenkompetenz in der Agrarforschung under Grant No. 16DKWN089.

Acknowledgments

We thank the ph-BB project for providing data support. The *ph-BB project* is funded under the *EIP-Agri funding guideline* by the *European Agricultural Fund for Rural Development* (EAFRD) and the State of Brandenburg.

Open access funding enabled and organized by Projekt DEAL.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Angelopoulou, T., N. Tziolas, A. Balafoutis, G. Zalidis, and D. Bochtis. 2019. "Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review." *Remote Sensing* 11, no. 6: 676. https://doi.org/10.3390/rs11060676.

Ankenbauer, K. J., and S. P. Loheide. 2017. "The Effects of Soil Organic Matter on Soil Water Retention and Plant Water Use in a Meadow of the Sierra Nevada, CA: Soil Organic Matter Affects Plant Water Use." *Hydrological Processes* 31, no. 4: 891–901. https://doi.org/10.1002/hyp.11070.

Asgarzadeh, H., M. R. Mosaddeghi, A. R. Dexter, A. A. Mahboubi, and M. R. Neyshabouri. 2014. "Determination of Soil Available Water for Plants:

Consistency Between Laboratory and Field Measurements." *Geoderma* 226–227: 8–20. https://doi.org/10.1016/j.geoderma.2014.02.020.

Bongiovanni, R., and J. Lowenberg-DeBoer. 2004. "Precision Agriculture and Sustainability." *Precision Agriculture* 5, no. 4: 359–387. https://doi.org/10.1023/B:PRAG.0000040806.39604.aa.

Botula, Y. D., W. M. Cornelis, G. Baert, and E. Van Ranst. 2012. "Evaluation of Pedotransfer Functions for Predicting Water Retention of Soils in Lower Congo (D.R. Congo)." *Agricultural Water Management* 111: 1–10. https://doi.org/10.1016/j.agwat.2012.04.006.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45, no. 1: 5–32. https://doi.org/10.1023/A:1010933404324.

Brisson, N. 1998. "An Analytical Solution for the Estimation of the Critical Available Soil Water Fraction for a Single Layer Water Balance Model Under Growing Crops." *Hydrology and Earth System Sciences* 2, no. 2/3: 221–231. https://doi.org/10.5194/hess-2-221-1998.

Chen, S., D. Arrouays, V. Leatitia Mulder, et al. 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.

Conrad, O., B. Bechtel, M. Bock, et al. 2015. "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4." *Geoscientific Model Development* 8, no. 7: 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Dedeoğlu, M., L. Başayiğit, M. Yüksel, and F. Kaya. 2020. "Assessment of the Vegetation Indices on Sentinel–2A Images for Predicting the Soil Productivity Potential in Bursa, Turkey." *Environmental Monitoring and Assessment* 192, no. 1: 16. https://doi.org/10.1007/s10661-019-7989-8.

De Jong Van Lier, Q. 2023. "StochHyProp: A Software to Generate Stochastic Realizations of Correlated Parameters Dedicated to Soil Hydraulic Properties." *Software Impacts* 15: 100483. https://doi.org/10.1016/j.simpa.2023.100483.

De Melo, M. L. A., L. Inforsato, E. A. R. Pinheiro, and Q. De Jong Van Lier. 2023. "Plant Available Water Predicted by a Flux-Based Approach." *Geoderma* 429: 116253. https://doi.org/10.1016/j.geoderma.2022.116253.

D'Errico, J. 2024. NearestSPD (Version 1.1.0.0): MATLAB Central File Exchange [Computer software]. The MathWorks, Inc. Retrieved June 10, 2024. https://www.mathworks.com/matlabcentral/fileexchange/42885-nearestspd.

De Vos, B., S. Lettens, B. Muys, and J. A. Deckers. 2007. "Walkley–Black Analysis of Forest Soil Organic Carbon: Recovery, Limitations and Uncertainty." *Soil Use and Management* 23, no. 3: 221–229. https://doi.org/10.1111/j.1475-2743.2007.00084.x.

Ellinger, M., I. Merbach, U. Werban, and M. Ließ. 2019. "Error Propagation in Spectrometric Functions of Soil Organic Carbon." *Soilless* 5, no. 2: 275–288. https://doi.org/10.5194/soil-5-275-2019.

European Space Agency. 2015. *SENTINEL–2 User Handbook*. European Space Agency (ESA). https://sentinel.esa.int/documents/247904/685211/sentinel-2_user_handbook.

Ewert, F., R. P. Rötter, M. Bindi, et al. 2015. "Crop Modelling for Integrated Assessment of Risk to Food Production From Climate Change." *Environmental Modelling and Software* 72: 287–303. https://doi.org/10.1016/j.envsoft.2014.12.003.

Givi, J., S. O. Prasher, and R. M. Patel. 2004. "Evaluation of Pedotransfer Functions in Predicting the Soil Water Contents at Field Capacity and Wilting Point." *Agricultural Water Management* 70, no. 2: 83–96. https://doi.org/10.1016/j.agwat.2004.06.009.

Gómez, A. M. R., Q. De Jong Van Lier, N. E. Q. Silvero, et al. 2023. "Digital Mapping of the Soil Available Water Capacity: Tool for the Resilience of Agricultural Systems to Climate Change." *Science of the Total Environment* 882: 163572. https://doi.org/10.1016/j.scitotenv.2023.163572.

Graß, R., B. Thies, K.-C. Kersebaum, and M. Wachendorf. 2015. "Simulating Dry Matter Yield of Two Cropping Systems With the Simulation Model HERMES to Evaluate Impact of Future Climate Change." *European Journal of Agronomy* 70: 1–10. https://doi.org/10.1016/j.eja.2015.06.005.

Groß, J., N. Gentsch, J. Boy, et al. 2023. "Influence of small-scale spatial variability of soil properties on yield formation of winter wheat." *Plant and Soil* 493, no. 1–2: 79–97. https://doi.org/10.1007/s11104-023-06212-2.

Heuvelink, G. B. M., P. A. Burrough, and A. Stein. 1989. "Propagation of Errors in Spatial Modelling With GIS." *International Journal of Geographical Information Systems* 3, no. 4: 303–322. https://doi.org/10.1080/02693798908941518.

Higham, N. J. 1988. "Computing a Nearest Symmetric Positive Semidefinite Matrix." *Linear Algebra and Its Applications* 103: 103–118. https://doi.org/10.1016/0024-3795(88)90223-6.

Ho, T. K. 1995. "Random Decision Forests." *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1: 278–282. https://doi.org/10.1109/ICDAR.1995.598994.

Holzworth, D. P., N. I. Huth, P. G. deVoil, et al. 2014. "APSIM—Evolution Towards a New Generation of Agricultural Systems Simulation." *Environmental Modelling and Software* 62: 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009.

Inforsato, L., and Q. de Jong van Lier. 2021. "Polynomial Functions to Predict Flux-Based Field Capacity From Soil Hydraulic Parameters." *Geoderma* 404: 115308. https://doi.org/10.1016/j.geoderma.2021.115308.

Jones, J. W., G. Hoogenboom, C. H. Porter, et al. 2003. "The DSSAT Cropping System Model." *European Journal of Agronomy* 18, no. 3–4: 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7.

Kersebaum, K. C. 2007. "Modelling Nitrogen Dynamics in Soil–Crop Systems With HERMES." *Nutrient Cycling in Agroecosystems* 77, no. 1: 39–52. https://doi.org/10.1007/s10705-006-9044-8.

Kohavi, R. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." *International Joint Conference on Artificial Intelligence* 14, no. 2: 1137–1145.

Kumari, A., and S. Karthikeyan. 2023. "Sentinel–2 Data for Land Use/Land Cover Mapping: A Meta-Analysis and Review." *SN Computer Science* 4, no. 6: 815. https://doi.org/10.1007/s42979-023-02214-0.

Lagacherie, P., and A. B. McBratney. 2006. "Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping." In *Developments in Soil Science*, edited by M. Busse, C. P. Giardina, D. M. Morris, and D. S. Page-Dumroese, 3–22. Elsevier. https://doi.org/10.1016/S0166-2481(06)31001-X.

Lehtinen, T., N. Schlatter, A. Baumgarten, et al. 2014. "Effect of Crop Residue Incorporation on Soil Organic Carbon and Greenhouse Gas Emissions in European Agricultural Soils." *Soil Use and Management* 30, no. 4: 524–538. https://doi.org/10.1111/sum.12151.

Lewis, J. P. 1995. "Fast Normalized Cross-Correlation." Vision Interface, 1995 (VI'95), pp. 120–123. Canadian Image Processing and Pattern Recognition Society.

Li, H., C. A. Calder, and N. Cressie. 2007. "Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model." *Geographical Analysis* 39, no. 4: 357–375. https://doi.org/10.1111/j.1538-4632.2007.00708.x.

McBratney, A. B., M. L. Mendonça Santos, and B. Minasny. 2003. "On Digital Soil Mapping." *Geoderma* 117, no. 1–2: 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Minasny, B., and A. B. McBratney. 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., and A. B. McBratney. 2018. "Limited Effect of Organic Matter on Soil Available Water Capacity." *European Journal of Soil Science* 69, no. 1: 39–47. https://doi.org/10.1111/ejss.12475.

Mondal, A., D. Khare, S. Kundu, S. Mondal, S. Mukherjee, and A. Mukhopadhyay. 2017. "Spatial Soil Organic Carbon (SOC) Prediction by Regression Kriging Using Remote Sensing Data." *Egyptian Journal of Remote Sensing and Space Sciences* 20, no. 1: 61–70. https://doi.org/10.1016/j.ejrs.2016.06.004.

Nendel, C., M. Berg, K. C. Kersebaum, et al. 2011. "The MONICA Model: Testing Predictability for Crop Growth, Soil Moisture and Nitrogen Dynamics." *Ecological Modelling* 222, no. 9: 1614–1625. https://doi.org/10.1016/j.ecolmodel.2011.02.018.

Nendel, C., K. C. Kersebaum, W. Mirschel, and K. O. Wenkel. 2014. "Testing Farm Management Options as Climate Change Adaptation Strategies Using the MONICA Model." *European Journal of Agronomy* 52: 47–56. https://doi.org/10.1016/j.eja.2012.09.005.

Papoulis, A. 1991. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill.

Parvizi, Y., and S. Fatehi. 2025. "Geospatial Digital Mapping of Soil Organic Carbon Using Machine Learning and Geostatistical Methods in Different Land Uses." *Scientific Reports* 15, no. 1: 4449. https://doi.org/10.1038/s41598-025-88062-9.

Pedometrics. 2018. In *Progress in Soil Science* edited by A. B. McBratney, B. Minasny, and U. Stockmann, Springer International Publishing. https://doi.org/10.1007/978-3-319-63439-5.

Pinheiro, E. A. R., and Q. de Jong van Lier. 2021. "Propagation of Uncertainty of Soil Hydraulic Parameterization in the Prediction of Water Balance Components: A Stochastic Analysis in Kaolinitic Clay Soils." *Geoderma* 388: 114910. https://doi.org/10.1016/j.geoderma.2020.114910.

Plante, A. F., and W. J. Parton. 2007. "The Dynamics of Soil Organic Matter and Nutrient Cycling." In *Soil Microbiology, Ecology and Biochemistry*, edited by E. A. Paul, 433–467. Elsevier. https://doi.org/10.1016/B978-0-08-047514-1.50020-2.

Qiao, J., Y. Zhu, X. Jia, L. Huang, and M. Shao. 2019. "Pedotransfer Functions for Estimating the Field Capacity and Permanent Wilting Point in the Critical Zone of the Loess Plateau, China." *Journal of Soils and Sediments* 19, no. 1: 140–147. https://doi.org/10.1007/s11368-018-2036-x.

Rai, R. K., V. P. Singh, and A. Upadhyay. 2017. "Soil Analysis." In *Planning and Evaluation of Irrigation Projects*, edited by V. P. Singh, 505–523. Elsevier. https://doi.org/10.1016/B978-0-12-811748-4.00017-0.

Saltelli, A., and R. Bolado. 1998. "An Alternative Way to Compute Fourier Amplitude Sensitivity Test (FAST)." *Computational Statistics and Data Analysis* 26, no. 4: 445–460. https://doi.org/10.1016/S0167-9473(97)00043-1.

Söderström, M., G. Sohlenius, L. Rodhe, and K. Piikki. 2016. "Adaptation of Regional Digital Soil Mapping for Precision Agriculture." *Precision Agriculture* 17, no. 5: 588–607. https://doi.org/10.1007/s11119-016-9439-8.

Taylor, J. A. 2023. "Precision Agriculture." In *Encyclopedia of Soils in the Environment*, edited by D. Hillel, 710–725. Elsevier. https://doi.org/10.1016/B978-0-12-822974-3.00261-5.

Tóth, B., M. Weynants, A. Nemes, A. Makó, G. Bilas, and G. Tóth. 2015. "New Generation of Hydraulic Pedotransfer Functions for Europe." *European Journal of Soil Science* 66, no. 1: 226–238. https://doi.org/10.1111/ejss.12192.

Tran, K. H., H. K. Zhang, J. T. McMaine, X. Zhang, and D. Luo. 2022. "10 m Crop Type Mapping Using Sentinel-2 Reflectance and 30 m Cropland Data Layer Product." *International Journal of Applied Earth Observation and Geoinformation* 107: 102692. https://doi.org/10.1016/j.jag.2022.102692.

Turanyi, T. 1990. "Sensitivity Analysis of Complex Kinetic Systems. Tools and Applications." *Journal of Mathematical Chemistry* 5, no. 3: 203–248. https://doi.org/10.1007/BF01166355.

Vågen, T.-G., L. A. Winowiecki, C. Neely, S. Chesterman, and M. Bourne. 2018. "Spatial Assessments of Soil Organic Carbon for Stakeholder Decision-Making—A Case Study From Kenya." *Soilless* 4, no. 4: 259–266. https://doi.org/10.5194/soil-4-259-2018.

Vaudour, E., A. Gholizadeh, F. Castaldi, et al. 2022. "Satellite Imagery to Map Topsoil Organic Carbon Content Over Cultivated Areas: An Overview." *Remote Sensing* 14, no. 12: 2917. https://doi.org/10.3390/rs14122917.

Wadoux, A. M. C., B. Minasny, and A. B. McBratney. 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.

Wallor, E., K.-C. Kersebaum, D. Ventrella, et al. 2018. "The Response of Process-Based Agro-Ecosystem Models to Within-Field Variability in Site Conditions." *Field Crops Research* 228: 1–19. https://doi.org/10.1016/j.fcr.2018.08.021.

Wilks, D. S. 1997. "Resampling Hypothesis Tests for Autocorrelated Fields." *Journal of Climate* 10, no. 1: 65–82. https://doi.org/10.1175/1520-0442(1997)010⟨0065:RHTFAF⟩2.0.CO;2.

Willmott, C., and K. Matsuura. 2005. "Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance." *Climate Research* 30: 79–82. https://doi.org/10.3354/cr030079.

Wuest, S. B., and N. Durfee. 2024. "Temporal Variability Is a Major Source of Uncertainty in Soil Carbon Measurements." *Soil Science Society of America Journal* 88, no. 3: 830–845. https://doi.org/10.1002/saj2.20660.

Yuzugullu, O., N. Fajraoui, A. Don, and F. Liebisch. 2024. "Satellite-Based Soil Organic Carbon Mapping on European Soils Using Available Datasets and Support Sampling." *Science of Remote Sensing* 9: 100118. https://doi.org/10.1016/j.srs.2024.100118.

Zhou, T., Y. Geng, C. Ji, et al. 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." *Science of the Total Environment* 755: 142661. https://doi.org/10.1016/j.scitotenv.2020.142661.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: jpln70038-sup-0001-SuppMat.pdf.