

Article

# Spatial Prediction of Soil Particle-Size Fractions Using Digital Soil Mapping in the North Eastern Region of India

Roomesh Kumar Jena <sup>1</sup>, Pravash Chandra Moharana <sup>2,\*</sup>, Subramanian Dharumarajan <sup>3</sup>, Gulshan Kumar Sharma <sup>4</sup>, Prasenjit Ray <sup>5</sup>, Partha Deb Roy <sup>1</sup>, Dibakar Ghosh <sup>1</sup>, Bachaspati Das <sup>1</sup>, Amnah Mohammed Alsuhaibani <sup>6</sup>, Ahmed Gaber <sup>7</sup> and Akbar Hossain <sup>8,\*</sup>

- <sup>1</sup> ICAR-Indian Institute of Water Management, Bhubaneswar 751023, India; roomesh.jena@icar.gov.in (R.K.J.); partha.roy@icar.gov.in (P.D.R.); dibakar.ghosh@icar.gov.in (D.G.); bachaspati.das1@gmail.com (B.D.)
- <sup>2</sup> ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur 440033, India
- <sup>3</sup> ICAR-National Bureau of Soil Survey and Land Use Planning, Regional Centre, Bengaluru 560024, India; dharumarajan.s@icar.gov.in
- <sup>4</sup> ICAR-Indian Institute of Soil and Water Conservation, Research Centre, Kota 324002, India; gulshan.sharma@icar.gov.in
- <sup>5</sup> ICAR-Indian Agricultural Research Institute, New Delhi 110012, India; prasenjit.ray@icar.gov.in
- <sup>6</sup> Department of Physical Sport Science, College of Education, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; amalsuhaibani@pnu.edu.sa
- <sup>7</sup> Department of Biology, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; a.gaber@tu.edu.sa
- <sup>8</sup> Division of Soil Science, Bangladesh Wheat and Maize Research Institute, Dinajpur 5200, Bangladesh
- \* Correspondence: pravash.moharana@icar.gov.in (P.C.M.); akbar.hossain@bwmri.gov.bd (A.H.)

**Abstract:** Numerous applications in agriculture, climate, ecology, hydrology, and the environment are severely constrained by the lack of detailed information on soil texture. The purpose of this study was to predict soil particle-size fractions (PSF) in the Ri-Bhoi district of Meghalaya state, India, using a random forest model (RF). For the modeling of soil particle-size fractions, we employed 95 soil profiles (456 depth-wise layers) gathered from a recent national land resource inventory as well as currently accessible environmental variables. Sand, silt, and clay content were predicted using the Random Forest model at varied depths of 0–5, 5–15, 30–60, 60–100, and 100–200 cm. Our results showed the  $R^2$  for sand was found to be 0.30 (0–5 cm), 0.28 (5–15 cm), and 0.21 (15–30 cm). For the sand, silt, and clay fractions, respectively, the concordance correlation coefficient (CCC) was found to be greater in the 0–30 cm, 0–60 cm, and 0–15 cm depths. When there is a reasonably close monitoring of the coverage probability with a confidence level along the 1:1 line, prediction interval coverage probability (PICP) gives a decent indicator of what to anticipate. The most crucial variables for the prediction of sand and silt were channel network base level (CNBL) and LS-Factor, whereas Min Temperature of Coldest Month ( $^{\circ}\text{C}$ ) (BIO6) was discovered for clay prediction. For all three soil texture fractions, the range between the 5% lower and 95% higher prediction bounds was large, indicating that the existing spatial predictions may be improved. The maps of soil texture were significantly more precise, and they accurately depicted the spatial variations of particle-size fractions. Additionally, there is still a need to investigate novel methodologies for extensive digital soil mapping, which will be very advantageous for many international initiatives.

**Keywords:** digital soil mapping; environmental variables; random forest; uncertainty analysis; particle-size fractions



**Citation:** Jena, R.K.; Moharana, P.C.; Dharumarajan, S.; Sharma, G.K.; Ray, P.; Deb Roy, P.; 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. <https://doi.org/10.3390/land12071295>

Academic Editor: Chuanrong Zhang

Received: 15 May 2023  
Revised: 16 June 2023  
Accepted: 24 June 2023  
Published: 27 June 2023



**Copyright:** © 2023 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

As Digital soil mapping (DSM) is gaining popularity, there is a high need for data sharing as an outcome. Many digital soil maps are accessible, each with a different quality concerning resolution, extent, modeling technique, uncertainty, etc. [1,2]. Traditional soil mapping methods mostly rely on fieldwork and seldom offer details on the

spatial distribution of soil characteristics at the appropriate resolution throughout the area [3,4]. Additionally, it is challenging to accurately capture the real status of the dynamic soil properties when conducting traditional field surveys to map soil spatial variations, particularly at regional, national, or global scales [5,6]. As a result, reliable techniques and models are required to forecast soil characteristics at a certain scale or location. With the use of modern techniques such as digital soil mapping, soil characteristics can now be precisely predicted due to significant advancements in remote sensing techniques and machine learning approaches [3,7]. Artificial neural networks, decision trees, and support vector machines, among other machine learning approaches, have recently been suggested as alternatives to traditional soil mapping methods [8–10]. Using environmental factors, DSM approaches have been used to map the characteristics of soil [11]. With the use of correlations between soil and environmental factors deduced from topographical features and satellite images, these approaches were created to surpass the drawbacks of the traditional soil mapping approach [3,5]. A growing need for accurate and useful soil data led to the founding of the *GlobalSoilMap* consortium [1,10]. This partnership has taken on the challenge of creating soil property maps utilizing DSM methods at a fine resolution. Many nations have an extensive collection of historical soil data, which includes soil maps at various scales, soil point data gathered over many years, environmental covariate data, and a network of collaborators who have contributed to developing the soil information over time.

The soil's soil particle size fractions (PSF) are an essential physical characteristic that directly affects a variety of soil functions, such as soil fertility and water retention. Studies have shown that topography plays an important role in explaining differences in the PSF [8,12]. Soil texture/PSF has traditionally been expressed by polygons, with each polygon representing a textural class. The textural composition of the region marked within a polygon, however, may be highly ambiguous due to the existence of intra-polygon texture heterogeneity intra-polygon [13]. In order to properly measure intra-polygon textural diversity, a different approach to solving this issue is to numerically map various PSF. There are several modern techniques to map PSF utilizing a geostatistic-scoring kriging strategy [14–16]. McBratney et al. [3] used a Jenny-like formulation [17] to quantitatively explain the connections between soil and other spatially referenced elements. Seven criteria are taken into account: s (soil, other properties of the soil at a point); c (climate, climatic properties of the environment at a point); o (organisms, plants, or animals; human activity); r (topography, landscape qualities); p (parent material, lithology); a (age, the time factor); and n (space, spatial location) are the other features of the soil at a point.  $S_a = f(s; c; o; r; p; a; n)$  may thus be represented as the model to predict a soil attribute  $S_a$  at a given place, where the s stands for soil information from a soil map, remote or proximal sensing, or expert knowledge. These methods described soil formation factors and management inputs using a range of data mining techniques (i.e., linear models, regression trees, multivariate adaptive regression splines, and artificial neural networks) and covariates. The majority of the variables were acquired from remote sensing and digital elevation models. Hyperspectral remote sensing [18–20], multispectral remote sensing [21,22], and radar remote sensing [23] have all been used in certain attempts. Recently, some investigations utilizing DSM approaches have been carried out on the regional variation of soil parameters [24–26]. Hengl et al. [2,7] used DSM techniques to predict soil parameters such as organic carbon, pH, particle size fractions, and bulk density at continental (1 km) and global levels (250 m). Vagen et al. [27] used the DSM approach to map SOC, pH, sand, and the total of bases throughout Africa. In a low-relief location, Wang et al. [28] used land surface temperature from MODIS satellite data to estimate soil texture. Geostatistical techniques were used by Walder et al. [29] to forecast the contents of clay, silt, sand, and humus in a German floodplain.

The objectives of this investigation are to (1) model and map soil particle-size fractions (PSF) for the entire soil profile using the DSM approach, which was adopted from the *GlobalSoilMap* project, and (2) identify the key environmental covariates that affect the

particle-size fractions in the study area. We anticipate that our results will enhance and update the existing soil texture information system with fresh, high-resolution soil texture maps that might be beneficial to stakeholders and end users.

## 2. Materials and Methods

### 2.1. Study Area

The study was carried out in the Ri-Bhoi district of the northeastern Indian state of Meghalaya (Figure 1). The district lies between 25°48'54.36" N to 26°04'40.08" N latitude and 91°20'40.56" E to 92°16'33.96" E longitude and covers an area of about 2359 km<sup>2</sup>. The Shillong Plateau physiographic division encompasses the study area.

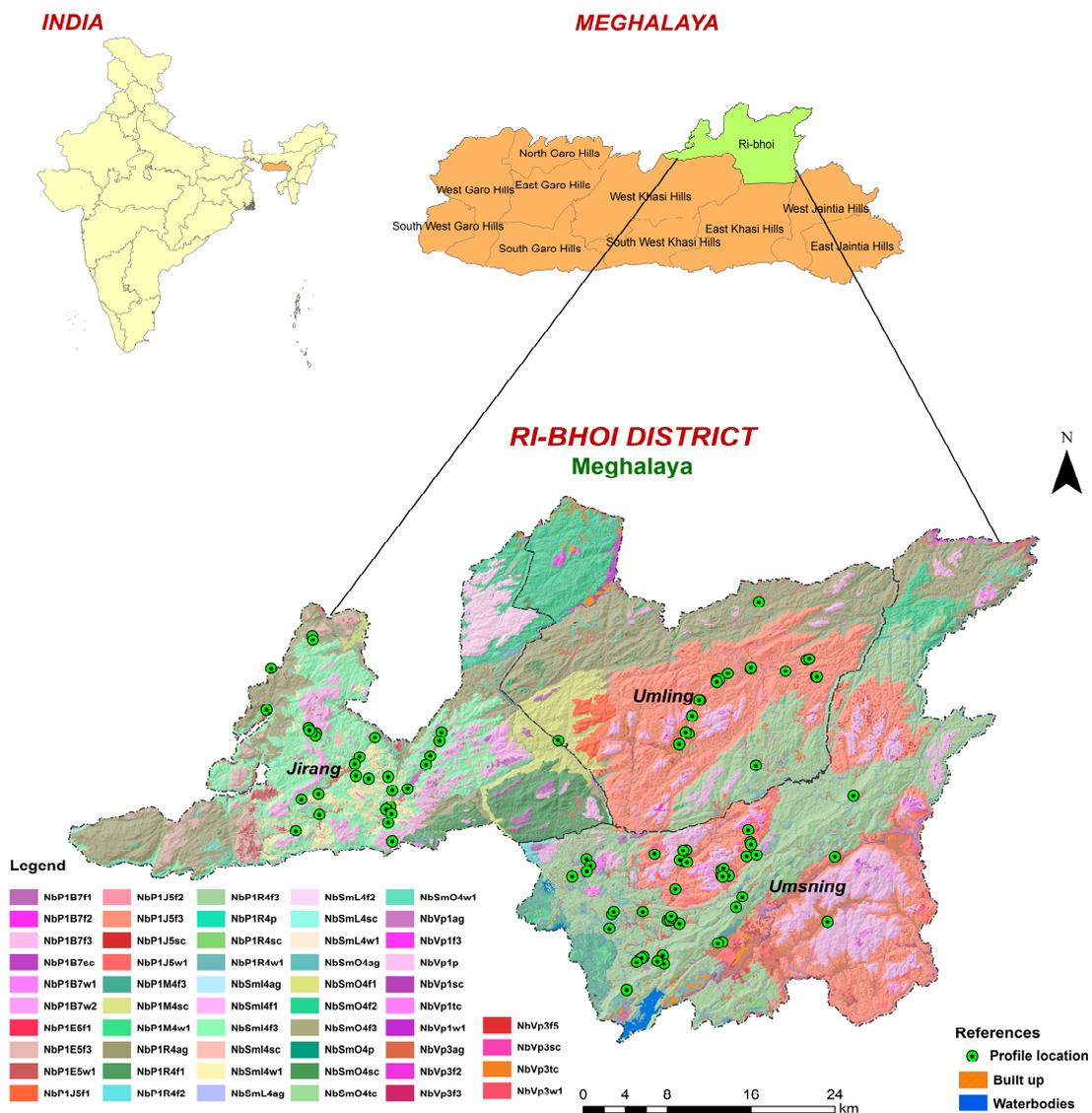


Figure 1. Location map and research area soil profile points.

The geography of the region is undulating, with a minimum elevation of 60 m on the northern side bordering Assam and a maximum elevation of 1400 m on the southern side. The research region is characterized by cold winters and hot, humid summers typical of a subtropical climate. Potential evapotranspiration (PET) ranges from 800 to 1300 mm, and the average annual rainfall ranges from 2000 to 4000 mm. Due to seasonal dry spells that occur throughout the post-monsoon period, or from November to February, the area has a water deficit of 250 to 350 mm [30]. There are 176 rainy days on average per year. The average summer temperature rises to 26 °C, and the average winter temperature drops to

9 °C. At higher elevations, the mean annual soil temperature (MAST) is less than 22 °C but more than 15 °C, but at lower elevations, it is greater than 22 °C. In sub-montane, valleys, and lower elevation parts of the physiographic unit, the soil temperature regime is hyperthermic, whereas the soil moisture regime is characterized by udic [31,32]. The study area forms a part of the stable Meghalaya plateau of northeastern India which is an extension of the Deccan plateau of southern India. Geomorphologically, Ri-Bhoi is a hilly district with inter-montane valleys. The lithology of the hills comprises Archaean to Proterozoic Gneissic complex rocks, which are highly deformed, fractured, and fissured in nature. The dominant geology of the district is granite gneiss, hornblende biotite gneiss, migmatite, banded granitoid, and unclassified gneiss with pegmatite and schist, followed by grey/pink granite.

## 2.2. Land Resource Inventory

Detailed land resource inventory (LRI) was carried out to study the soils for their field morphology under the project entitled “LRI of Ri-Bhoi district, Meghalaya for optimal agricultural land use planning using geospatial techniques” [33]. The soil survey was carried out by transect approach from higher elevation to lower elevation covering all the identified landscape ecological units (LEUs). The LEU map was generated by integrating land use and land cover information with topography (landforms and slope), considering the climate, relief, flora, and parent material as the most influencing factor compared to time, whereas the parent material and climate are the constant factors in the recent years as discussed earlier. A total of 456 soil samples from 95 profiles chosen in selected transects were studied. In each LEU, at least one soil profile pit (2.5 × 2.5 m and to a depth of underlying rock or weathered parent material) observation was made. According to soil survey recommendations, the horizon-wise morphological characteristics, such as depth, color, texture, structure, consistency, gravel content, the presence of mottles and nodules, etc., were recorded in the soil profile sheet [32].

## 2.3. Particle Size Analysis

Each pedological horizon in each profile was sampled independently, and the samples were then sent to the lab for examination. Before passing through a 2-mm sieve and being analyzed in the lab for particle size analysis using the international pipette technique, samples were air-dried at room temperature [34]. In the present study most frequently used so-called international system (IS) has been considered for particle size limits [35]. Sand (0.02–2 mm), silt (0.002–0.02 mm), and clay (0.002 mm) were analyzed and reported in percentage mass ( $\text{g } 100 \text{ g}^{-1}$ ). As the *GlobalSoilMap* project guidelines specified, these values were multiplied by 10 to convert the unit into  $\text{g kg}^{-1}$ .

## 2.4. Data Harmonization

To align the soil depth interval with the *GlobalSoilMap* depth criteria [1], the particle size of the profile datasets was preprocessed using equal-area spline functions [36]. For easier comparison of profile properties, the *GlobalSoilMap* program uses six standard depth horizons: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm [1,37]. The *mpspline2* R-package’s equal-area spline functions were used for this [38]. It was noticed that at least two horizons of soil data must be present for spline fitting. Although the soils of the study area were described as being from deep to very deep [33], all 95 soil profiles in the current study had at least two horizons, making it possible to harmonize the data into the six standard soil depths used in the *GlobalSoilMap* program. Particle size contents were determined during spline fitting in increments of 1 cm from the top surface of the soil to 200 cm below it.

## 2.5. Environmental Covariates

SRTM data were used to create a digital elevation model (DEM) with a 30m resolution, which was then processed using the ArcGIS 10 data management toolbox. Over the past ten years, many attempts have been made to choose and provide effective environmental

covariates in line with the desired soil properties and landscapes. In DSM, climate and terrain factors have been extensively used. Elevation, slope (Slp), aspect (Asp), convergence index (CI), valley depth (VD), channel network base level (CNBL), relative slope position (RSP), slope curvatures [downslope (DC), local (LC), local downslope (LDC), local upslope (LUC), and upslope (USC)], topographic wetness index (TWI), topographic ruggedness index (TRI), topographic position index (TPI), LS factor (LSF), Multi Normalized Difference Vegetation Index for Rabi Season (NDVI\_R), Summer Season (NDVI\_S), and Kharif Season (NDVI\_K) (MOD13Q1) were utilized as covariates in addition to DEM features to forecast the characteristics of the soil (Table 1). The world's grids for the research region were derived from the raster data on bioclimatic variables at 30 arc seconds resolution that was downloaded from <http://worldclim.org/current>; accessed on 25 May 2023 for the whole globe. In research on digital soil mapping, all 19 bioclimatic variables were used as covariates [39]. In the current study, the *Scorpan* factor included 18 relief, 3 species, and 19 climate elements (Figure 2).

**Table 1.** Covariates related to the environment are utilized to predict soil texture components.

Soil Forming Factors	Predictor	Abbreviation	Resolution	Min.	Max.	Mean
Relief	Elevation (m)	Elevation		76.0	1092.0	637.2
	Slope	Slope		0.0	0.5	0.1
	Aspect	Aspect		0.0	6.2	3.1
	Topographic Positioning Index	TPI		4.4	88.5	26.1
	Topographic Wetness index	TWI		6.2	18.2	10.6
	LS-factor	LSF		0.0	16.2	2.2
	Channel Network Base Level	CNBL		76.0	845.7	545.5
	Multi-resolution Index of Valley Bottom Flatness	MRVBF		0.0	3.4	1.2
	Multi-resolution Ridge Top Flatness	MRRTF	30 m	0.0	3.7	0.4
	Relative slope position	RSP		0.0	1.0	0.4
	Valley Depth	VD		0.0	417.6	137.8
	Vertical Distance to Channel	VDC		0.0	303.2	99.5
	Local Curvature	LC		−1.2	1.2	−0.2
	Downslope Curvature	DC		−1.1	0.2	−0.4
	Upslope Curvature	USC		0.1	1.2	0.2
	Convergence Index	CI		−58.2	70.6	−5.4
	Local Downslope Curvature	LDC		−0.8	1.3	0.1
Local Upslope Curvature	LUC		−1.3	−0.1	−0.5	
Vegetation	Normalized Difference Vegetation Index—Rabi	NDVI_R		3919.5	7188.8	5880.7
	Normalized Difference Vegetation Index—Summer	NDVI_S	250 m _16 days	4246.7	6722.1	5773.7
	Normalized Difference Vegetation Index—Kharif	NDVI_K		5706.1	7933.0	6983.3
Climate	Annual Mean Temperature (°C)	BIO1		18.5	23.8	20.9
	Mean Diurnal Range (Mean of monthly (max temp – min temp)) (°C)	BIO2		−0.4	8.9	8.3
	Isothermality (Bio_2/Bio_7) (× 100)	BIO3	1 km	−1.3	46.5	44.5
	Temperature Seasonality (standard deviation × 100) (°C)	BIO4		−0.8	382.9	364.5
	Max Temperature of Warmest Month (°C)	BIO5		0.0	32.6	29.2
	Min Temperature of Coldest Month (°C)	BIO6		−4.2	13.6	10.5

Table 1. Cont.

Soil Forming Factors	Predictor	Abbreviation	Resolution	Min.	Max.	Mean
	Temperature Annual Range (BIO5-BIO6) (°C)	BIO7		18.0	87.0	19.6
	Mean Temperature of Wettest Quarter (°C)	BIO8		21.8	6310.7	91.2
	Mean Temperature of Driest Quarter (°C)	BIO9		13.3	5958.4	78.7
	Mean Temperature of Warmest Quarter (°C)	BIO10		21.8	7090.0	99.5
	Mean Temperature of Coldest Quarter (°C)	BIO11		13.3	91.5	16.3
	Annual Precipitation (mm)	BIO12		25.9	4559.0	3392.9
	Precipitation of Wettest Month (mm)	BIO13		460.0	1177.0	848.7
	Precipitation of Driest Month (mm)	BIO14		6.0	9.0	7.1
	Precipitation Seasonality (Coefficient of Variation) (mm)	BIO15		90.3	106.2	101.8
	Precipitation of Wettest Quarter (mm)	BIO16		1173.0	2914.0	2137.8
	Precipitation of Driest Quarter (mm)	BIO17		35.0	45.0	40.7
	Precipitation of Warmest Quarter (mm)	BIO18		1030.0	2914.0	2130.2
	Precipitation of Coldest Quarter (mm)	BIO19		35.0	45.0	40.7

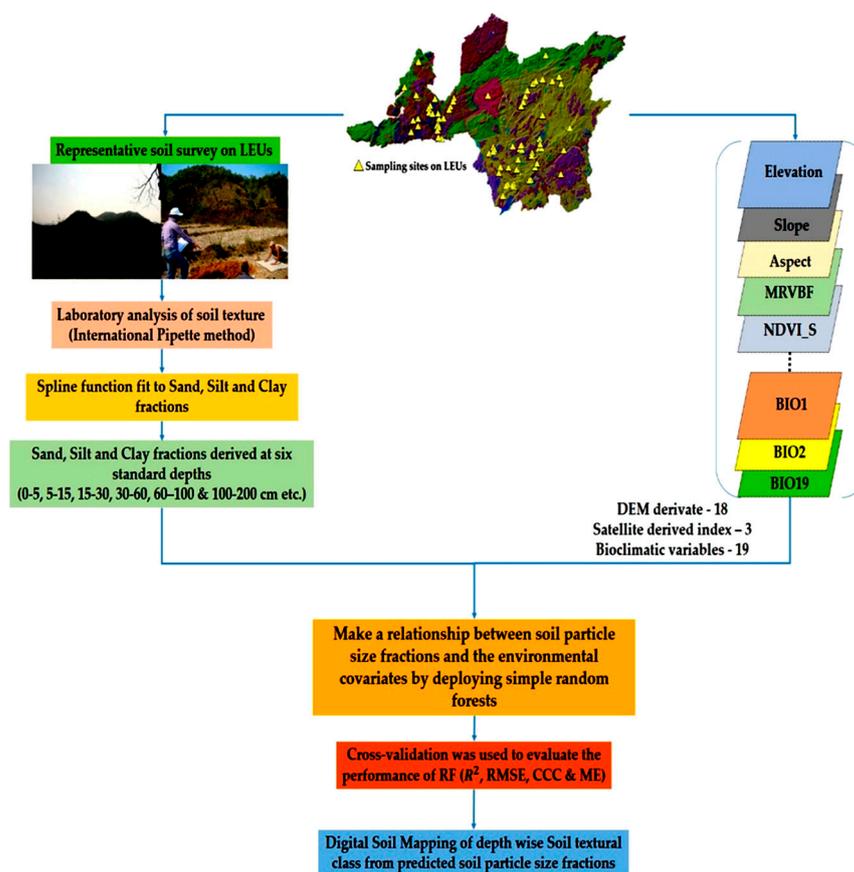


Figure 2. Methodology Flowchart of the study.

### 2.6. Modeling

The regression tree ensemble known as the random forest (RF) modeling creates a large number of regression trees before aggregating them to obtain a final prediction. The RF uses Out-Of-Bag (OOB) data to estimate errors. The word “mtry” refers to the number of environmental variables in each random subset, while “ntree” refers to the number of trees in the forest. By iterating mtry values from 1 to the complete number of significant variables and ntree values from 100 to 5000 by increments of 100, it was determined which values returned the lowest OOB error. The significance of every variable was likewise

estimated using RF. The values of the other variables remain the same while the values of each variable are randomly permuted in the OOB data. The changed OOB data were forecasted, and a variable importance metric was produced by comparing the MSE of the permuted and original OOB data. All environmental factors were first employed in the modeling process, and then, based on the relative importance, which has a range of 0 to 100%, variables with importance values less than 7% were deemed insignificant and eliminated from the models. The final models made use of the remaining variables. The ‘randomForest’ package of R 3.2.5 software [40] was used for all models [41].

### 2.7. Model Evaluation

The calibration dataset made up 80% of the total soil profiles, and the remaining soil profiles constituted validation data. The calibration and validation set samples were divided by using the caret package (Classification and REgression Training). In the current study, we calculated four generally used error metrics, including Lin’s concordance correlation coefficient (CCC), mean error (ME), root mean square error (RMSE), and coefficient of determination ( $R^2$ ). The following equations were provided:

$$\text{Coefficient of determination } (R^2) = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (\hat{p}_i - \hat{o}_i)^2} \quad (1)$$

$$\text{Mean error (ME)} = \frac{1}{n} \sum_{i=1}^n (p_i - o_i) \quad (2)$$

$$\text{Root mean squared error (RMSE)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2} \quad (3)$$

where,  $p_i$  and  $o_i$  are predicted and observed values,  $\hat{p}_i$  and  $\hat{o}_i$  are means of predicted and observed values.

$$\text{Lin's concordance correlation coefficient (CCC)} = \frac{2\rho\sigma_0\sigma_p}{\sigma_0^2 + \sigma_p^2 + (\mu_0 - \mu_p)^2} \quad (4)$$

$\mu_0$  and  $\mu_p$  are the means of observed and predicted values and  $\sigma_0^2$  and  $\sigma_p^2$  are corresponding variance  $\rho$  is the Pearson correlation coefficient.

### 2.8. Uncertainty Assessment

It is very obvious that any machine learning techniques generate estimates which have errors. To measure uncertainties in our study, we deployed the popular non-parametric approach called Bootstrap to quantify prediction uncertainties. More simply, one repeated sampling technique that uses replacement samples chosen at random is the bootstrap. The old dataset and the new bootstrapped dataset both have the same sample size and probability distribution. For the calibration datasets, we produced 50 bootstrapped datasets at each depth. The final assessment indicator and the projected outcome, respectively, were thought to be the average accuracy in validation datasets and the average PSF prediction. Maps of the 0.05 and 0.95 quantiles for each fraction and depth were created by the projection. This means that levels of uncertainty for the predicted percentages of clay, silt, and sand were present for each pixel position in the research region and for each depth.

We utilized the prediction interval coverage probability (PICP) criteria, which involve estimating the percentage of observations in the overall dataset that are contained within the prediction interval at a specific confidence level for that data point to assess the prediction uncertainties. In this investigation, the following confidence intervals were used: 99, 97.5, 95, 90, 80, 60, 40, 20, 10, and 5%. To assess if the defined uncertainties have been successfully calculated, one should anticipate the PICP value or proportion to be near the associated

confidence level [42,43]. The PICP is only the percentage of observations that are contained within the associated prediction interval for each depth [19,44].

$$\text{PICP} = \frac{1}{v} \text{ count } i \quad (5)$$

$$i = Cl_i^l \leq \text{obs} \leq Cl_i^u \quad (6)$$

'v' is the number of observations in the validation dataset, and  $Cl_i^l$  and  $Cl_i^u$  are the lower (5%) and the upper limits (95%) of the confidence interval, respectively. When the PICP value is near 100 (1 - a)%, such as the 90% confidence interval, the uncertainty is considered to be at its lowest level. When using the PICP, an estimate of 90% for a 90% prediction interval should be produced if the uncertainty estimates have been accurately defined. Additionally, the accuracy measurements' standard deviations ( $R^2$ , RMSE, and ME) derived from 50 bootstrapped datasets were also utilized to reflect, to some extent, the predictability of uncertainty.

### 3. Results and Discussion

#### 3.1. Descriptive Statistics

Table 2 provides statistical information about the proportions of sand, silt, and clay fitted using equal-area quadratic splines at various depths based on the soil profiles.

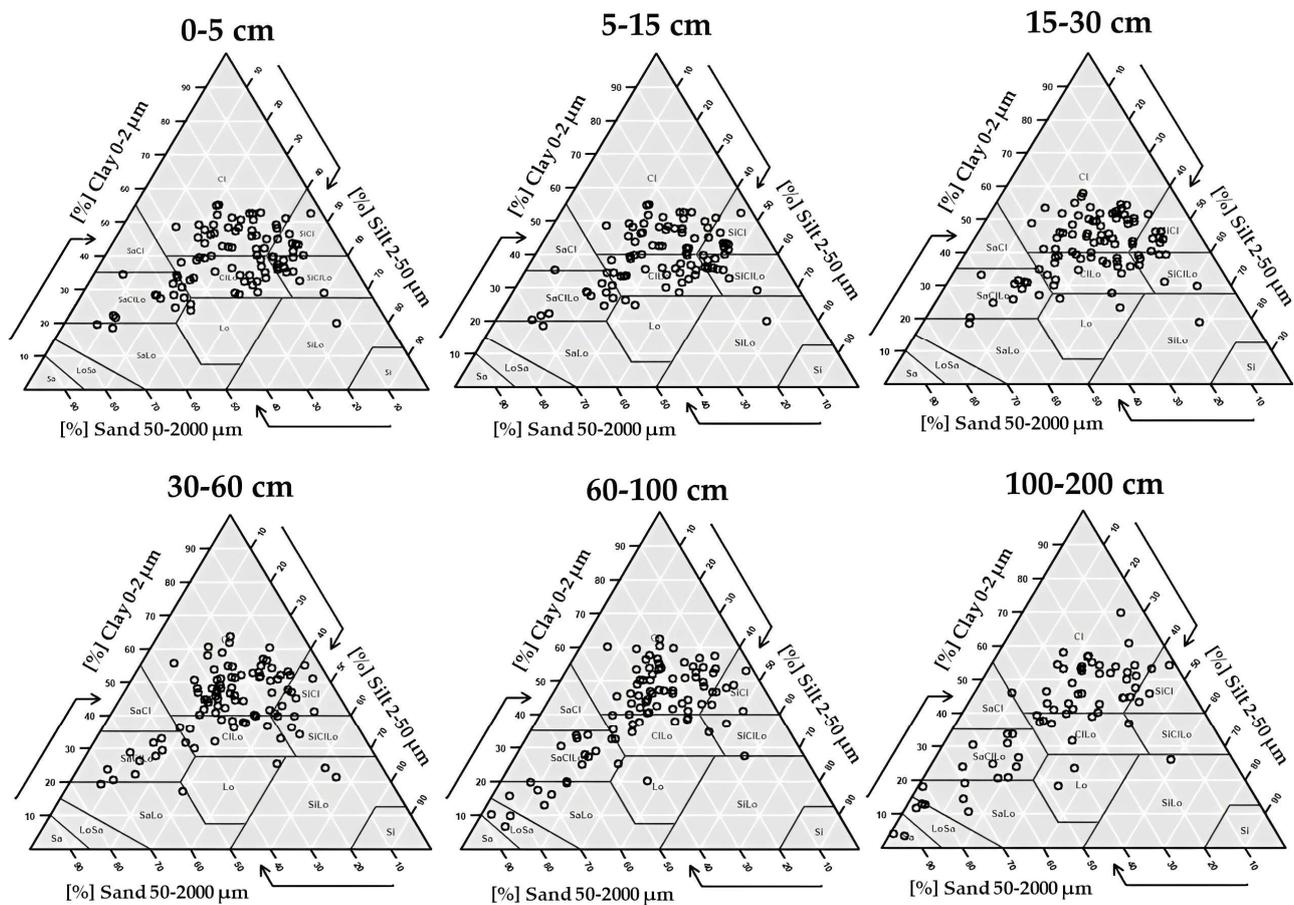
**Table 2.** The depth-wise statistical description of the spline-fitted sand, silt, and clay percentages.

Depth (cm)	Min (%)	Max (%)	Mean (%)	Median (%)	SD (%)	CV (%)	Skewness	Kurtosis
Sand								
0–5	2.93	72.29	28.81	25.32	14.73	51.14	0.96	0.54
5–15	3.01	70.73	28.67	25.40	14.55	50.74	0.96	0.58
15–30	8.28	70.04	27.98	24.85	13.94	49.83	0.96	0.56
30–60	3.90	72.65	27.68	25.73	14.93	53.94	1.09	0.98
60–100	1.94	87.22	32.32	26.65	19.48	60.28	1.09	0.62
100–200	1.57	95.03	35.56	28.36	22.85	64.24	0.94	−0.04
Silt								
0–5	7.13	67.26	32.11	30.32	11.44	35.64	0.20	−0.04
5–15	7.23	67.32	32.06	31.37	11.32	35.32	0.23	0.06
15–30	6.97	67.76	30.66	30.61	11.19	36.49	0.53	0.57
30–60	7.48	65.62	28.69	27.60	11.33	39.49	0.64	0.62
60–100	2.48	57.46	25.83	24.60	10.78	41.72	0.32	0.02
100–200	0.80	58.53	24.46	23.98	11.43	46.74	0.00	−0.04
Clay								
0–5	18.51	55.05	39.08	39.50	9.06	23.18	−0.24	−0.70
5–15	18.47	54.89	39.27	39.73	8.84	22.52	−0.28	−0.60
15–30	18.41	57.86	41.35	42.90	9.17	22.18	−0.47	−0.40
30–60	17.09	63.73	43.62	45.23	10.64	24.40	−0.56	−0.29
60–100	7.91	62.48	41.89	44.40	12.94	30.89	−0.82	0.02
100–200	3.46	69.67	39.99	43.86	14.84	37.10	−0.67	−0.43

In this study, spline functions were used to estimate the amounts of sand, silt, and clay for each of the six depth intervals. The estimates were then rounded up to ensure that the PSF total was 100%. At every depth layer, the mean silt and clay contents were noticeably higher than the mean sand contents. Although the mean silt and sand contents both slightly decreased as depth increased, the mean clay content slightly increased. The average sand content was  $28.81 \pm 14.73$ ,  $28.67 \pm 14.55$ ,  $27.98 \pm 13.94$ ,  $27.68 \pm 14.93$ ,  $32.32 \pm 19.48$ , and  $35.56 \pm 22.85\%$ , and the average silt content was  $32.11 \pm 11.44$ ,  $32.06 \pm 11.32$ ,  $30.66 \pm 11.19$ ,  $28.69 \pm 11.33$ ,  $25.83 \pm 10.78$ , and  $24.46 \pm 11.43\%$  for 0–5, 5–15, 15–30, 30–60, 60–100, and

100–200 cm, respectively. With increasing depth, the clay fraction's standard deviation (SD) showed an upward trend. All depth layers' sand and silt contents varied more, with coefficients of variation (CV%) ranging from 49.83 to 64.24% and 35.32 to 46.74%, respectively. Conversely, clay content had lower variability, with the CVs between 22.18 and 37.10%.

Figure 3 shows the texture triangle for observed values where we could observe that the dominant soil textural classes were clay and clay loam across all the soil depths, which was about 75% of the samples under the study, followed by silty clay loam, sandy clay loam, and silty clay spread over different landforms, e.g., denudational high and low hills, upper and lower plateaus. In some valleys, heavy textured soils were observed because clay illuviation processes were prominent in this region as the lateral, as well as the surface flow of water, is more.



**Figure 3.** Textural diagram of the sand, silt, and clay contents of the splined soil profile.

Correlation analysis between sand fraction and covariates used in the model (Table 3) showed that BIO2, BIO9, BIO17, BIO19, CNBL, LUC, and TWI have a significant correlation with sand fraction in the surface soil, whereas slope, NDVI\_R, TWI, and elevation have a significant correlation with silt fraction. BIO2, BIO3, BIO6, BIO17, BIO19, LUC, and MRVBF have a significant correlation with clay fraction.

**Table 3.** Correlation analysis between covariates and surface soil textural fractions.

Covariates	Sand	Silt	Clay
Aspect	0.156	−0.064	−0.173
BIO1	0.181	−0.101	−0.169
BIO2	−0.204 *	0.004	0.332 **
BIO3	−0.198	0.056	0.256 *
BIO4	−0.011	−0.044	0.074
BIO5	0.154	−0.088	−0.141
BIO6	0.200	−0.092	−0.212 *
BIO7	−0.021	−0.051	0.099
BIO8	0.164	−0.094	−0.150
BIO9	0.195	−0.101	−0.192
BIO10	0.163	−0.093	−0.149
BIO11	0.195	−0.101	−0.192
BIO12	−0.059	0.085	−0.011
BIO13	−0.043	0.072	−0.021
BIO14	0.025	0.116	−0.190
BIO15	−0.096	0.033	0.118
BIO16	−0.076	0.087	0.014
BIO17	0.243 *	−0.103	−0.267 **
BIO18	−0.068	0.078	0.013
BIO19	0.243 *	−0.103	−0.267 **
CI	0.021	−0.047	0.026
CNBL	−0.222 *	0.156	0.167
DC	−0.051	0.016	0.064
Elevation	−0.167	0.113	0.131
LC	−0.084	0.054	0.069
LDC	−0.083	0.037	0.089
LSF	0.026	0.066	−0.128
LUC	−0.234 *	0.113	0.240 *
MRRTF	−0.008	−0.005	0.020
MRVBF	−0.042	−0.111	0.212 *
NDVI_K	0.155	−0.149	−0.063
NDVI_R	0.157	−0.236 *	0.045
NDVI_S	0.163	−0.144	−0.083
RSP	0.161	−0.108	−0.125
Slope	−0.112	0.228 *	−0.108
TPI	−0.041	0.149	−0.124
TWI	0.245 *	−0.238 *	−0.098
USC	−0.078	0.085	0.020
VD	−0.119	0.102	0.064
VDC	0.178	−0.137	−0.115

\*\* Correlation is significant (0.01 level); \* Correlation is significant (0.05 level).

### 3.2. Performance of Model Prediction and Uncertainty Estimation

To determine the model's accuracy, the average values of the four evaluation indices ( $R^2$ , RMSE, CCC, and ME) in the validation dataset were chosen (Table 4). To account for dimensional effects, we determined the relative RMSE (RRMSE) by dividing the RMSE by the mean measured values. The explanatory power of the prediction findings to the measured values is represented by the  $R^2$  statistic. Our results showed the  $R^2$  of RF for sand was 0.30 (0–5 cm depth), 0.28 (5–15 cm depth), 0.21 (15–30 cm depth), 0.02 (30–60 cm depth), 0.02 (60–100 cm depth), and 0.14 (100–200 cm depth). We found that the top two depths, 0–5 cm and 5–15 cm, had obvious improvement; below these depths, the accuracy of the RF model was inconsistent. In the case of sand and clay fractions, the accuracy decreased with depth, but prediction accuracy was better in the case of silt fraction across all the six depths, i.e., 11.3, 11.2, 11.7, 11.5, 13.6, and 9.9%. The CCC was found higher on the surface three layers (0.29–0.35), four layers (0.26–0.29), and two layers (0.33–0.34) in the case of sand, silt, and clay fractions, respectively. In three different sites in Kentucky, Thompson et al. [45]

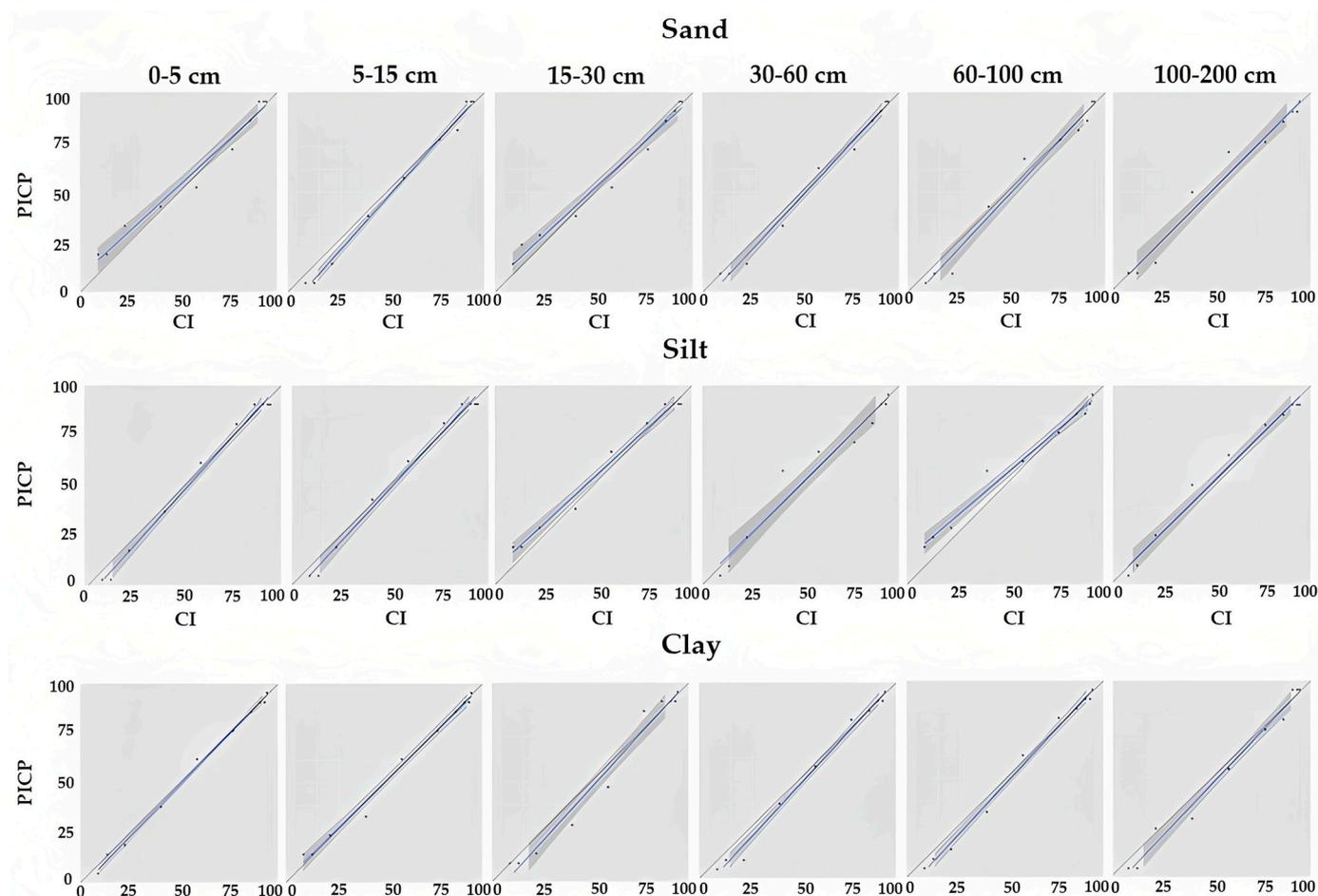
predicted the surface sand and clay contents and discovered RMSE values of 1.92 to 50.89 for sand and values of 2.61 to 22.73 for silt. The intricacy of soil property fluctuations and soil formation circumstances in the particular topography is demonstrated by the comparatively high RMSE values in this study. Wang et al. [28] utilized DSM to forecast soil parameters and found that for sand, clay, and physical clay, the RMSE values were 10.69, 4.57, and 12.99, respectively.

**Table 4.** Prediction accuracy (Investigation of 50 bootstrap iterations for various soil texture fractions).

Depth (cm)	R <sup>2</sup>	CCC	RMSE	ME	RRMSE
Sand					
0–5	0.30 ± 0.02	0.35 ± 0.01	11.20 ± 0.07	7.51 ± 0.14	0.39
5–15	0.28 ± 0.02	0.34 ± 0.01	11.10 ± 0.14	7.23 ± 0.13	0.38
15–30	0.21 ± 0.01	0.29 ± 0.01	10.70 ± 0.12	7.06 ± 0.09	0.38
30–60	0.02 ± 0.01	0.06 ± 0.01	15.20 ± 0.17	9.62 ± 0.19	0.55
60–100	0.02 ± 0.01	0.01 ± 0.01	21.50 ± 0.05	10.26 ± 0.13	0.66
100–200	0.14 ± 0.02	0.12 ± 0.02	20.80 ± 0.26	4.18 ± 0.34	0.58
Silt					
0–5	0.28 ± 0.02	0.29 ± 0.01	11.30 ± 0.11	−5.17 ± 0.05	0.35
5–15	0.28 ± 0.02	0.29 ± 0.01	11.20 ± 0.14	−5.17 ± 0.08	0.35
15–30	0.22 ± 0.02	0.28 ± 0.02	11.70 ± 0.19	−4.23 ± 0.18	0.38
30–60	0.26 ± 0.01	0.26 ± 0.01	11.50 ± 0.07	−5.99 ± 0.05	0.40
60–100	0.03 ± 0.01	0.09 ± 0.01	13.60 ± 0.08	−6.86 ± 0.11	0.52
100–200	0.01 ± 0.00	−0.02 ± 0.01	9.90 ± 0.08	1.62 ± 0.14	0.40
Clay					
0–5	0.25 ± 0.02	0.34 ± 0.01	7.80 ± 0.09	−1.42 ± 0.12	0.20
5–15	0.26 ± 0.02	0.33 ± 0.01	7.80 ± 0.08	−1.44 ± 0.06	0.19
15–30	0.11 ± 0.01	0.17 ± 0.01	9.30 ± 0.03	−2.04 ± 0.07	0.22
30–60	0.05 ± 0.01	0.11 ± 0.02	10.70 ± 0.13	−3.03 ± 0.1	0.24
60–100	0.01 ± 0.00	0.02 ± 0.01	12.50 ± 0.14	−2.59 ± 0.21	0.29
100–200	0.04 ± 0.01	−0.09 ± 0.01	15.30 ± 0.08	−6.82 ± 0.09	0.38

R<sup>2</sup>: coefficient of determination; CCC: concordance correlation coefficient; RMSE: root mean square error; ME: mean error; RRMSE: Relative RMSE.

A way to evaluate how well the predicted uncertainties perform during testing is the PICP approach. Uncertainty representation is a crucial aspect of DSM. DSM models are not only expected to deliver accurate soil predictions at a given location, but their suitability to deliver maps should encompass the ability to predict uncertainty. PICP was used to evaluate the uncertainty of the prediction. The PICP is simply the proportion of observations at each depth that are encapsulated by the corresponding prediction interval. This is accomplished simply by evaluating the prediction intervals' coverage at various levels of confidence around an observed value. The plots in Figure 4 give a clear picture of what to anticipate when the coverage probability and confidence level are tracked along the 1:1 line relatively close. In the case of silt, the prediction interval was closer to the 1:1 line as compared to sand and clay fractions. The probabilities above the 1:1 line indicate mild overprediction of the uncertainty range, which may be due to the uncertainty method itself [46].



**Figure 4.** Based on the validation of the model for various soil textural fractions, depth-wise plots between the prediction interval coverage probability (PICP) and confidence level (CI).

### 3.3. Importance of Environmental Variables

Slope, CNBL, TWI, LSF and elevation in terrain and relief factor, NDVI\_S and NDVI\_R in vegetation factor, bioclimatic factors such as BIO18, BIO16, BIO15, BIO12, BIO6, and BIO3 were important variables in this study. Figure 5 depicts the importance of the variable for the 0–5 cm of soil depth. CNBL and LSF was the most important variable for the prediction of sand and silt, while BIO6 was found in clay prediction. The aspect and CNBL were also identified by Rudiyanto et al. [47] as significant variables for predicting sand and silt with RF. The channel network base level (CNBL), BIO6, BIO3, and NDVI\_S were the most important variable for the prediction of clay across all depths. In lower depths, topographic factors such as downslope curvature, aspect, and convergence index played a vital role over bioclimatic variables in the prediction of clays. Similar findings were observed in the case of sand fractions below 15 cm depth of soil. These results suggest that flood depositions of the low-lying area and channel networks have an important role in driving clay variations in the study area. Higher values of clay were located closer to streams compared to areas that are far away from water bodies, a result which implies the influence of topography and the regulation of water and sediment redistribution. Temperature-derived bioclimatic variables were the most important bioclimatic variable for clay in the surface and subsurface layers, whereas in the lowest horizon, precipitation-related bio variables played an important role (BIO14 and BIO15).

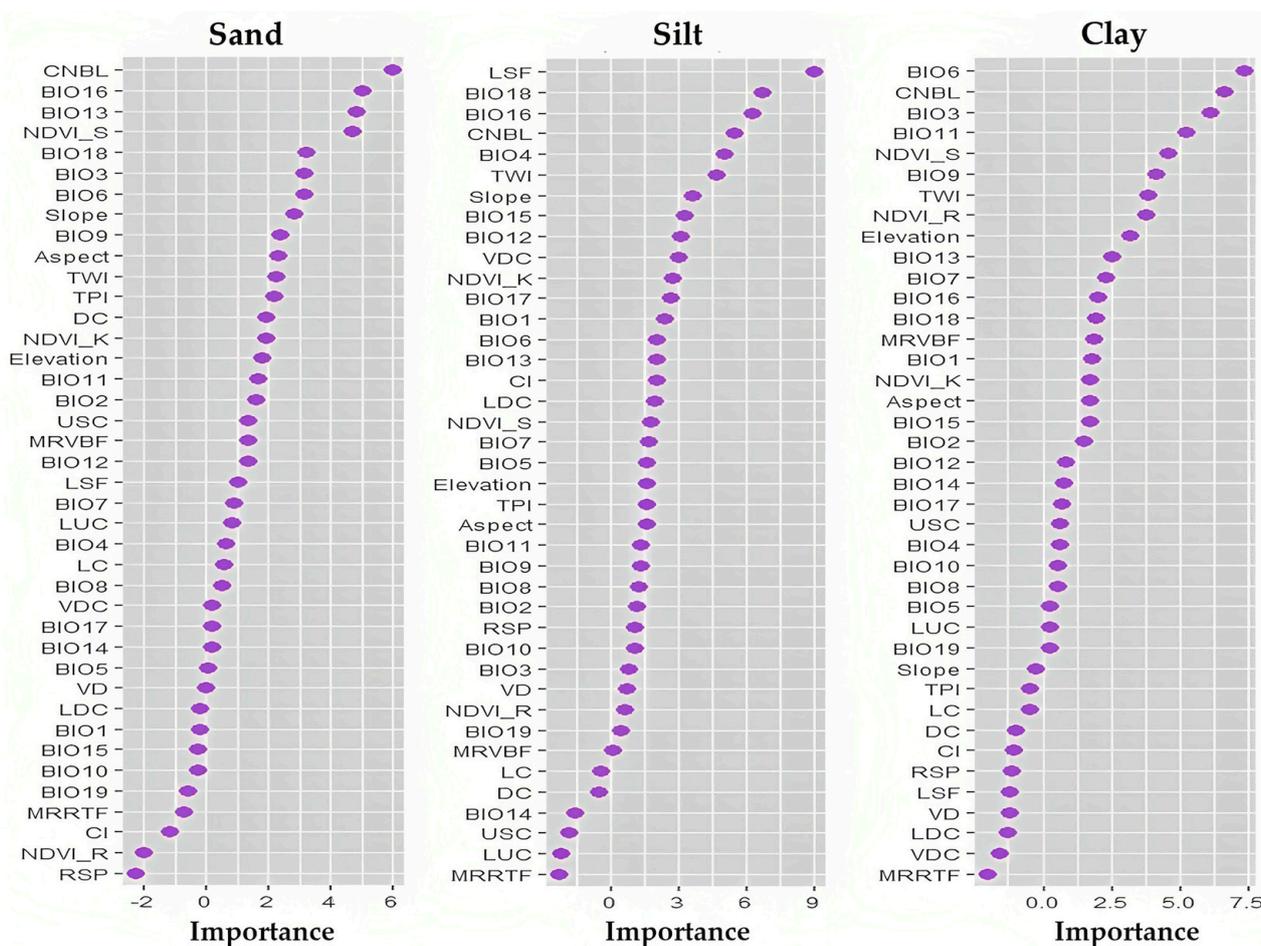
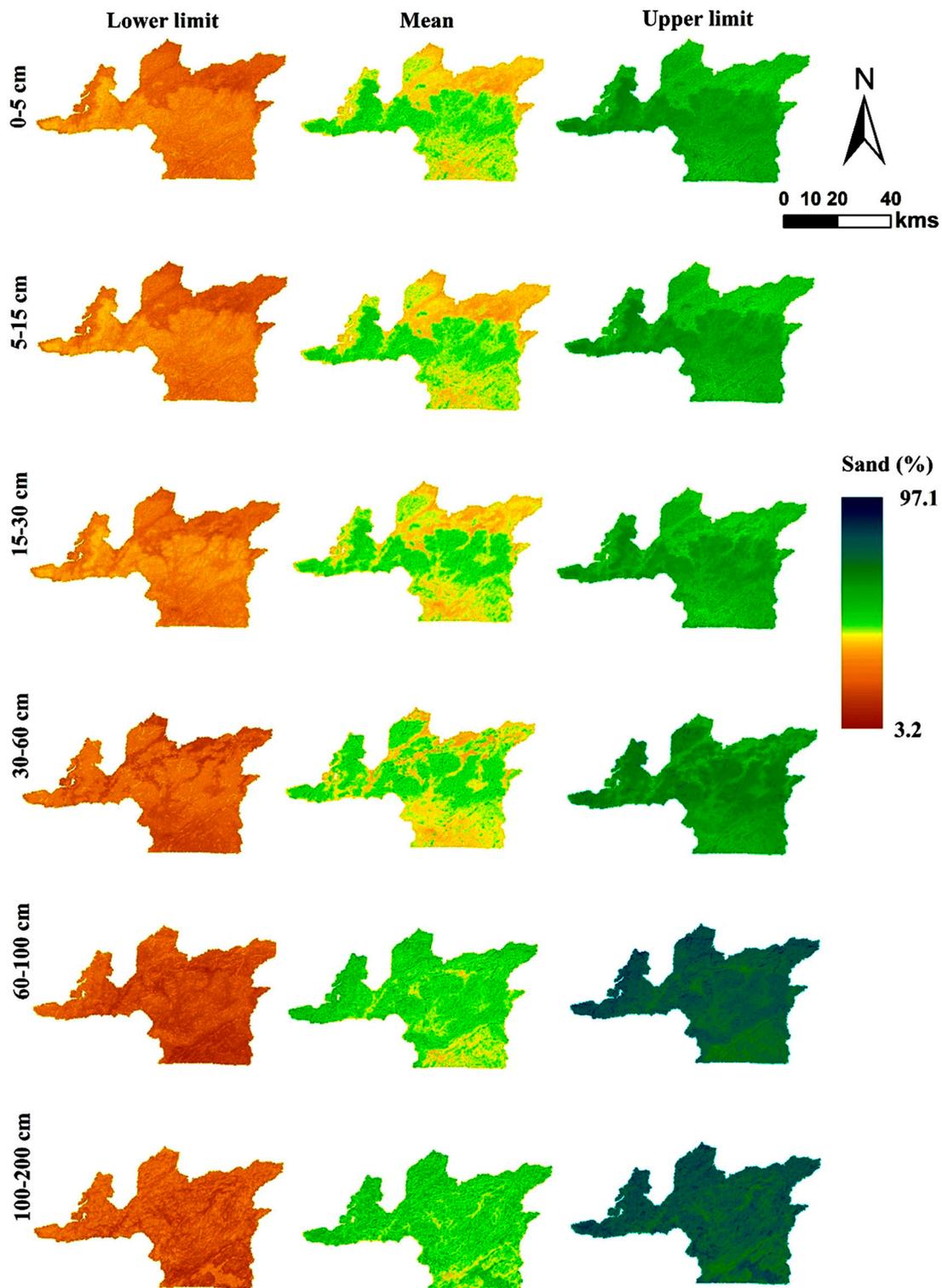


Figure 5. Importance factors in surface soil (0–5 cm) for predicting soil textural fractions.

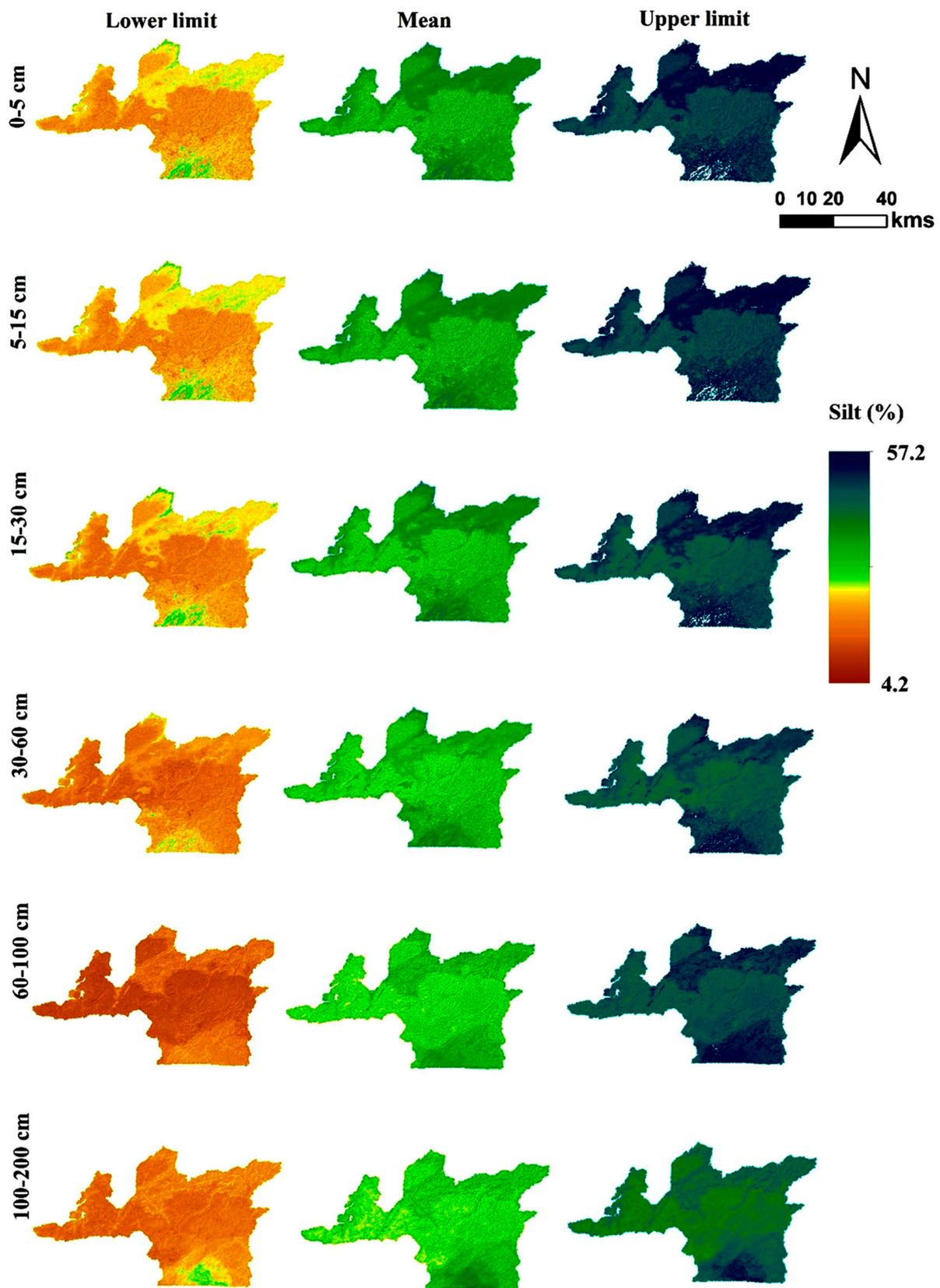
The dominance of bioclimatic variables coupled with channel networks is a sign of how terrain and climate interact at agricultural locations, resulting in hydrology and topo-sequences of soil characteristics. Under tropical rains, gravitational potential energy gradients in the landscape cause water and sediment to move and accumulate in lower-slope areas. Transport of sediments from upper slope regions to lower slope areas, which are often characterized by more fine particles, is caused by these topography-driven erosional processes [48,49]. As a result, deeper soils with a high clay percentage are found on these low slopes. The key covariates among the terrain and relief variables were CNBL, aspect, VDC, and DC across all the soil depths. In the case of precipitation-related bio variables such as annual precipitation (BIO12), precipitation of wettest month (BIO13), and precipitation of wettest quarter (mm) (BIO16) played significant roles for the prediction of sand up to 60cm of soil depth, whereas temperature related bio variables dominated in the lower horizons (annual mean temperature (°C)—BIO1, Isothermality (°C)—BIO3 and min temperature of the coldest month (°C)—BIO6). For silt fraction, CNBL, LSF, and slope played a significant role among all the terrain variables. The effects of bioclimatic variables were similar to the sand fraction. Considering vegetation as one of the important factors in the present study, it was found that NDVI\_S had a major impact on all spatial distribution of all these PSF across all depths, followed by NDVI\_R. These indicators relate to either water availability or water stress conditions, and both have an impact on the prediction of particle size fraction throughout all sampled soil depths.

### 3.4. Spatial Predictions and Their Uncertainty

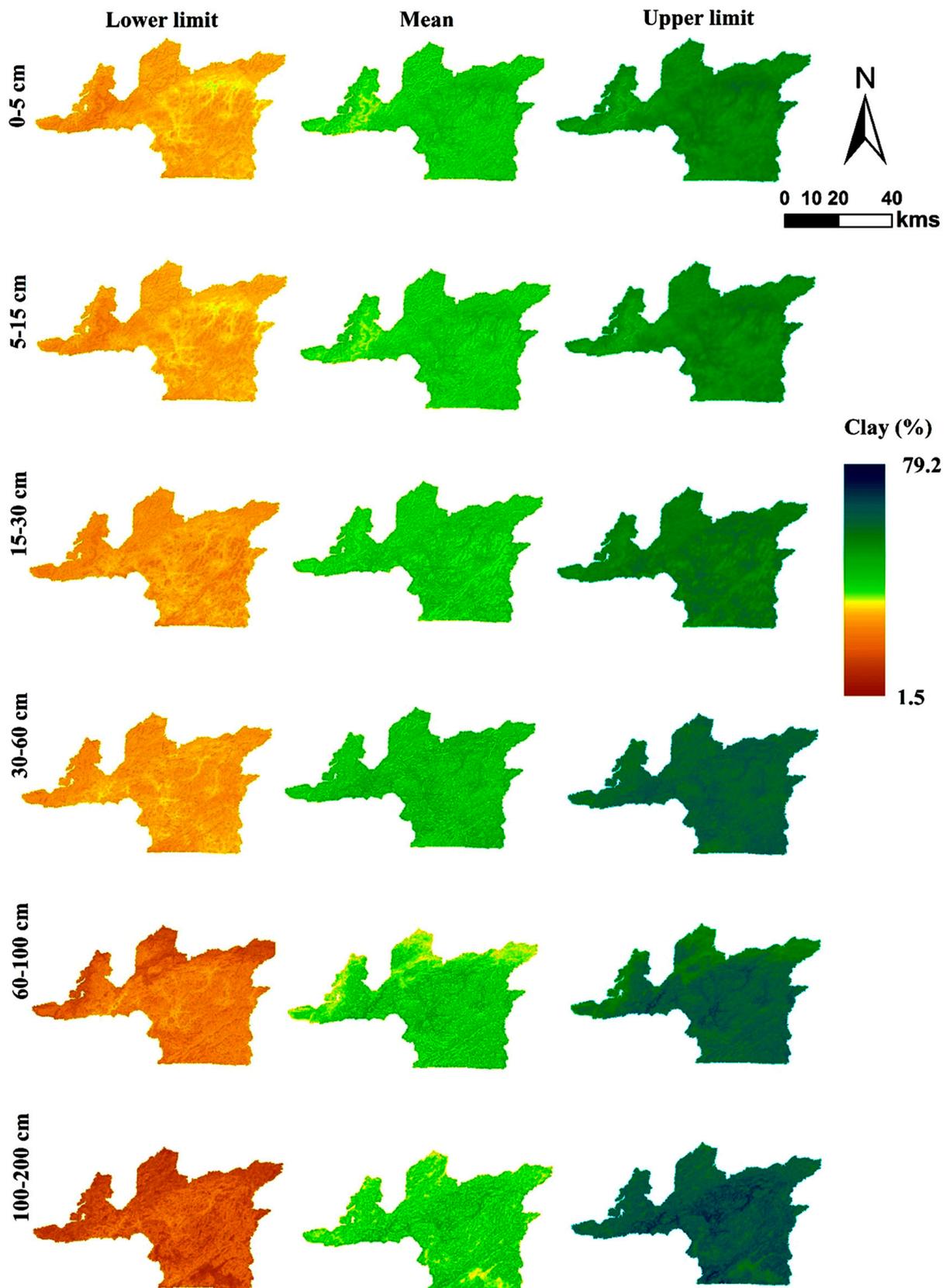
The distribution of the sand, silt, and clay fractions in space for each of the six-depth intervals is shown in Figures 6–8, along with the quantified uncertainties that are shown as upper and lower prediction limits.



**Figure 6.** The lower prediction limit (5th percentile) and higher prediction limit (95th percentile) are also displayed on maps showing the mean predicted sand content over the study area at six-depth intervals.



**Figure 7.** The lower prediction limit (5th percentile) and higher prediction limit (95th percentile) are also displayed on maps showing the mean predicted silt content over the study area at six-depth intervals.



**Figure 8.** The lower prediction limit (5th percentile) and higher prediction limit (95th percentile) are also displayed on maps showing the mean predicted clay content over the study area at six-depth intervals.

Lower and upper prediction limits with a 90% confidence interval were used to indicate the uncertainty. For all three soil texture fractions, the range between the 5% lower and 95% higher prediction limits was large, indicating that the existing spatial predictions may be improved. The textural variance in the expected values is seen in Figure 9. Most of the outcomes in the clay loam and clay classes have similar patterns. The clay loam and clay classes have higher projected values. If we look closely at the sand, silt, and clay fraction spatial patterns, we can see that the expected mean map values are restricted to the center of the legends. For the sand and clay fractions, the extremely lower and higher limits were seen in the lower horizons. The research area's center and southwest regions had the greatest sand prediction rates. This area presumably has a significant concentration of sand because, during rainy episodes, sand quickly settles out of suspension. The middle and southwestern regions of the research area were expected to have minimal silt content. The northern, northeastern, and southern regions have higher silt values. Clay composition predominates in the research area's northern region. Since this part of the study area is far from the river, likely, the slower settling time of silt and clay from runoff deposits is the cause of the higher clay concentration in these soils.

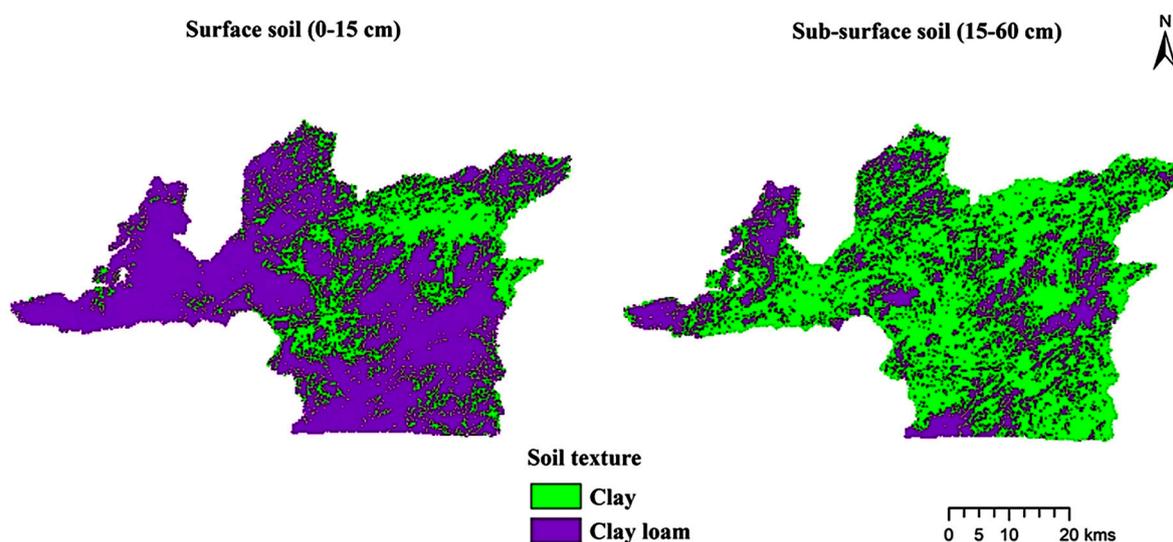


Figure 9. Predicted soil texture map of the study area.

#### 4. Conclusions

The research demonstrates that geographic variation in soil particle-size fractions may be accurately predicted both on a national scale and a detailed level using RF approaches in conjunction with currently known high-resolution soil formative environmental variables. In the Ri-Bhoi district of Meghalaya, we presented the initial edition of 30 m resolution maps of soil texture fractions. Compared to the current soil texture maps, results were significantly more precise and detailed and effectively represented the spatial variations of particle-size fractions. Additionally, there is still a need to investigate novel methodologies for extensive digital soil mapping, which will be very advantageous for many international initiatives. We observed that the topography had a significant determining factor in the spatial patterns of particle-size fractions. The physical and chemical weathering caused by the water has fueled erosion processes, which are principally responsible for shaping the clay content pattern. The results offer suggestions for creating mechanistic models of soil evolution to replicate the spatiotemporal evolution of soil texture on a regional scale. The simulation can aid in national soil management to guarantee the soil's long-term security. This is especially crucial in light of the accelerating effects of climate change and the intensification of human activity.

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

**Funding:** This research was funded by ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur (NRMA NBSSLUP SIL2014 01200362), India and Meghalaya Basin Development Authority (MBDA), Shillong, Govt. of Meghalaya, India and also partially funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2023R65), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Most of the data are available in all Tables and Figures of the manuscripts.

**Acknowledgments:** We are grateful to ICAR-National Bureau of Soil Survey and Land Use Planning, Nagpur, India; Meghalaya Basin Development Authority (MBDA), Shillong, Govt. of Meghalaya, India and Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia for supporting the current research.

**Conflicts of Interest:** Authors would hereby like to declare that there is no conflict of interest for the article.

## References

1. Arrouays, D.; Grundy, M.G.; Hartemink, A.E.; Hempel, J.W.; Heuvelink, G.B.M.; Hong, S.Y.; Lagacherie, P.; Lelyk, G.; McBratney, A.B.; McKenzie, N.J.; et al. Globalsoilmap: Toward a fine-resolution global grid of soil properties. *Adv. Agron.* **2014**, *125*, 93–134.
2. Hengl, T.; de Jesus, J.M.; MacMillan, R.A.; Batjes, N.H.; Heuvelink, G.B.; Ribeiro, E.; Samuel-Rosa, A.; Kempen, B.; Leenaars, J.G.; Walsh, M.G. SoilGrids1km—Global soil information based on automated mapping. *PLoS ONE* **2014**, *9*, e105992. [[CrossRef](#)] [[PubMed](#)]
3. McBratney, A.B.; Mendonca-Santos, L.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
4. Minasny, B.; McBratney, A.B.; Malone, B.P.; Wheeler, I. Digital mapping of soil carbon. *Adv. Agron.* **2013**, *118*, 1–47.
5. Minasny, B.; Hartemink, A.E. Predicting soil properties in the tropics. *EarthSci. Rev.* **2011**, *106*, 52–62. [[CrossRef](#)]
6. Dharumarajan, S.; Hegde, R.; Janani, N.; Singh, S.K. The need for digital soil mapping in India. *Geoderma Reg.* **2019**, *16*, e00204. [[CrossRef](#)]
7. Hengl, T.; Heuvelink, G.B.; Kempen, B.; Leenaars, J.G.; Walsh, M.G.; Shepherd, K.D.; Sila, A.; MacMillan, R.A.; de Jesus, J.M.; Tamene, L.; et al. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE* **2015**, *10*, e0125814. [[CrossRef](#)]
8. Zhao, Z.; Chow, T.L.; Rees, H.W.; Yang, Q.; Xing, Z.; Meng, F.R. Predict soil texture distributions using an artificial neural network model. *Comput. Electron. Agric.* **2009**, *65*, 36–48. [[CrossRef](#)]
9. Guo, P.-T.; Li, M.-F.; Luo, W.; Tang, Q.-F.; Liu, Z.-W.; Lin, Z.-M. Digital mapping of soil organic matter for rubber plantation at regional scale: An application of random forest plus residuals kriging approach. *Geoderma* **2015**, *237–238*, 49–59. [[CrossRef](#)]
10. Taghizadeh-Mehrjardi, R.; Nabiollahi, K.; Kerry, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* **2016**, *266*, 98–110. [[CrossRef](#)]
11. Jena, R.K.; Bandyopadhyay, S.; Pradhan, U.K.; Moharana, P.C.; Kumar, N.; Sharma, G.K.; Roy, P.D.; Ghosh, D.; Ray, P.; Padua, S.; et al. Geospatial Modelling for Delineation of Crop Management Zones Using Local Terrain Attributes and Soil Properties. *Remote Sens.* **2022**, *14*, 2101. [[CrossRef](#)]
12. Brown, D.J.; Clayton, M.K.; McSweeney, K. Potential terrain controls on soil color, texture contrast and grain-size deposition for the original catena landscape in Uganda. *Geoderma* **2004**, *122*, 51–72. [[CrossRef](#)]
13. Heuvelink, G.; Huisman, J. Choosing Between Abrupt and Gradual Spatial Variation? In *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing*; CRC Press: Boca Raton, FL, USA, 2000; pp. 111–117.
14. Adhikari, K.; Kheir, R.B.; Greve, M.B.; Bocher, P.K.; Malone, B.P.; Minasny, B.; McBratney, A.B.; Greve, M.H. High-resolution 3-D mapping of soil texture in Denmark. *Soil Sci. Soc. Am. J.* **2013**, *77*, 860–876. [[CrossRef](#)]
15. Akpa, S.I.C.; Odeh, I.O.A.; Bishop, T.F.A.; Hartemink, A.E. Digital mapping of soil particle-size fractions for Nigeria. *Soil Sci. Soc. Am. J.* **2014**, *78*, 1953–1966. [[CrossRef](#)]

16. Chagas, C.d.S.; de Carvalho Junior, W.; Bhering, S.B.; CalderanoFilho, B. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena* **2016**, *139*, 232–240. [[CrossRef](#)]
17. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; Dover Publications: New York, NY, USA, 1941; 281p.
18. Selige, T.; Boehner, J.; Schmidhalter, U. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* **2006**, *136*, 235–244. [[CrossRef](#)]
19. Lagacherie, P.; Baret, F.; Feret, J.-B.; Netto, J.M.; Robbez-Masson, J.M. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. *Remote Sens. Environ.* **2008**, *112*, 825–835. [[CrossRef](#)]
20. Ouerghemmi, W.; Gomez, C.; Naceur, S.; Lagacherie, P. Semi-blind source separation for the estimation of the clay content over semi-vegetated areas using VNIR/SWIR hyperspectral airborne data. *Remote Sens. Environ.* **2016**, *181*, 251–263. [[CrossRef](#)]
21. Ben-Dor, E. Quantitative remote sensing of soil properties. *Adv. Agron.* **2022**, *75*, 173–243.
22. Mulder, V.; de Bruin, S.; Schaepman, M.; Mayr, T. The use of remote sensing in soil and terrain mapping—A review. *Geoderma* **2011**, *62*, 1–19. [[CrossRef](#)]
23. Niang, M.; Nolin, M.; Jogo, G.; Perron, I. Digital mapping of soil texture using RADARSAT-2 polarimetric synthetic aperture radar data. *Soil Sci. Soc. Am. J.* **2014**, *78*, 673–684. [[CrossRef](#)]
24. Zeng, C.; Yang, L.; Zhu, A.X.; Rossiter, D.G.; Liu, J.; Liu, J.; Qin, C.; Wang, D. Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method. *Geoderma* **2016**, *281*, 69–82. [[CrossRef](#)]
25. Ratnayake, R.R.; Karunaratne, S.B.; Lessels, J.S.; Yogenthiran, N.; Rajapaksha, R.P.S.K.; Gnanavelrajah, N. Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka. *Geoderma Reg.* **2016**, *7*, 167–176. [[CrossRef](#)]
26. Camera, C.; ZomeniaZomeni, Z.; Jay, S.; Noller, J.S.; Andreas, M.; Zissimos, A.M.; Christoforou, I.C.; Bruggeman, A. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. *Geoderma* **2017**, *285*, 35–49. [[CrossRef](#)]
27. Vagen, T.G.; Leigh, A.; Winowiecki, L.A.; Tondoh, J.E.; Desta, L.T.; Gumbricht, T. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma* **2016**, *263*, 216–225. [[CrossRef](#)]
28. Wang, D.C.; Zhang, G.L.; Zhao, M.S.; Pan, X.Z.; Zhao, Y.G.; Li, D.C.; Macmillan, B. Retrieval and mapping of soil texture based on land surface diurnal temperature range data from MODIS. *PLoS ONE* **2015**, *10*, e0129977. [[CrossRef](#)]
29. Walder, K.; Walder, O.; Rinklebe, J.; Menz, J. Estimation of soil properties with geostatistical methods in floodplains. *Arch. Agron. Soil Sci.* **2008**, *54*, 275–295. [[CrossRef](#)]
30. Singh, R.S.; Baruah, U.; Sarkar, D.; Butte, P.S.; Gajbhiye, K.S. *Soil Series of Meghalaya*; National Bureau of Soil Survey and Land Use Planning: Nagpur, India, 2005; p. 86.
31. Singh, R.S.; Maji, A.K.; Sehgal, J.; Velayutham, M. *Soils of Meghalaya for Optimising Land Use Planning*; National Bureau of Soil Survey and Land Use Planning: Nagpur, India, 1999; p. 29.
32. Soil Survey Staff. *Keys to Soil Taxonomy*, 12th ed.; U.S.D.A.—Natural Resources Conservation Service: Washington, DC, USA, 2014.
33. Jena, R.K.; Padua, S.; Bandyopadhyay, S.; Ramachandran, S.; Ray, P.; Deb Roy, P.; Chatterjee, S.; Sah, K.D.; Baruah, U.; Singh, S.K.; et al. *Land Resource Inventory of Ri-Bhoi District, Meghalaya at 1: 10000 Scales for Optimal Agricultural Land Use Planning Using Geospatial Technique*; Report No. 1124; ICAR-National Bureau of Soil Survey and Land Use Planning: Nagpur, India, 2020; pp. 1–231.
34. Jackson, M.L. *Soil Chemical Analysis*; Prentice Hall of India Private Limited: New Delhi, India, 1973; p. 498.
35. Corral-Pazos-de-Provens, E.; Rapp-Arraras, I.; Domingo-Santos, J.M. Estimating Textural Fractions of the USDA Using Those of the International System: A Quantile Approach. *Geoderma* **2022**, *416*, 115783. [[CrossRef](#)]
36. Bishop, T.F.A.; McBratney, A.B.; Laslett, G.M. Modeling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **1999**, *91*, 27–45. [[CrossRef](#)]
37. Malone, B.P.; McBratney, A.; Minasny, B.; Laslett, G. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **2009**, *154*, 138–152. [[CrossRef](#)]
38. O'Brien, L. Mpspline2: Mass-Preserving Spline Functions for Soil Data. 2022. Available online: <https://CRAN.R-project.org/package=mpspline2> (accessed on 25 December 2022).
39. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **2005**, *25*, 1965–1978. [[CrossRef](#)]
40. Liaw, A.; Wiener, M. Classification and regression by random forest. *R News* **2002**, *2*, 18–22.
41. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019; Available online: <https://www.R-project.org/> (accessed on 31 January 2023).
42. Solomatine, D.P.; Shrestha, D.L. A novel method to estimate model uncertainty using machine learning techniques. *Water Resour. Res.* **2009**, *45*. [[CrossRef](#)]
43. 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](#)]
44. Dharumarajan, S.; Kalaiselvi, B.; Suputhra, A.; Lalitha, M.; Vasundhara, R.; Anil Kumar, K.S.; Nair, K.M.; Hegde, 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](#)]
45. Thompson, J.A.; Pena-Yewtukhiw, E.M.; Grove, J.H. Soil-landscape modelling across a physiographic region: Topographic patterns and model transportability. *Geoderma* **2006**, *133*, 57–70. [[CrossRef](#)]

46. Malone, B.P.; Searle, R. Updating the Australian digital soil texture mapping: Part Re-calibration of field soil texture class centroids. *Soil Res.* **2021**, *59*, 419–434. [[CrossRef](#)]
47. Rudiyanto; Minasny, B.; Setiawan, B.I.; Arif, C.; Saptomo, S.K.; Chadirin, Y. Digital mapping for cost-effective and accurate prediction of the depth and carbon stocks in Indonesian peatlands. *Geoderma* **2016**, *117*, 3–52. [[CrossRef](#)]
48. Pahlavan-Rad, M.R.; Akbarimoghaddam, A. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *Catena* **2018**, *160*, 275–281. [[CrossRef](#)]
49. Wubie, M.A.; Assen, M. Effects of land cover changes and slope gradient on soil quality in the Gumarawatershed, Lake Tana basin of North–West Ethiopia. *Model. Earth Syst. Environ.* **2020**, *6*, 85–97. [[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.