



remote sensing

IMPACT
FACTOR
4.2

CITESCORE
8.3

Article

Exploiting Soil and Remote Sensing Data Archives for 3D Mapping of Multiple Soil Properties at the Swiss National Scale

Felix Stumpf, Thorsten Behrens, Karsten Schmidt and Armin Keller

Special Issue

Recent Advances in Remote Sensing of Soil Science

Edited by

Dr. Nikolaos L. Tsakiridis, Dr. Uta Heiden and Dr. Nikolaos Tziolas



<https://doi.org/10.3390/rs16152712>



Article

Exploiting Soil and Remote Sensing Data Archives for 3D Mapping of Multiple Soil Properties at the Swiss National Scale

Felix Stumpf *, Thorsten Behrens, Karsten Schmidt  and Armin Keller

Competence Center for Soils, School of Agricultural, Forest and Food Sciences (HAFL), Bern University of Applied Sciences, Länggasse 85, 3052 Zollikofen, Switzerland

* Correspondence: felixmichael.stumpf@bfh.ch; Tel.: +41-78-974-3066

Abstract: Soils play a central role in ecosystem functioning, and thus, mapped soil property information is indispensable to supporting sustainable land management. *Digital Soil Mapping (DSM)* provides a framework to spatially estimate soil properties. However, broad-scale *DSM* remains challenging because of non-purposively sampled soil data, large data volumes for processing extensive soil covariates, and high model complexities due to spatially varying soil–landscape relationships. This study presents a three-dimensional *DSM* framework for Switzerland, targeting the soil properties of clay content (*Clay*), organic carbon content (*SOC*), pH value (*pH*), and potential cation exchange capacity (*CEC_{pot}*). The *DSM* approach is based on machine learning and a comprehensive exploitation of soil and remote sensing data archives. *Quantile Regression Forest* was applied to link the soil sample data from a national soil data base with covariates derived from a *LiDAR*-based elevation model, from climate raster data, and from multispectral raster time series based on satellite imagery. The covariate set comprises spatially multiscale terrain attributes, climate patterns and their temporal variation, temporarily multiscale land use features, and spectral bare soil signatures. Soil data and predictions were evaluated with respect to different landcovers and depth intervals. All reference soil data sets were found to be spatially clustered towards croplands, showing an increasing sample density from lower to upper depth intervals. According to the R^2 value derived from independent data, the overall model accuracy amounts to 0.69 for *Clay*, 0.64 for *SOC*, 0.76 for *pH*, and 0.72 for *CEC_{pot}*. Reduced model accuracies were found to be accompanied by soil data sets showing limited sample sizes (e.g., *CEC_{pot}*), uneven statistical distributions (e.g., *SOC*), and low spatial sample densities (e.g., woodland subsoils). Multiscale terrain covariates were highly influential for all models; climate covariates were particularly important for the *Clay* model; multiscale land use covariates showed enhanced importance for modeling *pH*; and bare soil reflectance was a major driver in the *SOC* and *CEC_{pot}* models.

Keywords: 3D Digital Soil Mapping; multiscale spatial and temporal data; multispectral raster timeseries; Landsat; Sentinel; Switzerland



Citation: Stumpf, F.; Behrens, T.; Schmidt, K.; Keller, A. Exploiting Soil and Remote Sensing Data Archives for 3D Mapping of Multiple Soil Properties at the Swiss National Scale. *Remote Sens.* **2024**, *16*, 2712. <https://doi.org/10.3390/rs16152712>

Academic Editor: Dominique Arrauys

Received: 29 May 2024

Revised: 19 July 2024

Accepted: 21 July 2024

Published: 24 July 2024



Copyright: © 2024 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/>).

1. Introduction

1.1. Soil Information Needs

Soils are an intrinsic component of terrestrial ecosystems, being formed and exposed at the intersections of the lithosphere, biosphere, atmosphere, and hydrosphere [1]. Therefore, soils determine the provision and regulation of numerous ecosystem services and functions related to the quality of air and water, food production, climate change, and biodiversity, among others [2,3]. The central role of soils for ecosystem functioning entails the need for extensive soil property information to support sustainable land management [4–6].

1.2. Digital Soil Mapping

Digital Soil Mapping (DSM) emerged over the last two decades as a labor- and cost-efficient, quantitative, reproducible, and updateable method to map soil properties [7].

DSM describes the creation of spatial soil information systems based on numerical models, linking soil observations with covariates to infer soil properties and associated uncertainties at unsampled locations [8–10]. Conceptually, *DSM* is rooted in Jenny's *clorpt* model for soil formation, where soil at a specific location is described by climate, vegetation, relief, parent material, and time [11]. The *DSM* principle acknowledges the spatial variation of soils, where a soil property is described as a function of the so-called *scorpan* factors, thus covariates that are related to soil information (s), climate (c), vegetation (o), relief (r), parent material (p), time (a), and the geographic position (n) [12].

In the past decade, *DSM* has been increasingly applied at broader scales to address extensive and soil-related environmental issues, as well as to support local and regional soil sampling campaigns [13–16]. However, *DSM* at broader scales often requires coping with the spatially varying complexity and non-stationarity of the soil-landscape relationship, large data volumes, as well as sparse, unevenly distributed, and spatially clustered soil sample data [10,17].

Broad-scale *DSM* most commonly utilizes machine learning, a technique in data science that applies statistical and mathematical methods to correlate data and build models with potentially high computational efficiency [5,18,19]. Among the large group of machine learning algorithms, tree-based ensemble models, such as *Quantile Regression Forest (QRF)*, a variant of *Random Forest (RF)*, are most popular [20–24]. Both *RF* and *QRF* aggregate the results of numerous randomized regression trees for the final predictions. The randomization is based on a bootstrap sample for each tree and a random selection of covariates for each tree node. *RF* records the mean of the conditional distribution of the predicted target variable for all end nodes of a fitted tree. Instead, *QRF* preserves the entire conditional distribution for all end nodes, allowing us to infer a lower and upper quantile and, thus, a prediction interval (*PI*) for any location. In *DSM*, *PIs* are usually leveraged to describe the prediction uncertainty, indicated by the range in which the true value is expected to be related to a given probability. Several studies have approved *QRF* to produce plausible patterns of uncertainty variations [10,25]. Moreover, the success of the tree-based ensemble models can be attributed to further characteristics of the algorithms. The models are reported to be relatively resistant to overfitting, varying sample sizes, multicollinearity, and high data dimensionality [26–28]. Furthermore, no assumptions on data distributions are required; only a small number of hyperparameters need to be tuned, while the algorithms capture nonlinear relationships [29–31]. The combination of these benefits enables a robust representation of the soil-landscape relationship based on the *scorpan* concept, as well as efficient workflows with respect to data preprocessing, model calibration, and prediction [18]. Recently, tree-based ensemble models have been increasingly applied for broad-scale *DSM* [32–34]. Moreover, numerous *DSM* studies reported the performance superiority of these models compared to other machine learning algorithms, such as *Support Vector Machines*, *Multiple Linear Regression*, *K-Nearest Neighbor*, or *Artificial Neural Networks* [35–37].

According to a recent review by Chen et al. [5], broad-scale *DSM* studies of the past two decades predominantly applied 2D model designs, thus calibrating a model referring to a single soil depth layer only [38–40]. The recently prevalent 2.5D design acknowledges the depth dimension of soil by combining independent models that are fitted for different depth layers [41–43]. Most recently, an increasing number of *DSM* studies have applied 3D model designs that incorporate the depth dimension more explicitly by spatially fitting depth functions or integrating depth as a covariate [44–47]. The 3D design, which applies depth as a covariate, is associated with a straightforward and flexible workflow, allowing predictions at any location and depth based on a single calibrated model. Moreover, the calibration sample size is maximized, while complex assumptions about depth trends are avoided [33,34,48]. However, Nauman and Duniway (2019) [49] stated that these 3D designs might lead to increased prediction uncertainties for soil properties with an enhanced depth gradient. Ma et al. [50] found unrealistically stepped depth functions. Nevertheless, model performances between 2.5D and 3D designs are

reported to be generally comparable, while any design preference might be case-specific, depending on the soil property and sample size [5,32,50].

Besides the advent of machine learning *DSM*, the increasing availability of air- and spaceborne remote sensing data as well as high-performance computing facilities has initiated major progress for broad-scale *DSM* [51–53]. Particularly, airborne *LiDAR*-based digital elevation models (*DEM*) in high spatial resolution and multispectral satellite imagery in high spatio-temporal resolution, combined with methodological advancements, allow for the extensive and more accurate processing of *scorpan* covariates [54–56].

1.3. Terrain Covariates for *DSM*

Generally, the *scorpan* factors determine soil formation by describing soil-forming processes that interact and operate across multiple spatial scales [57–59]. In this context, Behrens et al. [60] emphasized that spatially multiscale terrain attributes from high-resolution *DEMs* represent proxies for multiple *scorpan* factors related to gravimetry, hydrology, climate, land use, and lithology. This has been acknowledged by several *DSM* studies, proposing methods to process multiscale terrain covariates based on resampling techniques, expanding convolution kernels, or wavelet transformations [60–63]. Although the aforementioned studies revealed improved model performances, the majority of *DSM* studies make use of terrain covariates at a single spatial scale only, and thus, soil-forming processes operating on a different scale might be inadequately described [64–66].

1.4. Multispectral Covariates for *DSM*

Recently, long-term multispectral raster time series, derived from imagery archives of the *Landsat* (*LA*) and *Sentinel-2* (*S2*) satellite missions, have been leveraged to process *scorpan* covariates related to soil surface and land use characteristics [67–69].

Referring to soil surface characteristics, the spectral signature of bare soil surfaces correlates with mineralogical, physical, and chemical soil properties [70–72]. These bare soil covariates are approved to be expedient for *DSM*, particularly concerning topsoil clay content, pH, organic carbon, and cation exchange capacity [40,56,73]. The processing of bare soil covariates comprises a filtering of the multispectral time series data for pixels that represent bare soil surfaces and a subsequent aggregation of the filtered raster stack across time [38,74,75].

Referring to land use characteristics, the spectral signature of vegetation, as a land use indicator, correlates with biological, chemical, physical, and hydraulic soil properties [20,76,77]. Related *scorpan* covariates can be processed as averages across time based on vegetation indices derived from the multispectral raster time series [28,78,79]. These land use covariates describe temporarily static vegetation conditions for one defined period, are commonly applied for *DSM*, and are being approved to be particularly relevant for mapping organic carbon, pH, carbonates, texture, and cation exchange capacity [19,80,81]. However, vegetation characteristics tend to change over time and across multiple temporal scales [82–84]. For example, the temporal variability of vegetation differs across seasonal, inter-annual, or inter-decadal time scales, and thus, it is related to different and interacting soil-forming processes [54,85]. Similar to multiscale terrain analysis, the vast majority of *DSM* studies disregard this temporal multiscale character of land use by using temporarily static land use covariates based on a single time scale only. However, Heuvelink et al. [82] applied *DSM* to map soil organic carbon in space and time, using temporarily multiscale land use covariates based on a seasonal vegetation index processed for each year and retrospectively across multiple time scales, defined by an exponential decay function.

1.5. Study Outline and Objectives

This study provides a broad-scale 3D *DSM* framework for Switzerland, addressing all crop-, grass-, and woodland areas ('*Cropland*', '*Grassland*', and '*Woodland*') and three depth intervals ('*Topsoil*': 0–30 cm; '*Subsoil*': 30–60 cm; and '*Deep Subsoil*': 60–120 cm). The target soil properties are clay content (*Clay*, %), soil organic carbon concentration (*SOC*, %), pH

value (pH , $CaCl_2$), and potential cation exchange capacity (CEC_{pot} , $mm_c \cdot kg^{-1}$), which were derived from archived soil sample data sets. *Scorpan* covariates were derived by exploiting remote sensing data archives, thus a *LiDAR*-based *DEM*, two multispectral raster time series based on *LA* and *S2* imagery, respectively, as well as climate raster data. The covariate set comprises spatially multiscale terrain attributes, temporarily multiscale land use features, spectral bare soil signatures, as well as climate patterns and their temporal variation. We applied the machine learning algorithm *QRF* for a 3D model design, linking each soil sample set with the *scorpan* covariates and using soil depth as an additional covariate. Hence, the introduced framework for broad-scale *DSM* combines recent advancements related to the engineering of *scorpan* covariates with a state-of-the-art 3D model design. The latter allows us to efficiently derive map predictions and associated uncertainties at any location and depth based on a single calibrated model. Moreover, the evaluation scheme of the applied 3D model design allows for the derivation of model accuracies across landcovers and depth intervals based on independent soil data. The specific objectives of the study are defined by (i) analyzing soil sample data sets with respect to spatial sample density and statistical distributions' (ii) evaluating predicted maps for *Clay*, *SOC*, *pH*, and *CEC_{pot}*, referring to landcover and depth intervals; and (iii) assessing the influence of covariates related to terrain, bare soil, land use, and climate on model performances.

2. Materials and Methods

2.1. Study Area

Switzerland is a landlocked country in the center of Europe, covering 41,291 km². Altitudes range from 196 m to 4634 m (a.s.l.). Annual precipitation ranges from 530 mm to 3300 mm, while the mean temperature in January amounts to 1 °C in January and 17 °C in July [86–88]. With an areal proportion of 45%, the largest proportion of the Swiss territory is devoted to agriculture. This encompasses croplands (9%; 'Cropland') and grasslands (36%; 'Grassland'), which include permanent grasslands (14%), and extensively managed alpine grasslands (22%). Forest areas ('Woodland') and settlements cover 28% and 8% of the study area, respectively. Approximately 19% of the Swiss territory is unproductive due to rock, water, and ice surfaces. The study area comprises 'Cropland', 'Grassland', and 'Woodland', covering approximately 30,394 km² (Figure 1) [89].

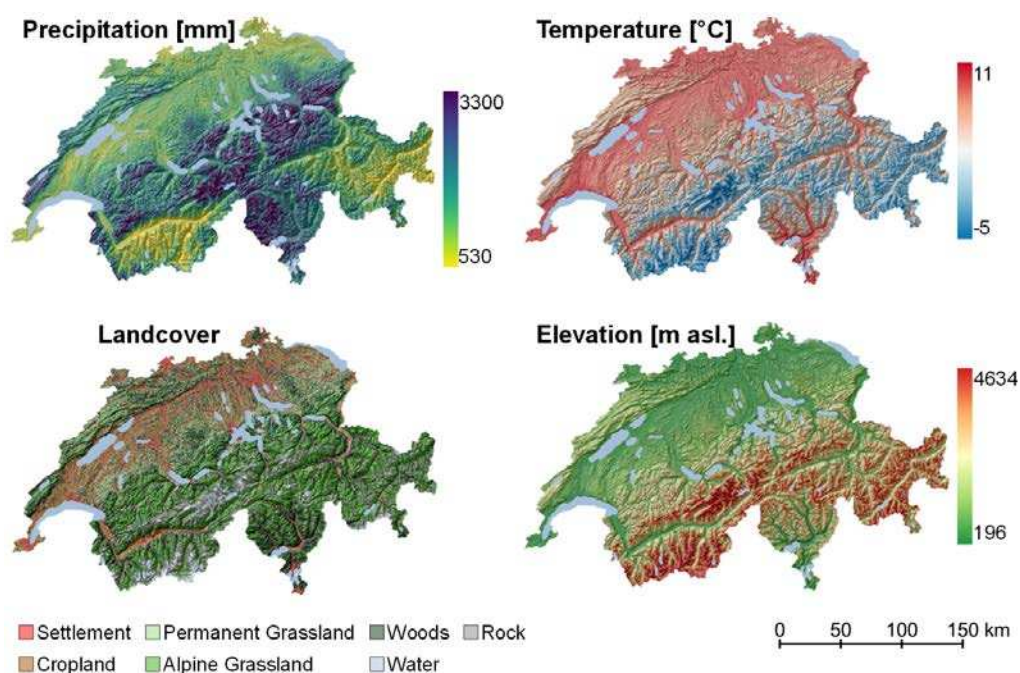


Figure 1. Switzerland with respect to landcover and topography, as well as to precipitation and temperature means across 1980 to 2022.

2.2. Soil Model Design

We used QRF to link soil sample data for *Clay*, *SOC*, *pH*, and *CEC_{pot}* with *scorpan* covariates. The final prediction was derived based on the median of the conditional distribution recorded for each end node of a fitted tree. The prediction uncertainty is based on the QRF prediction intervals, defined by the 5%- and 95%-quantiles of the conditional distribution. The final prediction uncertainty (*RPI*) was then defined by the ratio between the *PI* and the median of the conditional distribution (cf. 1.2).

The subset of sample data, which is excluded for the calibration of a single QRF tree, allows us to derive the model accuracy as the average across all trees (*OOB_{error}*). The hyperparameters '*mtry*' and '*numtrees*' define the size of the covariate subset at each node and the number of trees in the ensemble, respectively. These hyperparameters were determined by grid tuning based on the *OOB_{error}* and using the *Coefficient of Determination* (R^2 ; Equation (1)) as an accuracy indicator [90]. The R^2 is a measure of covariation between observed and predicted values, as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i^{\text{observed}} - y_i^{\text{predicted}})^2}{\sum_{i=1}^n (y_i^{\text{observed}} - \bar{y}^{\text{observed}})^2} \quad (1)$$

The *scorpan* covariates represent extensively available raster data sets related to terrain, climate, land use, and the spectral signature of bare soil. All covariates were resampled to a common raster cell size of 30×30 m and assigned to the soil sample data sets. Covariates related to terrain, climate, and bare soil were assumed to be static in time and assigned to the soil data based on the geographic position only. Covariates for land use describe temporarily dynamic phenomena and are annually available. These covariates were assigned to the soil data based on the geographic position and the year of soil data acquisition.

The soil depth served as an additional covariate and was primarily harmonized to the standard depth intervals of '*Topsoil*' (0–30 cm), '*Subsoil*' (30–60 cm), and '*Deep Subsoil*' (60–120 cm). The depth harmonization is based on fitting an *equal-area quadratic spline* down the soil profile of each sample location and soil property. The soil property value within each depth interval was defined by the mean of the fitted spline [91,92]. Since the fitting of a spline requires more than one data value, sample values at locations with only one sampled depth were transferred to their associated standard depth interval [93].

The accuracy evaluation is based on five-fold stratified random data splitting. Thus, the soil data sets were stratified by the depth intervals and landcover classes '*Cropland*', '*Grassland*', and '*Woodland*'. Within each stratum, the soil data sets were randomly split by a ratio of 80/20 to form a calibration set and an independent validation set not used in the model (*IV*) for all strata. The R^2 (Equation (1)) and the *Root Mean Square Error* (*RMSE*; Equation (2)), absolute measures of the average error, were used as accuracy indicators.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^{\text{predicted}} - y_i^{\text{observed}})^2}{n}} \quad (2)$$

The R^2 and *RMSE* were derived using the *OOB_{error}*, a 10-fold cross-validation approach (*CV10*), and *IV*. Moreover, both accuracy indicators and, additionally, the uncertainty indicator *RPI* were derived for each stratum of the *IV*. Based on different data splits, the modeling was repeated fivefold to generate independent submodels. The overall model accuracy and uncertainty were defined by the mean of the respective indicators across all submodels. Additionally, we evaluated the model robustness by analyzing the standard deviation for each accuracy indicator across all submodels (*SD_{R2}*; *SD_{RMSE}*), as well as by comparing the results of the *OOB_{error}*, *CV10*, and *IV*.

The best submodel was defined by the closest distance to the mean across all submodels, referring to R^2 and *RMSE* based on the *IV*. This model was then used to derive spatial

predictions of the soil properties and their respective *RPI* for all depth intervals using the most recent temporarily dynamic covariates.

A covariate importance analysis was conducted to assess the influence of the covariates on the model performance. Covariates related to terrain, climate, land use, and bare soil were each defined as a group. In the next step, the performance of additional calibrated models, for which all covariates of a particular group were randomly permuted, was compared to the performance of the original models, which were calibrated with the original data [94]. The decline in model performance therefore indicates the importance of the particular covariate group. For this specific analysis, the model performance was assessed by R^2 based on the OOB_{error} . The covariate importance was derived for each covariate group and submodel, while the final covariate importance was defined by the mean across all submodels.

2.3. Soil Sample Data

The soil sample data for *Clay* (%), *SOC* (%), *pH* (CaCl₂), and *CEC_{pot}* (mm_ckg⁻¹) stem from two sources. First, the 'National Soil Information System NABODAT' collects, merges, and maintains soil profile data from various soil surveys at regional scales across Switzerland since the 1960s. We used version 6 of the 'National Soil Dataset' [95]. Second, the 'Swiss Biodiversity Monitoring' (BDM), which includes a soil survey conducted between 2011 and 2015 based on a regular grid network of 6 by 4 km spanning across Switzerland. At around 1200 sampling plots, 2–4 topsoil samples (0–20 cm) were collected from the outer rim in divergent directions from a center point and finally averaged [96–98].

The soil sample data represent a conglomerate of the most diverse soil surveys of the past six decades and show increased heterogeneity with respect to underlying methodological and technical standards. This requires several data harmonization and cleansing steps. Hence, soil property labels and units were standardized, while data from surveys prior to 1985 were excluded. Profile data with overlapping soil horizons, duplicates, and implausible values were identified and removed. Data associated with sealed surfaces, according to current land cover models and referring to the time of data acquisition, indicate uncertain georeferencing and were removed. With respect to all soil properties, the soil sample data comprise 17,283 unique sample locations (n_{xy}).

Clay sample data were analyzed according to the sedimentation method [99]. The majority of the *SOC* data (81%) were analyzed using the Swiss standard method based on wet oxidation and retitration of potassium dichromate [100]. Moreover, 19% of the *SOC* sample data were analyzed by dry combustion using a CN-analyzer [101]. Since *SOC* values determined by wet oxidation are generally lower compared to an analysis by dry combustion, we adapted these samples by applying a conversion factor of 1.18 [102]. Moreover, *SOC* values from organic soils are inherently higher compared to *SOC* values from mineral soils, while available data prevent a definite distinction. Thus, we removed *SOC* data with values exceeding three times the interquartile range from the mean [79,103]. Sample data for *pH* were measured potentiometrically in a 0.01 M CaCl₂ solution [104]. The majority of the *CEC_{pot}* data (89%) were analyzed by extracting exchangeable cations (Ca, Mg, K, and Na⁺) in a 0.1 M BaCl₂ or a 0.1 M HCl + H₂SO₄ solution. Elemental concentrations were measured by atomic absorption spectroscopy (AAS) or a flame photometer. The summarized exchangeable charges of Ca, Mg, K, Na, and H correspond to *CEC_{pot}* [104]. Exchangeable cations of fewer *CEC_{pot}* data (11%) were extracted in a 1 M NH₄Cl solution, including for Al and Fe, while atomic emission spectroscopy (ICP-AES) was applied to measure elemental concentrations [105–107].

Finally, the soil sample data were analyzed according to their spatial sample density (*SSD*; n_{xy} *km⁻²) and statistical distributions with respect to landcover classes and depth intervals.

2.4. Terrain Covariates

The data source for terrain covariates is the digital elevation model (DEM) *swissALTI*^{3D} with a spatial resolution of 2×2 m. For areas below 2000 m a.s.l., *swissALTI*^{3D} is based on airborne *LiDAR* measurements and shows an accuracy of 0.5 m with respect to position and elevation. For areas above 2000 m a.s.l., the *swissALTI*^{3D} is predominantly based on stereo orthophoto correlation of overlapping aerial images, showing an accuracy of 1–3 m [88].

We conducted a multiscale terrain analysis, where terrain attributes were derived at different spatial scales using mixed scaling [108]. Therefore, the DEM was rescaled to multiple versions of increasingly coarser resolutions. Then, terrain attributes were derived from each of the rescaled DEMs. In a final step, the terrain attributes were rescaled to the initial resolution of the DEM (Figure S1) [60,108,109].

We spatially extended the *swissALTI*^{3D} data by 30 km across the Swiss external border to reduce artefacts that would arise from processing terrain attributes close to the border. For that purpose, we used elevation data from the Shuttle Topography Radar Mission (SRTM), available at a spatial resolution of 30×30 m [110]. The *swissALTI*^{3D} and the SRTM were resampled to a consistent spatial resolution of 8×8 m using bilinear interpolation and merged to form the pivotal database for the multiscale terrain analysis.

In total, we derived 11 terrain attributes (Table 1) at 14 different spatial scales between 8 m and 4880 m. The selected terrain attributes were approved to be relevant for the target soil properties [15,50,111]. The spatial scales represent the initial resolution of the DEM factorized by the Fibonacci numbers n ($n = 1, 2, \dots, 610$). In total, the multiscale terrain analysis resulted in 154 terrain covariates.

Table 1. List of terrain attributes used for the multiscale terrain analysis.

Terrain Attribute	Reference
Elevation (a.s.l.) [m]	[88]
Slope Inclination [°]	[112]
Terrain Ruggedness Index [-]	[113]
Terrain Roughness Index [-]	[114]
Topographic Position Index [-]	[115]
Profile Curvature [-]	[112]
Planform Curvature [-]	[112]
Flow Accumulation, log ₁₀ [-]	[116]
Convergence Index [-]	[117]
Eastness [-], sin (aspect in rad)	[114]
Northness [-], cos (aspect in rad)	[114]

2.5. Climate Covariates

Climate covariates are based on annual raster data for precipitation totals, mean relative sunshine duration, as well as means of daily average, maximum, and minimum temperatures. These data sets are available at a spatial resolution of 1000×1000 m for the period from 1980 to 2021 (Table 2) [87]. According to Frei et al. [118–120], the precipitation and temperature grids were interpolated based on a network of 430 rain gauges and 125 temperature stations across Switzerland. The relative sunshine duration represents the ratio between the effective sunshine and the maximally possible sunshine duration, assuming that no clouds are present. The grids for relative sunshine duration are based on 70 measurement stations and a clearness index derived from *Meteosat* satellite imagery. We derived climate covariates as the mean and the mean interannual change in each climate parameter across 1980 to 2021 (Figure S2). In total, the analysis results in 10 individual climate covariates.

Table 2. List of annually available (1980–2021) climate grid data used to derive long-term means and mean interannual changes.

Climate Raster	Reference
Precipitation totals [mm]	[118]
Temperature daily average [°C]	[119]
Temperature daily maximum [°C]	[119]
Temperature daily minimum [°C]	[119]
Relative sunshine duration [%]	[120]

2.6. Spectral Covariates

2.6.1. Multispectral Raster Time Series

The data source for covariates related to land use consists of two raster time series based on *LA* and *S2* data, respectively. The data source for covariates related to the spectral signature of bare soil is based only on the *LA* raster time series [121,122].

The *LA* raster time series was compiled from all available scenes of all *LA* sensors (*TM*, *ETM+*, and *OLI*) for the period from 1985 to 2021. The *LA* sensors record a blue (*BLUE*), green (*GREEN*), red (*RED*), and near-infrared (*NIR*) band, as well as two shortwave-infrared (*SWIR1*, *SWIR2*) bands at a spatial resolution of 30 m × 30 m. *LA* scenes of the sensors *TM* and *ETM+* were readily corrected to surface reflectance according to the *Landsat Ecosystem Disturbance Adaptive Processing System* (LEDAPS) [123]. *LA* scenes of the *OLI* sensor were corrected according to the *Land Surface Reflectance Code* (LaSRC) [124]. All pixels covered by clouds and cloud shadows were masked using the *CFmask* algorithm [125]. Since there are differences in the reflective wavelengths of the *LA* sensors *TM/ETM+* and *OLI*, we harmonized the data using the *RMA* (Reduced Major Axis) regression coefficients provided by Roy et al. [126].

The spectral vegetation indexes *Normalized Difference Vegetation Index* (*NDVI*; Equation (3)) and *Normalized Burn Ratio 2* (*NBR2*; Equation (4)) were derived as additional bands for each scene of the *LA* raster time series [127,128]. The *NDVI* describes a proxy for plant condition and scales between −1 and 1. Negative *NDVI* values correspond to water, low positive values to nearly non-vegetated areas, and increasing values to enhanced biomass and healthier vegetation.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (3)$$

The *NBR2* is based on the ratio of two shortwave-infrared bands (*SWIR1* and *SWIR2*) and serves as a proxy for litter deposit, soil moisture, and soil texture, among others [67]. The *NBR2* index also scales between −1 and 1, while positive values correspond to vegetation and negative values indicate bare surfaces [129].

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

The *S2* raster time series was compiled from all available scenes of the *S2* sensors for the period from 2017 to 2021. The *S2* sensors record a blue, green, red, and near-infrared band at a spatial resolution of 10 m × 10 m. *S2* scenes were corrected to surface reflectance using the *sen2cor* algorithm [130,131]. All pixels covered by clouds and cloud shadows were masked using the *S2cloudless* algorithm [132,133]. Moreover, we derived the *NDVI* as an additional band for each scene of the *S2* raster time series (Equation (3)).

Both the *LA* and *S2* raster time series were temporarily restricted to scenes acquired from the 1st of March to the 31st of October of each year. Also, the raster time series were spatially limited to areas of existing soil surfaces, and thus, sealed surfaces as well as areas covered by rock, snow, ice, and water were masked. The raster time series were obtained from the *Google Earth Engine Data Catalogue* and processed using the *Google Earth Engine API* [134,135].

2.6.2. Land Use Covariates

The land use covariates are based on multiple annual aggregations of the *LA* and *S2* *NDVI* raster time series using the median, maximum, minimum, and standard deviation (Table 3). For each aggregation, we computed multiple statistical trend functions retrospectively for each year of the time series and across multiple temporal scales. We defined the trend functions by the mean, standard deviation, and the *Theil-Sen* trend estimator [136,137]. The temporal scales were defined by five-year intervals for the *LA* and by annual intervals for the *S2* raster time series. Since the *LA* and *S2* raster time series span across 37 and 5 years, respectively, the number of temporal scales depends on the referred year of the time series. In total, we derived 1558 covariates related to land use (Figures S3 and S4).

Table 3. List of *NDVI* raster time series based on annually aggregated data used for the land use trend analysis (*NDVI*: Normalized Difference Vegetation Index; *S2*: Sentinel-2; *LA*: Landsat).

Data	Annual Aggregation	Timespan	Resolution
S2 LA	Median	2017–2022 1985–2022	10 m 30 m
S2 LA	Maximum	2017–2022 1985–2022	10 m 30 m
S2 LA	Minimum	2017–2022 1985–2022	10 m 30 m
S2 LA	Standard deviation	2017–2022 1985–2022	10 m 30 m

2.6.3. Spectral Bare Soil Covariates

The spectral bare soil covariates represent band-wise composites based on a median aggregation of the *LA* raster time series, using pixels that flag bare surfaces only (Table 4). Pixels showing *NDVI* values > 0.25 , which implies green vegetation and growing crops, were removed [40,73]. Pixels with *NBR2* values > 0.075 were also excluded since these pixels indicate straw, crop residues, or biased spectral signatures due to high soil moisture content [38,75]. Pixel stacks with a bare soil pixel frequency < 3 were rejected to prevent biased median aggregations due to the limited data population. Moreover, pixel stacks showing an elevation > 1200 m (a.s.l.) and a slope $> 25^\circ$ were excluded since topographic roughness combined with low solar angles causes biased spectral signatures due to shading effects [138,139].

Table 4. List of bare soil composites and their spectral range used to derive bare soil covariates.

Bare Soil Composite	Spectral Range [μm]
BS_BLUE	0.45–0.52
BS_GREEN	0.52–0.60
BS_RED	0.63–0.69
BS_NIR	0.76–0.90
BS_SWIR1	1.55–1.75
BS_SWIR2	2.08–2.35

The bare soil composites cover the study area only partially due to the existence of permanently vegetated areas (e.g., forests). These data gaps were filled for each bare soil composite using the spatial modeling approach of the aforementioned soil model design (cf. 2.2) with covariates related to terrain, climate, and land use (cf. 2.4–2.6). For this purpose, we extracted calibration sample sets from the bare soil composites, representing their geographical and spectral space evenly. Therefore, we derived spatial clusters using unsupervised *K-means* based on the geographic position and the spectral bare soil signature. Within these clusters, we randomly selected subsamples of size $n = 5000$,

which we combined to form the calibration sample set. The optimal number of clusters was determined using the *Silhouette Method* (Figure S5) [140].

3. Results

3.1. Soil Sample Data

The soil sample data were mainly acquired in the 1990s and generally cover the entire study area (Figure S6). The spatial sample density ($SSD, n_{xy} \cdot \text{km}^{-2}$) amounts to 0.34 for *Clay*, 0.37 for *SOC*, 0.56 for *pH*, and 0.06 for CEC_{pot} . All sample data sets show a generally decreasing SSD from 'Cropland' to 'Grassland' and 'Woodland', as well as from the 'Topsoil' to 'Subsoil' and 'Deep Subsoil'. The $SSDs$ of 'Cropland' are distinctively larger by approximately 80% compared to 'Grassland' and 'Woodland'. The declines of SSD from the upper to lower depth intervals are more continuous, ranging between 23% and 48% (Table 5).

Table 5. Spatial sample density SSD ($n_{xy} \cdot \text{km}^{-2}$; n_{xy} : Count of unique sample locations) for the soil sample data (*Clay*, *SOC*, *pH*, and CEC_{pot}). Provided values refer to the entire study area ('Total area') across all depth intervals, as well as to specific depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil') and landcover classes ('Cropland', 'Grassland', and 'Woodland').

Soil Property	Depth	Total Area	Cropland	Grassland	Woodland
Clay	0–120 cm	0.34	1.20	0.22	0.20
	0–30 cm	0.34	1.19	0.22	0.20
	30–60 cm	0.26	0.96	0.16	0.15
	60–120 cm	0.20	0.73	0.12	0.12
SOC	0–120 cm	0.37	1.25	0.23	0.25
	0–30 cm	0.37	1.24	0.22	0.24
	30–60 cm	0.20	0.62	0.11	0.17
	60–120 cm	0.12	0.33	0.06	0.12
pH	0–120 cm	0.56	1.84	0.36	0.37
	0–30 cm	0.56	1.83	0.35	0.37
	30–60 cm	0.29	0.97	0.17	0.19
	60–120 cm	0.21	0.72	0.12	0.15
CEC_{pot}	0–120 cm	0.06	0.21	0.04	0.04
	0–30 cm	0.06	0.20	0.04	0.04
	30–60 cm	0.04	0.15	0.02	0.02
	60–120 cm	0.03	0.13	0.02	0.01

The count of unique sample locations (n_{xy}) for the soil sample data amount to 10,467 for *Clay*, 11,329 for *SOC*, 17,052 for *pH*, and 1879 for CEC_{pot} . The descriptive statistics of the *Clay* (%) data show a marginally right-skewed distribution with a median of 22.04 and a standard deviation of 10.78. The *SOC* (%) data show an explicitly right-skewed distribution with a median and standard deviation of 1.75 and 1.76, respectively. The *pH* (CaCl_2) data set is left-skewed, with a median of 6.46 and a standard deviation of 1.14. The CEC_{pot} ($\text{mm}_c \cdot \text{kg}^{-1}$) data set is right-skewed with a median and standard deviation of 150.48 and 88.43, respectively (Table 6).

Table 6. Descriptive statistics of the soil sample data (*Clay*; *SOC*; *pH*; CEC_{pot} ; Min: Minimum; Max: Maximum; SD: Standard deviation; n_{xy} : Count of unique sample locations).

Soil Property	Min	Max	Mean	Median	SD	n_{xy}
Clay [%]	0.20	69.17	24.16	22.04	10.78	10,476
SOC [%]	0.01	10.19	2.14	1.75	1.76	11,329
pH [-]	2.50	9.00	6.24	6.46	1.14	17,052
CEC_{pot} [$\text{mm}_c \cdot \text{kg}^{-1}$]	1.04	554.35	170.46	150.48	88.43	1879

The frequency distributions of the soil sample data for *Clay* (%), *SOC* (%), and CEC_{pot} ($\text{mm}_c \cdot \text{kg}^{-1}$), stratified by landcover classes and depth intervals, are generally similar across

the strata. This similarity is confirmed for the pH ($CaCl_2$) sample data across depth intervals, as well as across ‘Cropland’ and ‘Grassland’, while showing a left-skewed distribution. However, the pH ($CaCl_2$) data set for ‘Woodland’ shows a bimodal distribution across all depth intervals. Apart from that, the distributions of the stratified data sets are comparable to those of the entire data population (Figure 2).

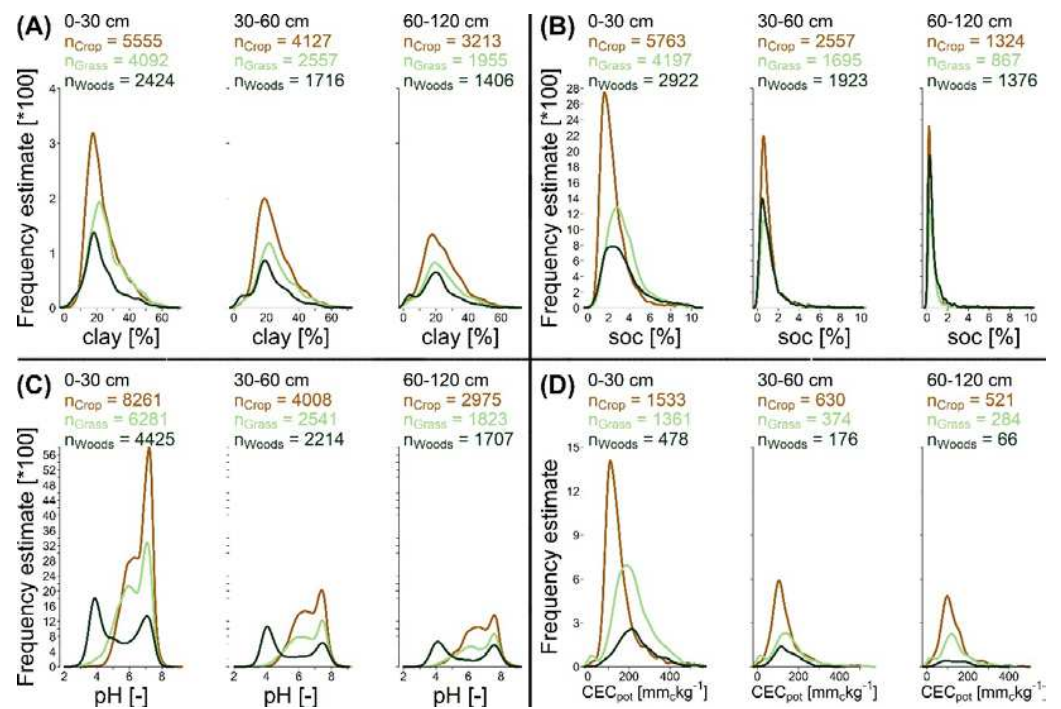


Figure 2. Frequency distributions of the soil sample data ((A) Clay in %; (B) SOC in %; (C) pH in $CaCl_2$; (D) CEC_{pot} in $mm_c.kg^{-1}$), stratified by depth intervals (‘Topsoil’, ‘Subsoil’, and ‘Deep Subsoil’) and landcover classes (brown: ‘Cropland’; light green: ‘Grassland’; dark green: ‘Woodland’; n: count of samples per stratum).

3.2. Clay Model

The Clay model shows an R^2 of 0.69, an $RMSE$ of 6.07, and an RPI of 0.89 as the mean across five independent submodels and based on the IV . With respect to depth intervals, the model achieved the highest accuracy and lowest uncertainty for ‘Subsoil’ ($R^2 = 0.76$; $RMSE = 5.39$; $RPI = 0.84$), followed by ‘Topsoil’ ($R^2 = 0.69$; $RMSE = 5.89$; $RPI = 0.91$) and ‘Deep Subsoil’ ($R^2 = 0.62$; $RMSE = 7.13$; $RPI = 0.92$). With respect to landcover classes and according to the R^2 , the best results were obtained for ‘Grassland’ ($R^2 = 0.76$), followed by a similar R^2 for ‘Cropland’ and ‘Woodland’ ($R^2 = 0.68$). According to the $RMSE$ and RPI , the results for ‘Cropland’ ($RMSE = 5.98$; $RPI = 0.77$) outperform ‘Grassland’ ($RMSE = 6.03$; $RPI = 0.86$), with a further performance decrease for ‘Woodland’ ($RMSE = 6.32$; $RPI = 1.03$; Table 7; Figures 3 and S7).

The Clay model’s performance is robust, showing low variability across the submodels with a SD_{R^2} of 0.01 and a SD_{RMSE} of 0.02. The accuracy indications based on the IV are confirmed, referring to the OOB_{error} ($R^2 = 0.69$; $SD_{R^2} = 0.01$; $RMSE = 5.99$; $SD_{RMSE} = 0.02$) and the $CV10$ ($R^2 = 0.68$; $SD_{R^2} = 0.01$; $RMSE = 6.13$; $SD_{RMSE} = 0.02$).

The covariate importance analysis revealed a high influence of covariates related to terrain and climate. The impact of spectral bare soil and land use covariates on the model performance was intermediate and low, respectively (Figure 3).

The Clay mapping shows consistent patterns across landcover classes and depth intervals. The mean Clay values range between 20% and 25% for all strata. However, referring to ‘Grassland’ and ‘Woodland’, Clay values are marginally larger for the ‘Subsoil’ (Figures 3, 4 and S8).

Table 7. Performance of the Clay model according to accuracy (R^2 and $RMSE$) and uncertainty (RPI) based on independent validation data. Provided values are averaged across five independent submodels and refer to the entire study area ('Total area') across all depth intervals, as well as to specific depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil') and landcover classes ('Cropland', 'Grassland', and 'Woodland').

Indicator	Depth	Total Area	Cropland	Grassland	Woodland
R^2	0–120 cm	0.69	0.68	0.76	0.68
	0–30 cm	0.69	0.68	0.73	0.62
	30–60 cm	0.76	0.75	0.77	0.77
	60–120 cm	0.62	0.61	0.63	0.64
$RMSE$	0–120 cm	6.07	5.98	6.03	6.32
	0–30 cm	5.89	5.58	5.91	6.51
	30–60 cm	5.39	5.35	5.40	5.44
	60–120 cm	7.13	7.27	6.97	6.95
RPI	0–120 cm	0.89	0.77	0.86	1.03
	0–30 cm	0.91	0.80	0.86	1.06
	30–60 cm	0.84	0.71	0.84	0.97
	60–120 cm	0.92	0.81	0.89	1.07

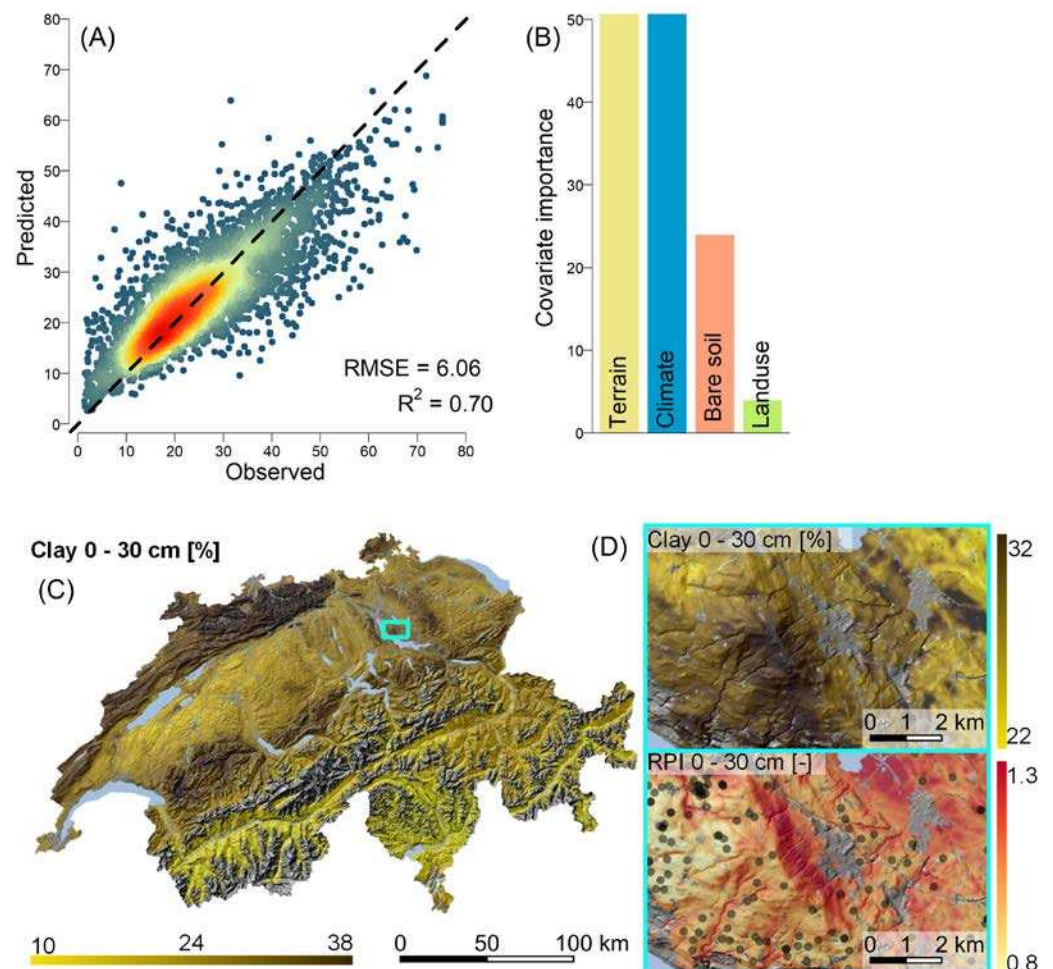


Figure 3. Clay model characteristics. (A) A density scatterplot of observed vs. predicted values based on five independent submodels and independent validation data (dashed line indicates 1:1 line; blue-yellow-red color scale indicates low to high density); (B) a barplot for the covariate importance of covariates related to terrain, climate, bare soil, and land use; (C) a topsoil prediction map at national scale; and (D) a zoomed-in map including RPI uncertainty (dots indicate calibration data).

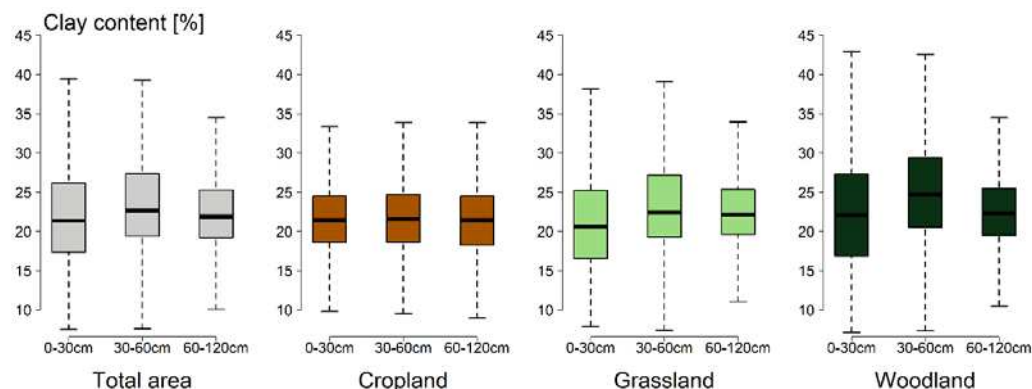


Figure 4. Boxplots of Clay predictions by landcover classes ('Total area', 'Cropland', 'Grassland', and 'Woodland') and depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil').

3.3. SOC Model

The SOC model shows an R^2 of 0.64, an $RMSE$ of 1.06, and an RPI of 3.04 as the mean across five independent submodels and based on the IV . With respect to depth intervals as well as according to R^2 and RPI , the model performs best for 'Topsoil' ($R^2 = 0.50$; $RPI = 1.57$), followed by 'Subsoil' ($R^2 = 0.45$; $RPI = 3.71$) and 'Deep Subsoil' ($R^2 = 0.37$; $RPI = 3.85$). According to $RMSE$, the best model performance is indicated for 'Subsoil' and 'Deep Subsoil' ($RMSE = 0.92$). With respect to landcover classes and according to R^2 , the model shows the highest accuracy for 'Grassland' ($R^2 = 0.66$), followed by 'Woodland' ($R^2 = 0.62$) and 'Cropland' ($R^2 = 0.61$). According to $RMSE$ and RPI , the model performs best on 'Cropland' ($RMSE = 0.94$; $RPI = 2.83$), followed by 'Grassland' ($RMSE = 1.09$; $RPI = 2.87$) and 'Woodland' ($RMSE = 1.21$; $RPI = 3.43$; Table 8; Figures 5 and S7).

Table 8. Performance of the SOC model according to accuracy (R^2 and $RMSE$) and uncertainty (RPI) based on independent validation data. Provided values are averaged across five independent submodels and refer to the entire study area ('Total area') across all depth intervals, as well as to specific depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil') and landcover classes ('Cropland', 'Grassland', and 'Woodland').

Indicator	Depth	Total Area	Cropland	Grassland	Woodland
R^2	0–120 cm	0.64	0.61	0.66	0.62
	0–30 cm	0.50	0.51	0.48	0.37
	30–60 cm	0.45	0.46	0.42	0.47
	60–120 cm	0.37	0.37	0.38	0.41
$RMSE$	0–120 cm	1.06	0.94	1.09	1.21
	0–30 cm	1.16	0.96	1.13	1.5
	30–60 cm	0.92	0.91	1.01	0.83
	60–120 cm	0.92	0.84	1.01	0.90
RPI	0–120 cm	3.04	2.83	2.87	3.43
	0–30 cm	1.57	1.49	1.47	1.75
	30–60 cm	3.71	3.33	3.67	4.12
	60–120 cm	3.85	3.67	3.47	4.42

The SOC model performance is robust, as indicated by the low variability across the submodels with a SD_{R^2} and a SD_{RMSE} of 0.01. The model accuracy indications based on the IV are comparable to those derived from OOB_{error} ($R^2 = 0.62$; $SD_{R^2} = 0.01$; $RMSE = 1.08$; $SD_{RMSE} = 0.01$) and $CV10$ ($R^2 = 0.63$; $SD_{R^2} = 0.01$; $RMSE = 1.10$; $SD_{RMSE} = 0.01$).

The covariate importance analysis indicated that covariates related to spectral bare soil and terrain are highly influential for the SOC model. Covariates related to climate and land use are likewise less important (Figure 5).

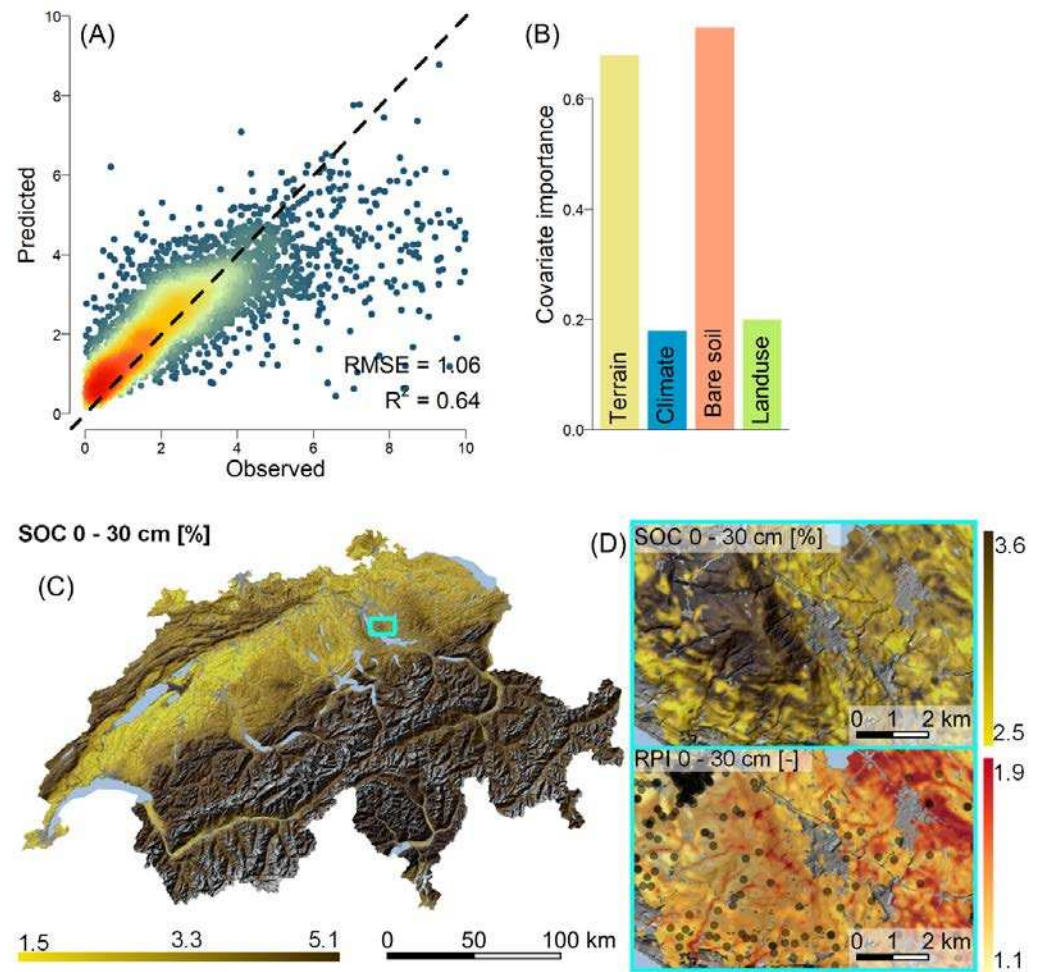


Figure 5. SOC model characteristics. (A) A density scatterplot of observed vs. predicted values based on five independent submodels and independent validation data (dashed line indicates 1:1 line; blue-yellow-red color scale indicates low to high density); (B) a barplot for the covariate importance of covariates related to terrain, climate, bare soil, and land use; (C) a topsoil prediction map at national scale; and (D) a zoomed-in map including RPI uncertainty (dots indicate calibration data).

With respect to predicted SOC values for landcover classes and depth intervals, the mean SOC decreases significantly from ‘Topsoil’ to ‘Subsoil’ by 89% and is less pronounced (59%) from ‘Subsoil’ to ‘Deep Subsoil’. Moreover, the mean SOC decreased for ‘Cropland’ (1.25%) compared to ‘Grassland’ (1.88%) and ‘Woodland’ (1.81%; Figures 5, 6 and S9).

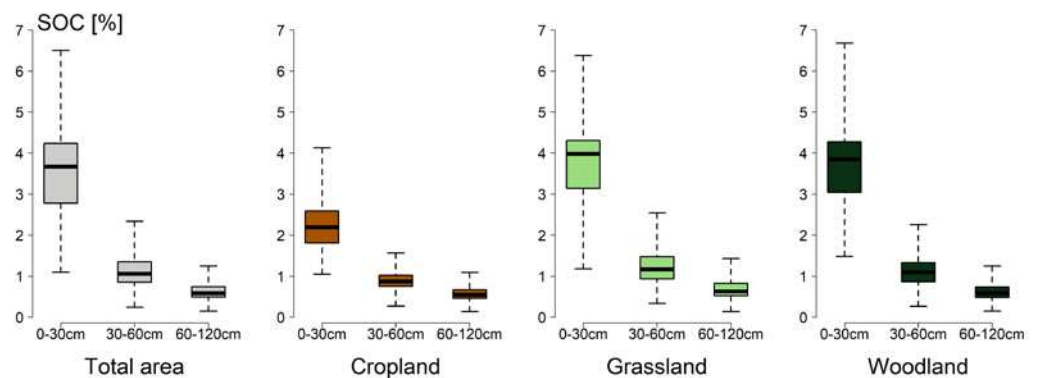


Figure 6. Boxplots of SOC predictions by landcover classes (‘Total area’, ‘Cropland’, ‘Grassland’, and ‘Woodland’) and depth intervals (‘Topsoil’, ‘Subsoil’, and ‘Deep Subsoil’).

3.4. pH Model

The pH model shows an R^2 of 0.76, an RMSE of 0.56, and an RPI of 0.39 as the mean across five independent submodels and based on the IV. With respect to depth intervals as well as according to R^2 and RMSE, the highest accuracy was found for 'Subsoil' ($R^2 = 0.86$; RMSE = 0.45), followed by 'Deep Subsoil' ($R^2 = 0.8$; RMSE = 0.57) and 'Topsoil' ($R^2 = 0.70$; RMSE = 0.60). The lowest uncertainty was indicated for 'Subsoil' (RPI = 0.36), while 'Topsoil' (RPI = 0.39) revealed a decreased uncertainty compared to 'Deep Subsoil' (RPI = 0.43). With respect to landcover classes, R^2 is comparable between 'Cropland' ($R^2 = 0.72$), 'Grassland' ($R^2 = 0.72$), and 'Woodland' ($R^2 = 0.71$). However, according to RMSE and RPI, the model performance is best for 'Cropland' (RMSE = 0.43; RPI = 0.25), followed by 'Grassland' (RMSE = 0.51; RPI = 0.39) and 'Woodland' (RMSE = 0.79; RPI = 0.53; Table 9; Figures 7 and S7).

Table 9. Performance of the pH model according to accuracy (R^2 and RMSE) and uncertainty (RPI) based on independent validation data. Provided values are averaged across five independent submodels and refer to the entire study area ('Total area') across all depth intervals, as well as to specific depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil') and landcover classes ('Cropland', 'Grassland', and 'Woodland').

Indicator	Depth	Total Area	Cropland	Grassland	Woodland
R^2	0–120 cm	0.76	0.72	0.72	0.71
	0–30 cm	0.70	0.61	0.66	0.63
	30–60 cm	0.86	0.85	0.83	0.84
	60–120 cm	0.80	0.79	0.75	0.81
RMSE	0–120 cm	0.56	0.43	0.51	0.79
	0–30 cm	0.60	0.47	0.53	0.88
	30–60 cm	0.45	0.34	0.43	0.61
	60–120 cm	0.57	0.44	0.56	0.77
RPI	0–120 cm	0.39	0.25	0.39	0.53
	0–30 cm	0.39	0.26	0.34	0.56
	30–60 cm	0.36	0.24	0.32	0.51
	60–120 cm	0.43	0.25	0.50	0.53

The pH model performance is robust with low variability across five independent submodels, as indicated by a SD_{R^2} and a SD_{RMSE} of 0.01. Moreover, accuracy levels based on the IV are similar to those derived from OOB_{error} ($R^2 = 0.75$; $SD_{R^2} = 0.01$; RMSE = 0.56; $SD_{RMSE} = 0.01$) and CV10 ($R^2 = 0.74$; $SD_{R^2} = 0.01$; RMSE = 0.58; $SD_{RMSE} = 0.01$).

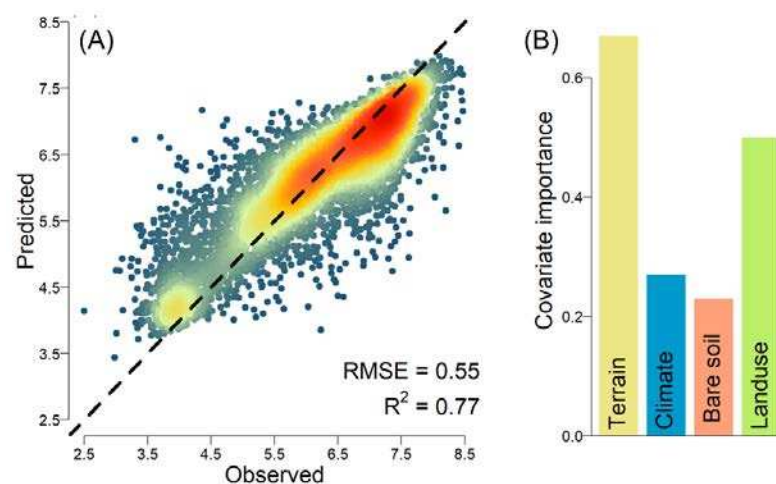


Figure 7. Cont.

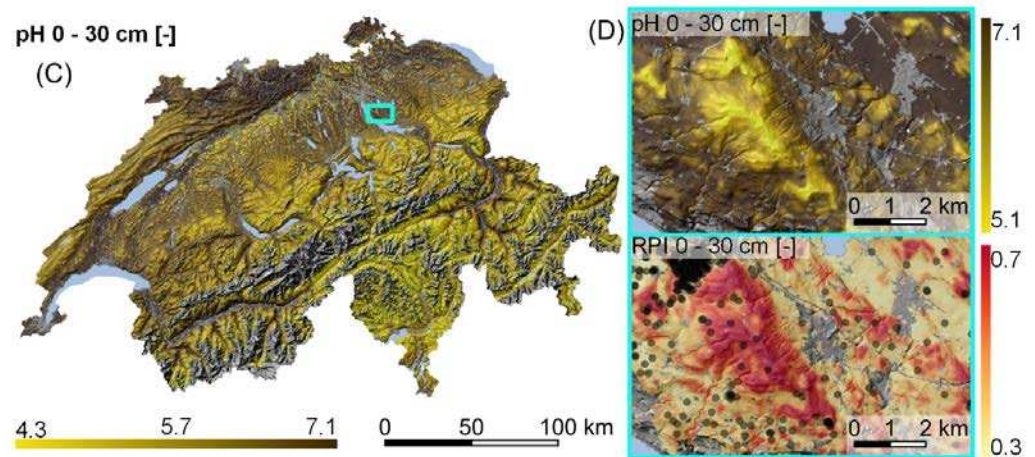


Figure 7. The pH model characteristics. (A) A density scatterplot of observed vs. predicted values based on five independent submodels and independent validation data (dashed line indicates 1:1; blue-yellow-red color scale indicates low to high density line); (B) a barplot for the covariate importance of covariates related to terrain, climate, bare soil, and land use; (C) a topsoil prediction map at national scale; and (D) a zoomed-in map including RPI uncertainty (dots indicate calibration data).

The covariates related to terrain and land use are indicated as highly influential for the performance of the pH model. The impact of covariates related to climate and spectral bare soil is less pronounced (Figure 7).

With respect to map values, the mean pH increases from 'Topsoil' to 'Subsoil' and 'Deep Subsoil' between 1% and 8% across landcover classes. Generally, the mean pH is larger for 'Cropland' (6.57) compared to 'Grassland' (5.87) and 'Woodland' (5.71; Figures 7, 8 and S10).

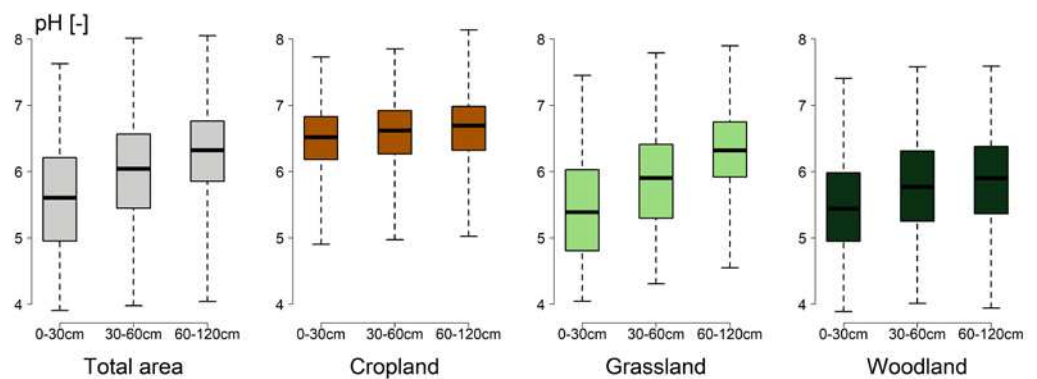


Figure 8. Boxplots of pH predictions by landcover classes ('Total area', 'Cropland', 'Grassland', and 'Woodland') and depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil').

3.5. CEC_{pot} Model

The CEC_{pot} model shows an R^2 of 0.72, an $RMSE$ of 48.61, and an RPI of 1.31 as the mean across five independent submodels and based on the IV . With respect to depth intervals and according to R^2 , the model performs best for 'Topsoil' ($R^2 = 0.71$), followed by 'Subsoil' ($R^2 = 0.67$) and 'Deep Subsoil' ($R^2 = 0.59$). According to $RMSE$, the highest accuracy was found for 'Subsoil' ($RMSE = 44.61$), followed by 'Deep Subsoil' ($RMSE = 45.08$) and 'Topsoil' ($RMSE = 50.65$). The RPI is lowest for 'Topsoil' ($RPI = 1.18$), followed by 'Deep Subsoil' ($RPI = 1.19$) and 'Subsoil' ($RPI = 1.55$). With respect to landcover classes, as well as according to R^2 and $RMSE$, the model performs best for 'Cropland' ($R^2 = 0.71$; $RMSE = 40.45$), followed by 'Grassland' ($R^2 = 0.69$; $RMSE = 54.17$) and 'Woodland' ($R^2 = 0.60$; $RMSE = 58.83$). The uncertainty increases from 'Cropland' ($RPI = 1.16$) to 'Woodland' ($RPI = 1.24$) and to 'Grassland' ($RPI = 1.52$; Table 10; Figures 9 and S7).

Table 10. Performance of the CEC_{pot} model according to accuracy (R^2 and $RMSE$) and uncertainty (RPI) based on independent validation data. Provided values are averaged across five independent submodels and refer to the entire study area ('Total area') across all depth intervals, as well as to specific depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil') and landcover classes ('Cropland', 'Grassland', and 'Woodland').

Indicator	Depth	Total Area	Cropland	Grassland	Woodland
R^2	0–120 cm	0.72	0.71	0.69	0.60
	0–30 cm	0.71	0.71	0.68	0.48
	30–60 cm	0.67	0.76	0.62	0.60
	60–120 cm	0.59	0.66	0.57	0.31
RMSE	0–120 cm	48.61	40.45	54.17	58.83
	0–30 cm	50.65	42.56	52.84	65.65
	30–60 cm	44.61	34.27	59.14	38.87
	60–120 cm	45.08	40.2	52.16	45.71
RPI	0–120 cm	1.31	1.16	1.52	1.24
	0–30 cm	1.18	1.04	1.33	1.17
	30–60 cm	1.55	1.38	1.94	1.32
	60–120 cm	1.19	1.06	1.29	1.22

The CEC_{pot} model performance is robust, referring to R^2 , showing low variability across independent submodels ($SD_{R^2} = 0.02$). However, the SD_{RMSE} is 1.96, which is relatively large. Accuracy indications based on the IV are comparable to those derived from OOB_{error} ($R^2 = 0.69$; $SD_{R^2} = 0.01$; $RMSE = 48.87$; $SD_{RMSE} = 0.55$) and $CV10$ ($R^2 = 0.69$; $SD_{R^2} = 0.01$; $RMSE = 49.71$; $SD_{RMSE} = 0.54$).

The covariate importance analysis revealed a high influence of covariates related to terrain and spectral bare soil. Covariates related to climate and land use were less important for the CEC_{pot} model (Figure 9).

With respect to map values, the mean CEC_{pot} decreases from 'Topsoil' to 'Subsoil' and 'Deep Subsoil' between 3% and 34%. The mean CEC_{pot} is lower for 'Cropland' ($137.76 \text{ mm}_c\text{kg}^{-1}$) compared to 'Grassland' ($179.21 \text{ mm}_c\text{kg}^{-1}$) and 'Woodland' ($184.79 \text{ mm}_c\text{kg}^{-1}$; Figures 9, 10 and S11).

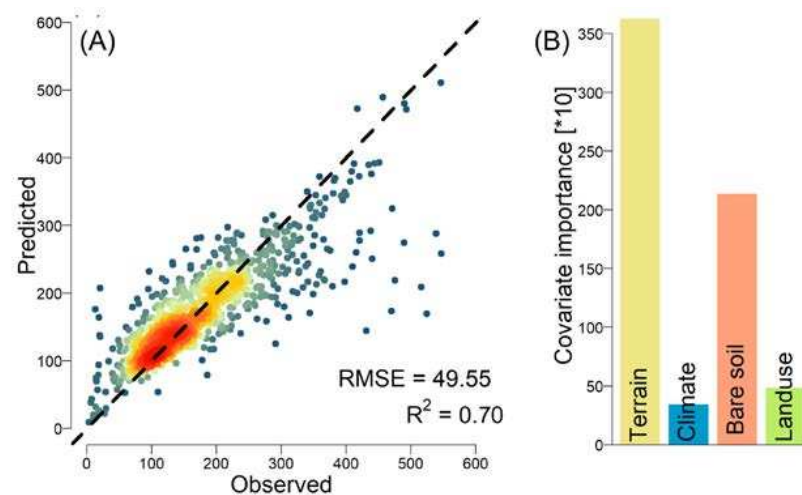


Figure 9. Cont.

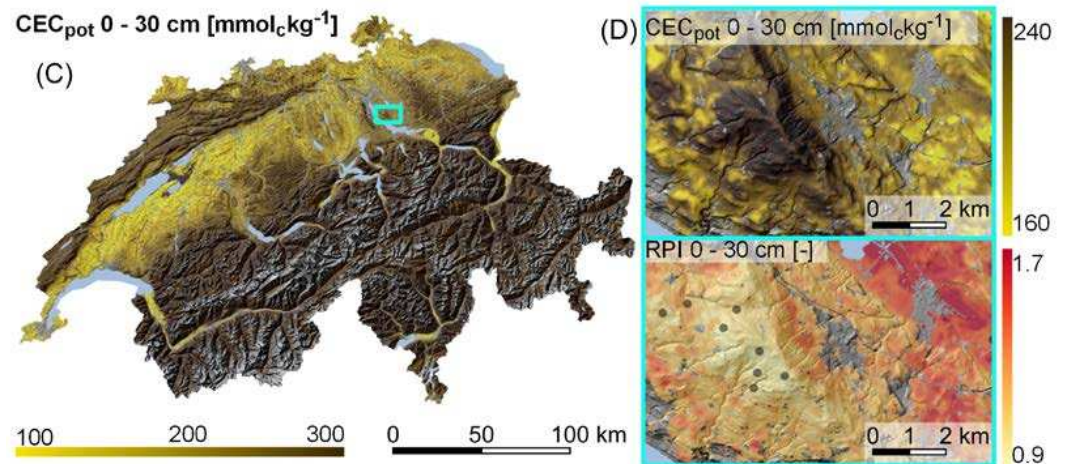


Figure 9. CEC_{pot} model characteristics. (A) A density scatterplot of observed vs. predicted values based on five independent submodels and independent validation data (dashed line indicates 1:1; blue-yellow-red color scale indicates low to high density line); (B) a barplot for the covariate importance of covariates related to terrain, climate, bare soil, and land use; (C) a topsoil prediction map at national scale; and (D) a zoomed-in map including RPI uncertainty (dots indicate calibration data).

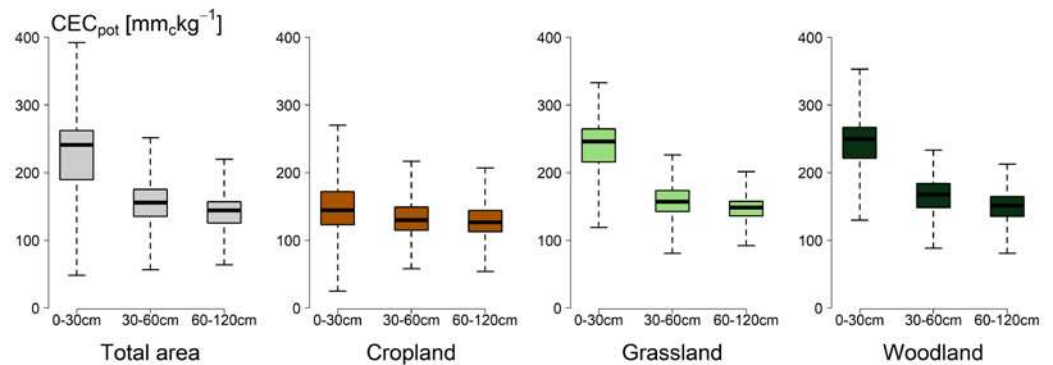


Figure 10. Boxplots of CEC_{pot} predictions by landcover classes ('Total area', 'Cropland', 'Grassland', and 'Woodland') and depth intervals ('Topsoil', 'Subsoil', and 'Deep Subsoil').

4. Discussion

4.1. Evaluation Strategy

We applied stratified random data splitting to determine the prediction accuracy, indicated by averaging the R^2 and $RMSE$ of multifold independent data splits and models, respectively. Data splitting, in combination with the aforementioned accuracy indicators, belongs to the prevalent accuracy estimation strategies for DSM [5,22]. The data splits were stratified, referring to depth intervals and landcover classes. This allows us to derive depth- and landcover-related accuracy estimations while using a single model calibrated across all strata. In comparison, recent DSM studies using a 3D model design as presented here disregarded stratification in the validation data and were limited to overall accuracy estimations across depth intervals or geographic regions [32–34].

The presented study provides a spatially explicit measure of the prediction uncertainty, framed by machine learning DSM for continuous soil property data. Generally, the uncertainty measure describes the limited ability of any DSM model to reproduce the soil–landscape relationship due to deficiencies in the soil sample and covariate data among others [15,141,142]. Uncertainty maps are particularly required for assessing broad-scale DSM products since the model is transferred to a variety of spatial domains, which may be inconsistently represented by the model [143–145]. Moreover, uncertainty maps, derived a priori based on archived soil data, can be leveraged as an information base for planning upcoming soil surveys. Accordingly, several studies have proposed sampling designs that

are optimized by targeting areas of high uncertainty for improved model performances combined with reduced field costs [13,15,146].

Since we applied *QRF*, *PIs* and thus our uncertainty measure (*RPI*) harness the built-in mechanism of *QRF* to predict conditional quantiles [141]. While the *QRF*-based uncertainty estimation has been generally approved [10,147], the *RPI* measure (*RPI*) is further approved by our results, showing corresponding trends between *RPI* and the accuracy indicators (R^2 , *RMSE*) across depth intervals and landcover classes.

Contrary to this, Kasraei et al. [25] proposed an uncertainty approach, applying a *QRF*-like quantile regression as a post-processing mechanism and thus being independent from the prediction model. Moreover, bootstrapping is another technique frequently applied in *DSM* that allows to derive uncertainty estimations independently from the type of prediction model. According to the bootstrapping method, multiple map realizations and thus *PIs* are derived based on randomized subsets of the training data [15,148]. Alternatively, Malone et al. [149] introduced an empirical method to derive spatially explicit *PIs* independently. The method clusters the covariate space based on a fuzzy K-means routine, then *PIs* are estimated based on the model errors within each cluster and finally regionalized according to the grade of membership to each cluster. Recently, Meyer and Pebesma [143] proposed a method to derive spatial uncertainties based on comparing the covariate spaces between a specific location and the calibration data rather than using *PIs*. This method acknowledges that prediction uncertainties increase when the covariate space of specific locations becomes increasingly dissimilar to the covariate space of the calibration data. The aforementioned approaches might be appealing since they are independent from the prediction model and applicable to a wide range of *DSM* designs. However, this also might include a more complex workflow, computational demand, and increased requirements related to data distributions and sample density, which are impractical in the context of broad-scale *DSM*. Instead, the use of *QRF* provides a straightforward workflow that integrates adequate model performance with reliable uncertainty estimations based on a single calibrated model [32].

4.2. Model Performance and Soil Sample Data

Numerous *DSM* studies identified the spatial sample density (*SSD*, $n_{xy} \cdot \text{km}^{-2}$) as a major driver for the model performance [150–152]. Accordingly, accuracy levels increase with an increasing *SSD* up to a certain threshold, which depends on the target soil property, model design, and environmental setting [64,153,154]. Chen et al. [50] reviewed 244 recent broad-scale *DSM* studies and reported an *SSD* variation between 0.0001 and 1, while only 5% of these studies outline an *SSD* larger than 0.1. Most broad-scale *DSM* studies in Europe showed an *SSD* between 0.01 and 0.1. A Czech *DSM* study by Žižala et al. [42] reported a relatively high *SSD* of 0.83 for *Clay* and 0.07 for *SOC*. A French *DSM* study by Mulder et al. [155] reported an *SSD* of 0.03. Zhou et al. [156] mapped *SOC* for Switzerland using a soil sample set with an *SSD* of 0.005 only. Generally, the *SSDs* of our study are considerably higher, ranging between 0.06 and 0.56 across the target soil properties.

Our results for CEC_{pot} revealed increasing *SSDs* combined with improved R^2 values from deeper to lower soil depths, as well as from 'Woodland' to 'Grassland' and to 'Cropland' (Tables 5 and 10). Since the *SSD* of CEC_{pot} is relatively low, the consistent trends indicate *SSD* as one driving factor for the *DSM* model's performance. However, referring to *Clay*, *SOC*, and *pH*, which show relatively high *SSDs*, the accuracy trends across depth intervals and landcover classes are ambiguous, most likely since *SSDs* are above a certain threshold to influence accuracy levels (Tables 5 and 7–9).

Besides a sufficiently high *SSD*, adequate *DSM* model performance depends on a uniform spatial distribution of the sample data to optimally represent the true soil variation within the study area [31,151]. However, similar to more than 50% of the broad-scale *DSM* studies, the present study is based on legacy soil data, which are not acquired purposefully for countrywide *DSM* [5]. Accordingly, our sample data show partly uneven frequency distributions and are spatially clustered towards 'Cropland' (Table 5; Figures 1, 2 and S6).

In consequence, similar to other broad-scale *DSM* studies, our model performances show spatial inconsistencies (Tables 7–9; Figures S8–S11). The deficiencies of our sample data, indicated by a depth-related decrease in *SSD* and spatial clustering, can be addressed by the purposeful and uncertainty-guided acquisition of additional soil samples [13,15,146].

4.3. Model Performance in Comparison

The review by Chen et al. [5] reported median R^2 values of broad-scale *DSM* studies, predominantly published between 2014 and 2021. In comparison, the R^2 values of the presented study are 14% to 33% higher across the target soil properties, referring to 3D *DSM* studies that disregard a depth-specific model evaluation (cf. 4.1). With respect to ‘Topsoil’ only, the R^2 values of the presented study are 2% to 95% higher (Table 11).

Table 11. Comparison of R^2 values between the presented study and the median of broad-scale *DSM* studies, reviewed by Chen et al. [5]. ‘Total’ refers to 3D *DSM* models that are not evaluated for specific depth intervals.

Soil Property	Depth	R^2 —Chen et al. [5]	R^2 —Presented Study	Change
Clay	0–30 cm	0.44	0.69	+57%
	Total	0.52	0.69	+33%
SOC	0–30 cm	0.49	0.50	+2%
	Total	0.56	0.64	+14%
pH	0–30 cm	0.60	0.70	+17%
	Total	0.60	0.76	+27%
CEC _{pot}	0–30 cm	0.37	0.71	+95%
	Total	0.59	0.72	+22%

The accuracy levels of the presented study are widely comparable with the most recent broad-scale *DSM* studies. Lu et al. [19] applied an extremely randomized *RF* model for mapping ‘Topsoil’ *pH* at the European scale and a spatial resolution of 250 m × 250 m, reporting an R^2 of 0.70. Žižala et al. [42] mapped *Clay*, *SOC*, and *pH* at the Czech national scale with a spatial resolution of 20 m × 20 m using a model design similar to the presented study. In comparison to Žižala et al. [42], our study shows higher R^2 values for modeling *Clay*, with +13% for ‘Topsoil’ and +33% for ‘Subsoil’. Referring to *SOC*, the R^2 values of our study are 25% higher for ‘Topsoil’ and 20% lower for ‘Subsoil’. The R^2 for our *pH* model is 59% higher. Furthermore, Liu et al. [157] mapped ‘Topsoil’ *pH*, *SOC*, and *CEC_{pot}* at a spatial resolution of 90 m × 90 m for the entire territory of China based on *QRF*. Referring to the *RMSE*, the accuracy levels of our study are higher, with +21% for *pH*, +13% for *SOC*, and +37% for *CEC_{pot}*.

In the Swiss context, the results of the presented study outperformed existing national-scale *DSM* studies. Baltensweiler et al. [35] mapped *pH*, *SOC*, and *Clay* for the Swiss ‘Woodland’ at a spatial resolution of 25 m × 25 m using *RF*. In comparison, our R^2 values for ‘Woodland’ *pH* predictions are higher, with +31% for ‘Topsoil’ and +68% for ‘Subsoil’. The *SOC* predictions of the presented study are higher, with +32% for ‘Topsoil’ and +104% for ‘Subsoil’. The R^2 values of our *Clay* predictions are higher, with +35% for ‘Topsoil’ and by 75% for ‘Subsoil’. Zhou et al. [156] applied *Boosted Regression Trees* to map *SOC* for the entire Swiss territory in a spatial resolution of 100 m × 100 m based on 150 topsoil samples only (0–20 cm). Our study shows a 6% higher R^2 value referring to ‘Topsoil’ (0–30 cm). Furthermore, Gupta et al. [64] mapped *SOC* and *Clay* for the entire territory of Switzerland at a spatial resolution of 30 m × 30 m using *QRF*. While their accuracy levels were not analyzed for specific depth intervals, our study shows a higher overall R^2 with +21% for the *SOC* and +73% for the *Clay* model.

4.4. Influential Covariates for the Soil Property Distributions

Due to the wide availability of DEMs, combined with the integrative character of the terrain to represent multiple *scorpan* factors, terrain covariates are widely used for DSM and often drive the model performance [5,60,158]. Numerous DSM studies found a dominant role for terrain covariates in predicting various soil properties in different environments [35,51,159]. Particularly, spatially multiscale terrain covariates are approved to be beneficial for DSM [42,60,61,160,161]. This is confirmed by the results of the presented study, which designate the group of multiscale terrain covariates as highly influential for the model performance across all target soil properties (Figures 3, 5, 7 and 9).

Climate influences soil formation due to interactions with hydrological cycles, whereby these interactions are related to changes in soil biology, soil physics, and soil hydrology [20,162,163]. Accordingly, the results of numerous DSM studies revealed a high influence of climate covariates in predicting *Clay*, *pH*, *SOC*, and *CEC_{pot}* [19,164–166]. In the presented study, climate covariates are a dominant factor for the *Clay* model but are medium-important for *pH* and less-important for predicting *SOC* and *CEC_{pot}*. However, since our model performances for *pH*, *SOC*, and *CEC_{pot}* are reasonable (Tables 8–10), the underlying soil-forming processes seem to be adequately described by the interactions of other covariates than those directly related to climate (Figures 3, 5, 7 and 9).

Land use and its change influence soil carbon and nutrient cycles, thereby changing soil biology and soil physics [77,167,168]. Several DSM studies reported the relatively high importance of land use covariates for predicting *Clay*, *pH*, *SOC*, and *CEC_{pot}* [19,80,81,169]. In the presented study, the covariate importance analysis indicates a lower influence of land use covariates for predicting *Clay*, *SOC*, and *CEC_{pot}*, while only the *pH* model is widely driven by covariates related to land use. However, land use patterns are directly influenced by specific climatic and terrain conditions, and thus, land use covariates are likely superimposed by those covariates (Figures 3, 5, 7 and 9) [20].

Covariates related to the spectral signature of bare soil correlate with soil properties such as *Clay*, *pH*, *SOC*, and *CEC_{pot}* in varying environments [38,40,56,73,170–172]. Thus, as suggested by Dematté et al. [67], spectral bare soil covariates are increasingly used as an integrative part of DSM covariate sets. For example, the Czech DSM study by Žížala et al. [42] reported a high dominance of spectral bare soil covariates for predicting *SOC* and a medium impact targeting *Clay* and *pH*. Also, a French DSM study by Urbina-Salazar et al. [75] identified covariates related to the spectral reflectance of bare soil as a dominant factor for mapping *SOC*. These findings are widely confirmed by the presented study, where spectral bare soil covariates are identified as a dominant factor for predicting *SOC* and *CEC_{pot}*, as well as medium influential for predicting *Clay* and *pH* (Figures 3, 5, 7 and 9). Since the use of spectral bare soil covariates for DSM emerged quite recently, the diversity of available techniques for filtering and compositing bare soil pixels is high [38,68,70,73,138,173]. In this context, Heiden et al. [174] emphasized that the selection of adequate spectral indices to filter bare soil pixels is of particular importance to challenge disturbing factors related to, e.g., varying soil moisture conditions and soil structures, as well as crop residue cover [67]. A wide range of spectral indices, including *NDVI*, *NBR2*, *Bare Soil Index (BSI)*, or adaptations thereof, have been applied [38,40,138,174,175]. In the present study, we used a combination of *NDVI* and *NBR2* as two of the most commonly used indices for bare soil detection with respect to *LA* data and various environments [38,40,69,70,176,177]. However, although our spectral bare soil covariates outlined a major importance for the soil model performances, a more detailed analysis of the spectral indices for bare soil detection is indicated with respect to future applications [174,175]. Furthermore, several studies reported promising results from leveraging hyperspectral imagery for bare soil detection and associated soil mapping [177–180]. Accordingly, the current advent of available hyperspectral data based on the satellite platforms *PRISMA* and *EnMAP* has high potential for future DSM applications at broad scales [177,181,182].

5. Conclusions

This study provides a *DSM* framework to map multiple soil properties in three dimensions for the entire territory of Switzerland based on machine learning as well as soil and remote sensing data archives. The *DSM* approach targets the soil properties of clay content, organic carbon concentrations, pH, and the potential cation exchange capacity for three depth intervals. A *Quantile Regression Forest* model was applied to link archived soil data with a comprehensive set of covariates derived from remote sensing data archives, a digital elevation model, and climate data, describing *scorpan* factors multiscale in space and time. Accordingly, the covariate set describes the spatially multiscale character of terrain, the spectral signature of bare soil, temporarily dynamic and multiscale patterns of land use, as well as climate and its associated temporal trends. The applied model design is particularly suitable for complex broad-scale *DSM*, allowing to efficiently derive predictions and uncertainties for multiple depth intervals based on a single calibrated model. Further, the model design allows us to evaluate the model's accuracy with respect to crop-, grass-, and woodland areas, as well as refer to specific depth intervals. The results show robust model performances, which are superior compared to previous *DSM* approaches at the national scale in Switzerland and in relation to the majority of recent broad-scale *DSM* approaches in other countries. Moreover, the results indicate that a spatially uniform and high soil sample density is a decisive factor for adequate *DSM* performance. The results suggest that in Switzerland, additional soil samples are particularly required for woodland and grassland areas, as well as for the subsoils. The multiscale remote sensing covariates were approved to be beneficial for *DSM*. Particularly, covariates related to the spectral signature of bare soil were found to be a major driving factor for *DSM* targeting organic carbon and cation exchange capacity. Generally, this study reveals the potential to update existing broad-scale *DSM* approaches by combining machine learning with a comprehensive exploitation of soil and remote sensing data archives.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16152712/s1>. Figure S1: Spatially multiscale terrain analysis using the example of flow accumulation and planar curvature at spatial scales of 8 m, 64 m, 272 m, and 712 m; Figure S2: Climate covariates as mean and mean interannual change across 1980 to 2021, referring to annual mean temperature, annual precipitation totals, and annual mean relative sunshine duration; Figure S3: NDVI land use covariates based on Sentinel-2 data, describing the mean of the annual median, standard deviation (SD), maximum (MAX), and minimum (MIN) across a five-year time scale (2017–2021; NDVI: Normalized Difference Vegetation Index); Figure S4: NDVI land use covariates based on Landsat data, describing the mean of the annual standard deviation (SD) across a 5- and 35-year time scale (NDVI: Normalized Difference Vegetation Index); Figure S5: Spectral bare soil covariates, describing the spectral signature of the blue (BLUE), green (GREEN), red (RED), near-infrared (NIR), and two shortwave infrared (SWIR1 and SWIR2) bands of the Landsat sensors; Figure S6: Spatial sample density and temporal distribution of soil data acquisition with respect to clay content (Clay; A), organic carbon content (SOC; B), pH (C), and potential cation exchange capacity (CECpot; D); nxy: Total number of unique sample locations); Figure S7: Model accuracies (R2: Coefficient of Determination; RMSE: Root Mean Square Error) referring to multiple soil properties, different depth intervals, and landcovers (large symbols refer to the average of the submodels that are indicated by small symbols); Figure S8: Spatial predictions and uncertainties (RPI) for the topsoil (0–30 cm) clay content (Clay). The ecoregions Jura, Central Plateau, Northern Alps, Central Alps, and Southern Alps are indicated by cyan line symbols and ordered from north to south; Figure S9: Spatial predictions and uncertainties (RPI) for the topsoil (0–30 cm) organic carbon concentrations (SOC). The ecoregions Jura, Central Plateau, Northern Alps, Central Alps, and Southern Alps are indicated by cyan line symbols and ordered from north to south; Figure S10: Spatial predictions and uncertainties (RPI) for the topsoil (0–30 cm) pH. The ecoregions Jura, Central Plateau, Northern Alps, Central Alps, and Southern Alps are indicated by cyan line symbols and ordered from north to south; Figure S11: Spatial predictions and uncertainties (RPI) for the topsoil (0–30 cm) potential cation exchange capacity (CECpot). The ecoregions Jura, Central Plateau, Northern Alps, Central Alps, and Southern Alps are indicated by cyan line symbols and ordered from north to south.

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

Funding: This research was funded by the Swiss Federal Office for the Environment (FOEN), the Swiss Federal Office for Agriculture (FOAG), and the Swiss Federal Office for Spatial Development (ARE).

Data Availability Statement: The soil sample data can be obtained from the *National Soil Information System NABODAT* (<https://www.nabodat.ch/index.php/de/>; accessed on 23 May 2024). The digital elevation model can be obtained from the *Swiss Federal Office of Topography, swisstopo* (<https://www.swisstopo.admin.ch/en>; accessed on 23 May 2024). Climate data are available at the Swiss Federal Office of Meteorology and Climatology, *meteoswiss* (<https://www.meteoswiss.admin.ch/#tab=forecast-map>; accessed on 23 May 2024). The spectral data from satellite imagery can be obtained via the Google Earth Engine cloud computing platform (<https://earthengine.google.com/>; accessed on 24 May 2024). The resulting soil property maps are available at the Swiss Competence Center for Soils (<https://ccsols.ch/de/home/>; accessed on 24 May 2024). For information about the code, please contact the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

- Adhikari, K.; Hartemink, A.E. Linking soils to ecosystem services—A global review. *Geoderma* **2016**, *262*, 101–111. [CrossRef]
- Lehmann, J.; Bossio, D.A.; Kögel-Knabner, I.; Rillig, M.C. The concept of future prospects of soil health. *Nat. Rev. Earth Environ.* **2020**, *1*, 544–553. [CrossRef] [PubMed]
- Pereira, P.; Bogunovic, I.; Muñoz-Rojas, M.; Brevik, E.C. Soil ecosystem services, sustainability, valuation and management. *Curr. Opin. Environ. Sci. Health* **2018**, *5*, 7–13. [CrossRef]
- Amelung, W.; Bossio, D.; Kögel-Knabner, I.; Lehmann, J.; Amundson, R.; Bol, R.; Collins, C.; Lal, R.; Leifeld, J.; Minasny, B.; et al. Towards a global-scale soil climate mitigation strategy. *Nat. Commun.* **2020**, *11*, 5427. [CrossRef]
- Chen, S.; Arrouays, D.; Mulder, V.L.; Poggio, L.; Minasny, B.; Roudier, P.; Libohova, Z.; Lagacherie, P.; Shi, Z.; Hannam, J.; et al. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma* **2022**, *409*, 115567. [CrossRef]
- Evans, D.L.; Janes-Bassett, V.; Borrelli, P.; Chenu, C.; Ferreira, C.S.S.; Griffiths, R.I.; Kalantari, Z.; Keestra, S.; Lala, R.; Panagos, P.; et al. Sustainable futures over the next decade are rooted in soil science. *Eur. J. Soil Sci.* **2021**, *73*, e13145. [CrossRef]
- Arrouays, D.; McBratney, A.; Bouma, J.; Libhova, Z.; Richer-de-Forges, A.C.; Morgan, C.L.S.; Roudier, P.; Poggio, L.; Mulder, V.L. Impression of digital soil maps: The good, the not so good, and the making them ever better. *Geoderma Reg.* **2020**, *20*, e00255. [CrossRef]
- Lagacherie, P.; McBratney, A.B. Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping. *Dev. Soil Sci.* **2006**, *31*, 3–22.
- Minasny, B.; McBratney, A.B. Digital soil mapping: A brief history and some lessons. *Geoderma* **2016**, *264*, 301–311. [CrossRef]
- Vaysse, K.; Lagacherie, P. Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma* **2017**, *291*, 55–64. [CrossRef]
- Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; McGraw Hill Book Company: New York, NY, USA, 1941; p. 281.
- McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]
- Blackford, C.; Heung, B.; Webster, K.L. Incorporating spatial uncertainty maps into soil sampling improves digital soil mapping classification accuracy in Ontario, Canada. *Geoderma Reg.* **2022**, *29*, e00495. [CrossRef]
- Stoorvogel, J.J.; Bakkenes, M.; Temme, A.J.A.M.; Batjes, N.H.; ten Brink, B.J.E. S-World: A global soil map for environmental modelling. *Land Degrad. Dev.* **2016**, *28*, 22–33. [CrossRef]
- Stumpf, F.; Schmidt, K.; Goebes, P.; Behrens, T.; Schönbrodt-Stitt, S.; Wadoux, A.; Xiang, W.; Scholten, T. Uncertainty-guided sampling to improve digital soil maps. *Catena* **2017**, *153*, 30–38. [CrossRef]
- Viscarra Rossel, R.A.; Lee, J.; Behrens, T.; Luo, Z.; Baldock, J.; Richards, A. Continental-scale soil carbon composition and vulnerability modulated by regional environmental controls. *Nat. Geosci.* **2019**, *12*, 547–552. [CrossRef]
- Vaysse, K.; Lagacherie, P. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Reg.* **2015**, *4*, 20–30. [CrossRef]
- Khaledian, Y.; Miller, B.A. Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model.* **2020**, *81*, 401–418. [CrossRef]

19. Lu, Q.; Tian, S.; Wei, L. Digital mapping of soil pH and carbonates at the European scale using environmental variables and machine learning. *Sci. Total Environ.* **2023**, *856*, 159171. [[CrossRef](#)]
20. Lamichhane, S.; Kumar, L.; Wilson, B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* **2019**, *352*, 395–413. [[CrossRef](#)]
21. Padarian, J.; Minasny, B.; McBratney, A.B. Machine learning and soil sciences: A review aided by machine learning tools. *Soil* **2020**, *6*, 35–52. [[CrossRef](#)]
22. Wadoux, A.M.J.-C.; Minasny, B.; McBratney, A.B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth Sci. Rev.* **2020**, *2010*, 103359. [[CrossRef](#)]
23. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
24. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.
25. Kasraei, B.; Heung, B.; Saurette, D.D.; Schmidt, M.G.; Bulmer, C.E.; Bethel, W. Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning. *Environ. Model. Softw.* **2021**, *144*, 105139. [[CrossRef](#)]
26. Chagas, C.S.; Carvalho Junior, W.; Bhering, S.B.; Filho, B.C. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena* **2016**, *139*, 232–240. [[CrossRef](#)]
27. Taghizadeh-Mehrjardi, R.; Mahdianpari, M.; Mohammadimanesh, F.; Behrens, T.; Toomanian, N.; Scholten, T.; Schmidt, K. Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. *Geoderma* **2020**, *376*, 114552. [[CrossRef](#)]
28. Zhang, Y.; Sui, B.; Shen, H.; Ouyang, L. Mapping stocks of soil total nitrogen using remote sensing data: A comparison of random forest models with different predictors. *Comput. Electron. Agric.* **2019**, *160*, 23–30. [[CrossRef](#)]
29. Belgiu, M.; Drăgut, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm.* **2016**, *114*, 24–31. [[CrossRef](#)]
30. Keskin, H.; Grunwald, S.; Harris, W.G. Digital mapping of soil carbon fractions with machine learning. *Geoderma* **2019**, *339*, 40–58. [[CrossRef](#)]
31. Wadoux, A.M.J.C.; Brus, D.J.; Heuvelink, G.B.M. Sampling design optimization for soil mapping with random forest. *Geoderma* **2019**, *355*, 113913.
32. Roudier, P.; Burge, O.R.; Richardson, S.J.; McCarthy, J.K.; Grealish, G.J.; Ausseli, A.G. National scale 3D mapping of soil pH using a data augmentation approach. *Remote Sens.* **2020**, *12*, 2872. [[CrossRef](#)]
33. Sanderman, J.; Hengl, T.; Fiske, G.; Solvik, K.; Adame, M.F.; Benson, L.; Bukoski, J.J.; Carnell, P.; Cifuentes-Jara, M.; Donato, D. A global map of mangrove forest soil carbon at 30 m spatial resolution. *Environ. Res. Lett.* **2018**, *13*, 055002. [[CrossRef](#)]
34. Sothe, C.; Gonsamo, A.; Arabian, J.; Snider, J. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. *Geoderma* **2022**, *405*, 115402. [[CrossRef](#)]
35. Baltensweiler, A.; Walthert, L.; Hanewinkel, M.; Zimmermann, S.; Nussbaum, M. Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. *Geoderma Reg.* **2021**, *27*, e00437. [[CrossRef](#)]
36. Heung, B.; Ho, H.C.; Zhang, J.; Knudby, A.; Bulmer, C.E.; Schmidt, M.G. An overview and comparison of machine learning techniques for classification purposes in digital soil mapping. *Geoderma* **2016**, *265*, 62–77. [[CrossRef](#)]
37. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* **2019**, *338*, 445–452. [[CrossRef](#)]
38. Castaldi, F.; Chabrilat, S.; Don, A.; van Wesemael, B. Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects. *Remote Sens.* **2019**, *11*, 2121. [[CrossRef](#)]
39. Liang, Z.; Chen, S.; Yang, Y.; Zhao, R.; Shi, Z.; Viscarra Rossel, R.A. National digital soil map of organic matter in topsoil and its associated uncertainty in 1980's China. *Geoderma* **2019**, *335*, 47–56. [[CrossRef](#)]
40. Safanelli, J.L.; Chabrilat, S.; Ben-Dor, E.; Demattê, J.A.M. Multispectral models from bare soil composites for mapping topsoil properties over Europe. *Remote Sens.* **2020**, *12*, 1369. [[CrossRef](#)]
41. Song, X.D.; Wu, H.Y.; Liu, F.; Yang, F.; Li, D.C.; Zhao, Y.G.; Yang, J.L.; Zhang, G.L. pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma* **2020**, *363*, 114145.
42. Žižala, D.; Minařík, R.; Skála, J.; Beitlerová, H.; Juřicová, J.; Reyes Rojas, J.; Penížek, V.; Zádorová, T. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. *Catena* **2022**, *212*, 106024.
43. Zhang, Y.; Saurette, D.D.; Easher, T.H.; Ji, W.; Adamchuk, V.I.; Biswas, A. Comparison of sampling designs for calibrating digital soil maps at multiple depths. *Pedosphere* **2022**, *32*, 588–601. [[CrossRef](#)]
44. Hengl, T.; de Jesus, J.M.; Heuvelink, G.B.M.; Gonzales, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, Q.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748.
45. Orton, T.G.; Pringle, M.J.; Bishop, T.F.A. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. *Geoderma* **2016**, *262*, 174–186. [[CrossRef](#)]
46. Ramcharan, A.; Hengl, T.; Nauman, T.; Brungard, C.; Waltman, S.; Wills, S.; Thompson, J. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. *Soil Sci. Soc. Am. J.* **2018**, *82*, 196–201.
47. Rentschler, T.; Gries, P.; Behrens, T.; Bruelheide, H.; Kühn, P.; Seitz, S.; Shi, X.; Trogisch, S.; Scholten, T.; Schmidt, K. Comparison of catchment scale 3D and 2.5D modelling of soil organic carbon stocks in Jiangxi province, PR China. *PLoS ONE* **2019**, *14*, e0220881.

48. Yamashita, N.; Ishizuka, S.; Hashimoto, S.; Ugawa, S.; Nanko, K.; Osone, Y.; Iwahashi, J.; Sakai, Y.; Inatomi, M.; Kawanishi, A.; et al. National-scale 3D mapping of soil organic carbon in Japanese forest considering microtopography and tephra deposition. *Geoderma* **2022**, *406*, 115534. [[CrossRef](#)]
49. Nauman, T.W.; Duniway, M.C. Relative prediction intervals reveal larger uncertainty in 3D approaches to predictive digital soil mapping of soil properties with legacy data. *Geoderma* **2019**, *347*, 170–184.
50. Ma, Y.; Minasny, B.; McBratney, A.; Poggio, L.; Fajardo, M. Predicting soil properties in 3D: Should depth be a covariate? *Geoderma* **2021**, *383*, 114794. [[CrossRef](#)]
51. Schillaci, C.; Acutis, M.; Lombardo, L.; Lipani, A.; Fantappiè, M.; Märker, M.; Saia, S. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. *Sci. Total Environ.* **2017**, *601–602*, 821–832. [[CrossRef](#)]
52. Tziolas, N.; Tsarkiridis, N.; Chabrilat, S.; Demattè, J.A.M.; Ben-Dor, E.; Gholizadeh, A.; Zalidis, G.; van Wesemael, B. Earth observation data-driven cropland soil monitoring: A review. *Remote Sens.* **2021**, *13*, 4439. [[CrossRef](#)]
53. Xiao, J.; Chevallier, F.; Gomez, C.; Guanter, L.; Hicke, J.A.; Heute, A.R.; Ichii, K.; Ni, W.; Pang, Y.; Rahman, A.F.; et al. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sens. Environ.* **2019**, *233*, 111383. [[CrossRef](#)]
54. Maynard, J.J.; Levi, M.R. Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability. *Geoderma* **2017**, *285*, 94–109. [[CrossRef](#)]
55. Rahmani, S.; Ackerson, J.P.; Schulze, D.; Adhikari, K.; Libohova, Z. Digital mapping of soil organic matter and cation exchange capacity in a low relief landscape using LiDAR data. *Agronomy* **2022**, *12*, 1338. [[CrossRef](#)]
56. Silvero, N.E.Q.; Demattè, J.A.M.; Amorim, M.T.A.; dos Santos, N.V.; Rizzo, R.; Safanelli, J.L.; Poppiel, R.R.; de Sousa Mendes, W.; Bonfatti, B.R. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. *Remote Sens. Environ.* **2021**, *252*, 112117. [[CrossRef](#)]
57. Biswas, A.; Cresswell, H.P.; Viscarra Rossel, R.A.; Si, B.C. Characterizing scale- and location-specific variation in non-linear soil systems using the wavelet transform. *Eur. J. Soil Sci.* **2013**, *64*, 706–715. [[CrossRef](#)]
58. Maynard, J.J.; Johnson, M.G. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. *Geoderma* **2014**, *230–231*, 29–40. [[CrossRef](#)]
59. Miller, B.A.; Koszinski, S.; Wehrhan, M.; Sommer, M. Impact of multi-scale predictor selection for modeling soil properties. *Geoderma* **2015**, *239–240*, 97–106. [[CrossRef](#)]
60. Behrens, T.; Schmidt, K.; MacMillan, R.A.; Viscarra Rossel, R.A. Multi-scale contextual spatial modelling with the Gaussian scale space. *Geoderma* **2018**, *310*, 128–137. [[CrossRef](#)]
61. Behrens, T.; Zhu, A.; Schmidt, K.; Scholten, T. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma* **2010**, *155*, 175–185. [[CrossRef](#)]
62. Behrens, T.; Schmidt, K.; Ramirez-Lopez, L.; Gallant, J.; Zhu, A.X.; Scholten, T. Hyper-scale digital soil mapping and soil formation analysis. *Geoderma* **2014**, *213*, 578–588. [[CrossRef](#)]
63. Sun, X.; Wang, H.; Zhao, Y.; Zhang, C.; Zhang, G. Digital soil mapping based on wavelet decomposed components of environmental covariates. *Geoderma* **2017**, *303*, 118–132. [[CrossRef](#)]
64. Gupta, S.; Hasler, J.K.; Alewell, C. Mining soil data of Switzerland: New maps for soil texture, soil organic carbon, nitrogen, and phosphorus. *Geoderma Reg.* **2024**, *36*, e00747. [[CrossRef](#)]
65. Kumar, A.; Moharana, P.C.; Jena, R.K.; Malyan, S.K.; Sharma, G.K.; Fagodiya, R.K.; Shabnam, A.A.; Jigyasu, D.K.; Kumari, K.M.V.; Doss, S.G. Digital mapping of soil organic carbon using machine learning algorithms in the upper Brahmaputra valley of northeastern India. *Land* **2023**, *12*, 1841. [[CrossRef](#)]
66. Zhou, T.; Geng, Y.; Chen, J.; Pan, J.; Haase, D.; Lausch, A. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning. *Sci. Total Environ.* **2020**, *729*, 138244. [[CrossRef](#)] [[PubMed](#)]
67. Demattè, J.A.M.; Fongaro, C.T.; Rizzo, R.; Silvero, N.E.Q.; Safanelli, J.L. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. *Remote Sens. Environ.* **2018**, *212*, 161–175. [[CrossRef](#)]
68. Rogge, D.; Bauer, A.; Zeidler, J.; Mueller, A.; Esch, T.; Heiden, U. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sens. Environ.* **2018**, *205*, 1–17. [[CrossRef](#)]
69. Safanelli, J.L.; Demattè, J.A.M.; Chabrilat, S.; Poppiel, R.R.; Rizzo, R.; Dotto, A.C.; Silvero, N.E.Q.; Mendes, W.S.; Bonfatti, B.R.; Ruiz, L.F.C.; et al. Leveraging the application of earth observation data for mapping cropland soils in Brazil. *Geoderma* **2021**, *396*, 115042. [[CrossRef](#)]
70. Demattè, J.A.M.; Safanelli, J.L.; Poppiel, R.R.; Rizzo, R.; Silvero, N.E.Q.; de Sousa Mendes, W.; Bonfatti, B.R.; Dotto, A.C.; Salazar, D.F.U.; Mello, F.A.D.O.; et al. Bare earth's surface spectra as a proxy for soil resource monitoring. *Sci. Rep.* **2020**, *10*, 4461. [[CrossRef](#)]
71. Bonfatti, B.R.; Demattè, J.A.M.; Marques, K.P.P.; Poppiel, R.R.; Rizzo, R.; Mendes, W.S.; Silvero, N.E.Q.; Safanelli, J.L. Digital mapping of soil parent material in a heterogeneous tropical area. *Geomorphology* **2020**, *367*, 107305. [[CrossRef](#)]

72. Viscarra Rossel, R.A.; Behrens, T.; Ben-Dor, E.; Brown, D.J.; Demattê, J.A.M.; Shepherd, K.D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; et al. A global spectral library to characterize the world's soil. *Earth Sci. Rev.* **2016**, *155*, 198–230. [CrossRef]
73. Sorenson, P.T.; Shirliff, S.J.; Bedard-Haughn, A.K. Predictive soil mapping using historic bare soil composite imagery and legacy soil survey data. *Geoderma* **2021**, *401*, 115316. [CrossRef]
74. Roberts, D.; Wilford, J.; Ghattas, O. Exposed soil and mineral map of the Australian continent revealing the land at its barest. *Nat. Commun.* **2019**, *10*, 5297. [CrossRef] [PubMed]
75. Urbina-Salazar, D.; Vaudour, E.; Riche-de-Forges, A.C.; Chen, S.; Martelet, G.; Baghdadi, N.; Arrouays, D. Sentinel-2 and Sentinel-1 bare soil temporal mosaics of 6-year periods for soil organic carbon content mapping in Central France. *Remote Sens.* **2023**, *15*, 2410. [CrossRef]
76. Delogu, E.; Olioso, A.; Alliès, A.; Demarty, J.; Boulet, G. Evaluation of multiple methods for the production of continuous evapotranspiration estimates from TIR remote sensing. *Remote Sens.* **2021**, *13*, 1086. [CrossRef]
77. Guo, L.; Fu, P.; Shi, T.; Chen, Y.; Zeng, C.; Zhang, H.; Wang, S. Exploring influence factors in mapping soil organic carbon on low-relief agricultural lands using time series of remote sensing data. *Soil Till. Res.* **2021**, *210*, 104982. [CrossRef]
78. Li, X.; Wang, X.; Wu, J.; Luo, W.; Tian, L.; Wang, Y.; Liu, Y.; Zhang, L.; Zhao, C.; Zhang, W. Soil moisture monitoring and evaluation in agricultural fields based on NDVI long time series and CEEMDAN. *Remote Sens.* **2023**, *15*, 5008. [CrossRef]
79. Stumpf, F.; Keller, A.; Schmidt, K.; Mayr, A.; Gubler, A.; Schaepman, M. Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. *Agric. Ecosyst. Environ.* **2018**, *258*, 129–142. [CrossRef]
80. Hounkpatin, K.O.L.; Bossa, A.Y.; Yira, Y.; Igue, M.A.; Sinsin, B.A. Assessment of the soil fertility status in Benin (West Africa)—Digital soil mapping using machine learning. *Geoderma Reg.* **2022**, *28*, e00444. [CrossRef]
81. Keshavarzi, A.; Sánchez del Árbol, M.Á.; Kaya, F.; Gyasi-Agyei, Y.; Rodrigo-Comino, J. Digital mapping of soil texture classes for efficient land management in the Piedmont plain of Iran. *Soil Use Manag.* **2022**, *38*, 1705–1735. [CrossRef]
82. Heuvelink, G.B.M.; Angelini, M.E.; Poggio, L.; Bai, Z.; Batjes, N.H.; Van den Bosch, R.; Bossio, D.; Estella, S.; Lehmann, J.; Olmedo, G.F.; et al. Machine learning in space and time for modelling soil organic carbon change. *Eur. J. Soil Sci.* **2020**, *72*, 1607–1623. [CrossRef]
83. Venter, Z.S.; Hawkins, H.J.; Cramer, M.D.; Mills, A.J. Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. *Sci. Total Environ.* **2021**, *771*, 145384. [CrossRef] [PubMed]
84. Wang, J.; Xiao, X.; Qin, Y.; Doughty, R.B.; Dong, J.; Zhou, Z. Characterizing the encroachment of juniper forests into sub-humid and semi-arid prairies from 1984 to 2010 using PALSAR and Landsat data. *Remote Sens. Environ.* **2018**, *205*, 166–179. [CrossRef]
85. Yang, H.; Zhang, X.; Xu, M.; Shao, S.; Wang, X.; Liu, W.; Wu, D.; Ma, Y.; Bao, Y.; Zhang, X.; et al. Hyper-temporal remote sensing data in bare soil period and terrain attributes for digital soil mapping in the Black soil regions of China. *Catena* **2020**, *184*, 104259. [CrossRef]
86. Federal Statistical Office, Switzerland—Land Use, Land Cover and Suitability. Available online: <https://www.bfs.admin.ch/bfs/en/home.html> (accessed on 2 April 2023).
87. Federal Office for Meteorology and Climatology, Switzerland—Spatial Climate Analysis. Available online: <https://www.meteoswiss.admin.ch/> (accessed on 10 April 2023).
88. Federal Office of Topography, Switzerland—swissALTI3D. Available online: <https://www.swisstopo.admin.ch/en/home.html> (accessed on 5 April 2023).
89. Bolliger, J.; Hagedorn, F.; Leifeld, J.; Böhl, J.; Zimmermann, S.; Soliva, R.; Kienast, F. Effects of land use change on carbon stocks in Switzerland. *Ecosystems* **2008**, *11*, 895–907. [CrossRef]
90. Yang, L.; Shami, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
91. Bishop, T.F.A.; McBratney, A.B.; Laslett, G.M. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **1999**, *91*, 27–45. [CrossRef]
92. Malone, B.P.; McBratney, A.B.; Minasny, B.; Laslett, G.M. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **2009**, *154*, 138–152. [CrossRef]
93. Lai, Y.R.; Orton, T.G.; Pringle, M.J.; Menzies, N.W.; Dang, Y.P. Increment-averaged kriging: A comparison with depth-harmonized mapping of soil exchangeable sodium percentage in a cropping region of eastern Australia. *Geoderma* **2020**, *363*, 114151. [CrossRef]
94. Gregorutti, B.; Michel, B.; Saint-Pierre, P. Grouped variable importance with random forests and application to multiple functional data analysis. *Comput. Stat. Data Anal.* **2015**, *90*, 15–35. [CrossRef]
95. Swiss Soil Information System: Soil Database for Switzerland—Version 6. Available online: <https://nabodat.ch/index.php/de/> (accessed on 10 April 2023).
96. Biodiversity Monitoring Switzerland. Available online: <https://www.biodiversitymonitoring.ch/index.php/en/> (accessed on 10 April 2023).
97. Meuli, R.G.; Wächter, D.; Schwab, P.; Kohli, L.; Zimmermann, R. Connecting biodiversity monitoring with soil inventory data—A Swiss case study. *Bull. BGS* **2017**, *38*, 65–69.
98. Descombes, P.; Wakthert, L.; Baltensweiler, A.; Meuli, R.G.; Karger, D.N.; Ginzler, C.; Zurell, D.; Zimmermann, N.E. Spatial modelling of ecological indicator values improves predictions of plant distributions in complex landscapes. *Ecography* **2020**, *43*, 1448–1463. [CrossRef]

99. Gee, G.W.; Bauder, J.; Klute, A. *Methods of Soil Analysis, Part 1, Physical and Mineralogical Methods*; Soil Science Society of America Book Series; American Society of Agronomy, Inc.: Madison, WI, USA, 1986; pp. 404–410.
100. Walkley, A.; Black, L.A. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Sci.* **1934**, *37*, 29–38. [[CrossRef](#)]
101. Weston, N.B.; Vile, M.A.; Neubauer, S.C.; Velinsky, D.J. Accelerated microbial organic matter mineralization following salt-water intrusion into tidal freshwater marsh soils. *Biogeochemistry* **2011**, *102*, 135–151. [[CrossRef](#)]
102. Gubler, A.; Wächter, D.; Schwab, P. Homogenization of series of soil organic carbon: Harmonizing results by wet oxidation (Swiss Standard Method) and dry combustion. *Agroscopie Sci.* **2018**, *62*, 1–9.
103. Heikkinen, J.; Ketoja, E.; Nuutinen, V.; Regina, K. Declining trend of carbon in Finnish cropland soils in 1974–2009. *Glob. Chang. Biol.* **2013**, *19*, 1456–1469. [[CrossRef](#)] [[PubMed](#)]
104. Agroscope, Swiss Lab Reference Methods. Available online: <https://www.agroscope.admin.ch/agroscope/en/home/topics/environment-resources/monitoring-analytics/referenzmethoden.html> (accessed on 10 April 2024).
105. Solly, E.F.; Weber, V.; Zimmermann, S.; Walthert, L.; Hagedorn, F.; Schmidt, M.W.I. A Critical Evaluation of the Relationship between the Effective Cation Exchange Capacity and Soil Organic Carbon Content in Swiss Forest Soils. *Front. For. Glob. Chang.* **2020**, *3*, 98. [[CrossRef](#)]
106. Walthert, L.; Graf Pannatier, E.; Meier, E.S. Shortage of nutrients and excess of toxic elements in soils limit the distribution of soil-sensitive tree species in temperate forests. *For. Ecol. Manag.* **2013**, *297*, 94–107. [[CrossRef](#)]
107. Walthert, L.; Meier, E.S. Tree species distribution in temperate forests is more influenced by soil than by climate. *Ecol. Evol.* **2017**, *7*, 9473–9484. [[CrossRef](#)]
108. Behrens, T.; Schmidt, K.; MacMillan, R.A.; Viscarra Rossel, R.A. Multi-scale digital soil mapping with deep learning. *Sci. Rep.* **2018**, *8*, 15244. [[CrossRef](#)]
109. Behrens, T.; Viscarra Rossel, R.A. On the interpretability of predictors in spatial data science: The information horizon. *Sci. Rep.* **2020**, *10*, 16737. [[CrossRef](#)] [[PubMed](#)]
110. Farr, T.G.; Rosen, P.G.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; et al. The Shuttle Radar Topography Mission. *Rev. Geophys.* **2007**, *45*, 2. [[CrossRef](#)]
111. Amirian-Chakan, A.; Minasny, B.; Taghizadeh-Mehrjardi, R.; Akbarifazli, R.; Darvishpasand, Z.; Khordehbin, S. Some practical aspects of predicting texture data in digital soil mapping. *Soil Till. Res.* **2019**, *194*, 104289. [[CrossRef](#)]
112. Zevenbergen, L.W.; Thorne, C.R. Quantitative analysis of land surface topography. *Earth Surf. Process. Landf.* **1987**, *12*, 47–56. [[CrossRef](#)]
113. Riley, S.J.; DeGloria, S.D.; Elliot, R. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermt. J. Sci.* **1999**, *5*, 23–27.
114. Wilson, M.F.J.; O’Connell, B.; Brown, C.; Guinan, J.C.; Grehan, A.J. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Mar. Geod.* **2007**, *30*, 3–35. [[CrossRef](#)]
115. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM versus CCA spatial modelling of plant species distribution. *Plant Ecol.* **1999**, *143*, 107–122. [[CrossRef](#)]
116. Quinn, P.F.; Beven, K.J.; Chevallier, P.; Planchon, O. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol. Process.* **1991**, *5*, 59–79. [[CrossRef](#)]
117. Kiss, R. Determination of drainage network in digital elevation model, utilities and limitations. *J. Hung. Geomath.* **2004**, *2*, 16–29.
118. Frei, C.; Schöll, R.; Fokutome, S.; Schmidli, J.; Vidale, P.L. Future change of precipitation extremes in Europe: An intercomparison of scenarios from regional climate models. *J. Geophys. Res.* **2006**, *111*, D06105. [[CrossRef](#)]
119. Frei, C. Interpolation of temperature in a mountainous region using non-linear profiles and non-Euclidean distances. *Int. J. Climatol.* **2014**, *34*, 1585–1605. [[CrossRef](#)]
120. Frei, C.; Willi, M.; Stöckli, R.; Dürri, B. Spatial analysis of sunshine duration in complex terrain by non-contemporaneous combination of station and satellite data. *Int. J. Climatol.* **2015**, *35*, 4471–4790. [[CrossRef](#)]
121. European Space Agency (ESA). Available online: <https://sentinels.copernicus.eu> (accessed on 8 April 2023).
122. U.S. Geological Survey. Available online: <https://www.usgs.gov/landsat-missions> (accessed on 8 April 2023).
123. Schmidt, G.; Jenkerson, C.; Masek, J.; Vermote, E.; Gao, F. *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description*; U.S. Geological Survey: Reston, VA, USA, 2013. [[CrossRef](#)]
124. Vermote, E.; Roger, J.C.; Franch, B.; Shakun, S. LaSRC (Land Surface Reflectance Code): Overview, application and validation using MODIS, VIIRS, LANDSAT and Sentinel 2 data’s. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018. [[CrossRef](#)]
125. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Hughes, M.J.; Laue, B. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [[CrossRef](#)]
126. Roy, D.P.; Kovalskyy, V.; Zhang, H.K.; Vermote, E.F.; Yan, L.; Kumar, S.S.; Egorov, A. Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sens. Environ.* **2016**, *185*, 57–70. [[CrossRef](#)]
127. García, M.J.L.; Caselles, V. Mapping burns and natural reforestation using thematic Mapper data. *Geocarto Int.* **1991**, *6*, 31–37. [[CrossRef](#)]

128. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS-Symp.* **1974**, *351*, 309.
129. Escuin, S.; Navarro, R.; Fernandez, P. Fire severity assessment by using NBR (normalized burn ratio) and NDVI (normalized difference vegetation index) derived from LANDSAT TM/ETM images. *Int. J. Remote Sens.* **2008**, *29*, 1053–1073. [[CrossRef](#)]
130. Louis, J.; Pflug, B.; Debaecker, V.; Mueller-Wilm, U.; Iannone, Q.; Boccia, V.; Gascon, F. Evolutions of Sentinel-2 Level-2A Cloud Masking Algorithm Sen2Cor Prototype First Results. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021. [[CrossRef](#)]
131. Sanchez, A.H.; Picoli, M.C.A.; Camara, G.; Andrade, P.R.; Chaves, M.E.D.; Lechler, S.; Soares, A.R.; Marajo, R.F.B.; Simões, R.E.O.; Ferreira, K.R.; et al. Comparison of cloud cover detection algorithms on Sentinel-2 images of the Amazon tropical forest. *Remote Sens.* **2020**, *12*, 1284. [[CrossRef](#)]
132. Nagy, D.; Warshavsky, Z.; Hughes, L.H. Improved Image Aggregation for Large-Scale Cloud-Free Image Creation. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021. [[CrossRef](#)]
133. Shakrun, S.; Wevers, J.; Brockmann, C.; Doxani, G.; Aleksandrov, M.; Batič, M.; Frantz, D.; Gascon, F.; Gómez-Chova, L.; Hagolle, O.; et al. Cloud mask intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. *Remote Sens. Environ.* **2022**, *274*, 112990.
134. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
135. Stumpf, F.; Schneider, M.K.; Keller, A.; Mayr, A.; Rentschler, T.; Meuli, R.G.; Schaepman, M.; Liebisch, F. Spatial monitoring of grassland management using multi-temporal satellite imagery. *Ecol. Indic.* **2020**, *113*, 106201. [[CrossRef](#)]
136. Sen, P.K. Estimates of the regression coefficient based on Kendall's Tau. *J. Am. Stat. Assoc.* **1968**, *63*, 1379–1389. [[CrossRef](#)]
137. Wu, D.; Wu, H.; Zhao, X.; Zhou, T.; Tang, B.; Zhao, W.; Jia, K. Evaluation of spatiotemporal variations of global fractional vegetation cover based on GIMMS NDVI data from 1982 to 2011. *Remote Sens.* **2014**, *6*, 4217–4239. [[CrossRef](#)]
138. Diek, S.; Fornallaz, F.; Schaepman, M.E.; De Jong, R. Barest pixel composite for agricultural areas using Landsat time series. *Remote Sens.* **2017**, *9*, 1245. [[CrossRef](#)]
139. Vázquez-Jiménez, R.; Romero-Calcerrada, R.; Ramos-Bernal, R.N.; Arrogante-Funes, P.; Novillo, C.J. Topographic correction to Landsat imagery through slope classification by applying the SCS+C method in mountainous forest areas. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 287. [[CrossRef](#)]
140. Batool, F.; Hennig, C. Clustering with average silhouette width. *Comput. Stat. Data Anal.* **2021**, *158*, 107190. [[CrossRef](#)]
141. Schmidinger, J.; Heuvelink, G.B.M. Validation of uncertainty predictions in digital soil mapping. *Geoderma* **2023**, *437*, 116585. [[CrossRef](#)]
142. Szatmári, G.; Pásztor, L. Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms. *Geoderma* **2019**, *337*, 1329–1340. [[CrossRef](#)]
143. Meyer, H.; Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods. Ecol. Evol.* **2021**, *12*, 1620–1633. [[CrossRef](#)]
144. Broeg, T.; Blaschek, M.; Seitz, S.; Taghizadeh-Mehrjardi, R.; Zepp, S.; Scholten, T. Transferability of covariates to predict soil organic carbon in cropland soils. *Remote Sens.* **2023**, *15*, 876. [[CrossRef](#)]
145. Rau, K.; Eggenberger, K.; Schneider, F.; Hennig, P.; Scholten, T. How can we quantify, explain, and apply the uncertainty of complex soil maps predicted with neural networks. *Sci. Total Environ.* **2024**, *944*, 173720. [[CrossRef](#)] [[PubMed](#)]
146. Huang, J.; McBratney, A.B.; Minasny, B.; Malone, B. Evaluating an adaptive sampling algorithm to assist soil survey in New South Wales, Australia. *Geoderma Reg.* **2020**, *21*, e00284. [[CrossRef](#)]
147. Fouedjio, F.; Klump, J. Exploring prediction uncertainty of spatial data in geostatistical and machine learning approaches. *Environ. Earth Sci.* **2019**, *78*, 38. [[CrossRef](#)]
148. Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning for digital soil mapping. *Soil* **2019**, *5*, 79–89. [[CrossRef](#)]
149. Malone, B.P.; McBratney, A.B.; Minasny, B. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma* **2011**, *160*, 614–626. [[CrossRef](#)]
150. Somarathna, P.D.S.N.; Minasny, B.; Malone, B.P. More data or better model? Figuring out what matters most for the spatial prediction of soil carbon. *Soil Sci. Soc. Am. J.* **2017**, *81*, 1413–1426. [[CrossRef](#)]
151. Lagacherie, P.; Arrouays, D.; Bourennane, H.; Gomez, C.; Nkuba-Kasanda, L. Analyzing the impact of soil spatial sampling on the performances of Digital Soil Mapping models and their evaluation: A numerical experiment on Quantile Random Forest using clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. *Geoderma* **2020**, *375*, 114503. [[CrossRef](#)]
152. Voltz, M.; Arrouays, D.; Bispo, A.; Lagacherie, P.; Laroche, B.; Lemerrier, B.; Richer-de-Forges, A.; Sauter, J.; Scnebelen, N. Possible futures of soil-mapping in France. *Geoderma Reg.* **2020**, *23*, e00334. [[CrossRef](#)]
153. Loiseau, T.; Arrouays, D.; Richer-de-Forges, A.C.; Lagacherie, P.; Ducommun, C.; Minasny, B. Density of soil observations in digital soil mapping: A study in the Mayenne region. *Geoderma Reg.* **2021**, *24*, e00358. [[CrossRef](#)]
154. Luan, J.; Zhang, C.; Xu, B.; Xue, Y.; Ren, Y. The predictive performance of random forest models with limited sample size and different species traits. *Fish. Res.* **2020**, *227*, 105534. [[CrossRef](#)]
155. Mulder, V.L.; Lacoste, M.; Richer-de-Forges, A.C.; Arrouays, D. GlobalSoilMap France: High resolution spatial modelling the soils of France up to two meter depth. *Sci. Total Environ.* **2016**, *573*, 1352–1369. [[CrossRef](#)] [[PubMed](#)]

156. Zhou, T.; Geng, Y.; Ji, C.; Xu, X.; Wang, H.; Pan, J.; Bumberger, J.; Haase, D.; Lausch, A. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. *Sci. Total Environ.* **2021**, *755*, 142661. [[CrossRef](#)]
157. Liu, F.; Wu, H.; Zhao, Y.; Li, D.; Yang, J.L.; Song, X.; Shi, Z.; Zhu, A.X.; Zhang, G.L. Mapping high resolution national soil information grids of China. *Sci. Bull.* **2022**, *67*, 328–340. [[CrossRef](#)] [[PubMed](#)]
158. Pouladi, N.; Gholizadeh, A.; Khosravi, V.; Borůvka, L. Digital mapping of soil organic carbon using remote sensing data: A systematic review. *Catena* **2023**, *232*, 107409. [[CrossRef](#)]
159. Adhikari, K.; Hartemink, A.E.; Minasny, B.; Bou Kheir, R.; Greve, M.B.; Greve, M.H. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE* **2014**, *9*, e105519. [[CrossRef](#)]
160. Ma, Y.; Minasny, B.; Wu, C. Mapping key soil properties to support agricultural production in Eastern China. *Geoderma Reg.* **2017**, *10*, 144–153. [[CrossRef](#)]
161. Sena, N.C.; Veloso, G.V.; Fernandes-Filho, E.I.; Francelino, M.R.; Schaefer, C.E.G.R. Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern Brazil. *Geoderma Reg.* **2020**, *21*, e00268. [[CrossRef](#)]
162. Brevik, E.C. The potential impact of climate change on soil properties and processes and corresponding influence on food security. *Agriculture* **2013**, *3*, 398–417. [[CrossRef](#)]
163. Sun, W.; Li, S.; Zhang, G.; Fu, G.; Qi, H.; Li, T. Effects of climate change and anthropogenic activities on soil pH in grassland regions on the Tibetan Plateau. *Glob. Ecol. Conserv.* **2023**, *45*, e2532. [[CrossRef](#)]
164. Dharumarajan, S.; Kalaiselvi, B.; Suputhra, A.; Lalitha, M.; Vasundhara, R.; Anil Kumar, K.S.; Nair, K.M.; Hedge, R.; Singh, S.K.; Lagacherie, P. Digital soil mapping of soil organic carbon stocks in Western Ghats, South India. *Geoderma Reg.* **2021**, *25*, e00387. [[CrossRef](#)]
165. Kaya, F.; Mishra, G.; Francaviglia, R.; Keshavarzi, A. Combining digital covariates and machine learning models to predict spatial variation of soil cation exchange capacity. *Land* **2023**, *12*, 819. [[CrossRef](#)]
166. Jena, R.K.; Moharana, P.C.; Dharumarajan, S.; Sharma, G.K.; Ray, P.; Roy, P.D.; Ghosh, D.; Das, B.; Alsuhaibani, A.M.; Gaber, A.; et al. Spatial prediction of soil particle size fractions using digital soil mapping in the north eastern region of India. *Land* **2023**, *12*, 1295. [[CrossRef](#)]
167. De Vries, F.T.; Thebault, E.; Liiri, M.; Birkhofer, K.; Tsiafouli, M.A.; Bjørnlund, L.; Jørgensen, H.B.; Brady, M.V.; Christensen, S.; de Ruiter, P.C.; et al. Soil food web properties explain ecosystem services across European land use systems. *Proc. Natl. Acad. Sci. USA* **2013**, *35*, 14296–14301. [[CrossRef](#)]
168. Haghghi, F.; Gorji, M.; Shorafa, M. A study of the effects of land use changes on soil physical properties and organic matter. *Land Degrad. Dev.* **2010**, *21*, 496–502. [[CrossRef](#)]
169. Hinge, G.; Surampalli, R.Y.; Goyal, M.K. Prediction of soil organic carbon stock using mapping approach in humid India. *Environ. Earth Sci.* **2018**, *77*, 172. [[CrossRef](#)]
170. Loiseau, T.; Chen, S.; Mulder, V.L.; Dobarco, M.R.; Richer-de-Forges, A.C.; Lehmann, S.; Bourenname, H.; Saby, M.P.; Vaudour, E.; Gomez, C.; et al. Satellite data integration for soil clay content modelling at a national scale. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *82*, 101905. [[CrossRef](#)]
171. Tziolas, N.; Tsakiridis, N.; Ben-Dor, E.; Theocharis, J.; Zalidis, G. Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. *Remote Sens.* **2020**, *12*, 1389. [[CrossRef](#)]
172. Žižala, D.; Minařík, R.; Zádorová, T. Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions. *Remote Sens.* **2019**, *11*, 2947. [[CrossRef](#)]
173. Vaudour, E.; Gomez, C.; Lagacherie, P.; Loiseau, T.; Baghdadi, N.; Urbina-Salazar, D.; Loubet, B.; Arrouays, D. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *96*, 102277. [[CrossRef](#)]
174. Heiden, U.; d'Angelo, P.; Schwind, P.; Karlshöfer, P.; Müller, R.; Zepp, S.; Wiesmeier, M.; Reinartz, P. Soil reflectance composites—Improved thresholding and performance evaluation. *Remote Sens.* **2022**, *14*, 4526. [[CrossRef](#)]
175. Zepp, S.; Heiden, U.; Bachmann, M.; Möller, M.; Wiemeier, M.; Van Wesemael, B. Optimized bare soil compositing for soil organic carbon prediction of topsoil croplands in Bavaria using Landsat. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 287–302. [[CrossRef](#)]
176. Zepp, S.; Heiden, U.; Bachmann, M.; Wiesmeier, M.; Steininger, M.; Van Wesemael, B. Estimation of soil organic carbon contents in croplands of Bavaria from SCMaP soil reflectance composites. *Remote Sens.* **2021**, *13*, 3141. [[CrossRef](#)]
177. Mzid, N.; Castaldi, F.; Tolomio, M.; Pascucci, S.; Casa, R.; Pignatti, S. Evaluation of agricultural bare soil properties retrieval from Landsat 8, Sentinel-2 and PRISMA satellite data. *Remote Sens.* **2022**, *14*, 714. [[CrossRef](#)]
178. Sun, W.; Liu, S.; Zhang, X.; Li, Y. Estimation of soil organic matter content using selected spectral subset of hyperspectral data. *Geoderma* **2022**, *409*, 115653. [[CrossRef](#)]
179. Liu, Q.; Guo, L.; Wang, M.; Deng, D.; Lv, P.; Wang, R.; Jia, Z.; Hu, Z.; Wu, G.; Shi, T. Digital mapping of soil organic carbon density using newly developed bare soil spectral indices and deep neural network. *Catena* **2022**, *219*, 106603. [[CrossRef](#)]
180. 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](#)]

-
181. Shaik, R.U.; Periasamy, S.; Zeng, W. Potential assessment of PRISMA hyperspectral imagery for remote sensing. *Remote Sens.* **2023**, *15*, 1378. [[CrossRef](#)]
 182. Chabrilat, 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](#)]

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.