remote sensing

MDPI

Technical Note

Constructing Soil–Landscape Units Based on Slope Position and Land Use to Improve Soil Prediction Accuracy

Changda Zhu, Fubin Zhu, Cheng Li, Wenhao Lu, Zihan Fang, Zhaofu Li and Jianjun Pan *

College of Resources and Environmental Sciences, Nanjing Agricultural University, No. 1 Weigang, Xuanwu District, Nanjing 210095, China; 2019203040@njau.edu.cn (C.Z.); 2020203042@stu.njau.edu.cn (F.Z.); licheng2022@njau.edu.cn (C.L.); lwh8126373@163.com (W.L.); 2020103087@stu.njau.edu.cn (Z.F.); lizhaofu@njau.edu.cn (Z.L.)

* Correspondence: jpan@njau.edu.cn

Abstract: Topography is one of the dominant factors in regional soil formation and development. Soil distribution has a certain pattern from high to low in space, and this pattern has a high degree of consistency with slope position. Most of the current research on soil mapping uses landscape types generated by existing methods directly as environmental covariates, and there are few landscape classification methods specifically oriented toward soil surveys. There is rarely any research on landform classification using relative slope position (RSP) and elevation. Therefore, we designed a landform classification method based on RSP and elevation, Terrainforms (TF), and combined the landform type with land use type to construct soil–landscape units for soil type and attribute spatial prediction. In this study, two commonly used landform classification methods, Geomorphons and Landforms, were also used to compare with this design method. It was found that the constructed soil–landscape units had a high consistency with the soil spatial distribution. The landform types based on RSP and elevation obtained the second-highest prediction accuracy in both soil type and soil organic carbon (SOC), and the constructed soil–landscape types obtained the highest prediction accuracy. The results show that the landform classification method based on RSP and elevation is not easily limited by the analysis scale, and is an efficient and accurate landform classification method. The TF landform type and its constructed soil–landscape types can be used as an important environmental variable in soil prediction and sampling, which can provide some guidance and reference for landform classification and digital soil mapping.

Keywords: soil–landscape units; soil space prediction; soil type and attribute; landform classification; relative slope position

check for updates

Citation: Zhu, C.; Zhu, F.; Li, C.; Lu, W.; Fang, Z.; Li, Z.; Pan, J. Constructing Soil–Landscape Units Based on Slope Position and Land Use to Improve Soil Prediction Accuracy. *Remote Sens.* **2024**, *16*, 4090. https://doi.org/10.3390/rs16214090

Academic Editor: Dusan Gleich

Received: 24 August 2024
Revised: 26 September 2024
Accepted: 7 October 2024
Published: 1 November 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

In conventional soil mapping, soil experts sampled at specific locations and along transects combined these observations with topographic information and image features, and then proposed a conceptual soil–landscape model [1]. The soil–landscape model is the basis of the soil survey, and environmental variables are the key to the quantitative expression of the soil–landscape relationship. Effective environmental variables can reflect the relationship between soil and the environment to a large extent, and reveal the role of soil-forming factors in soil development. Climatic and topographic factors play an important role in soil formation and development, and are the most widely used environmental variables in digital soil mapping (DSM). Climate is one of the dominant factors causing zonal soil spatial differentiation, and plays an important role in large-scale soil surveys. Topography has an important impact on both regional and zonal soil distribution, and is an indispensable environmental factor for soil prediction. In addition, in a certain local area, the climate usually changes with the change in terrain.

In the field of soil–landscape relationships, unique insights and models have been developed. In 1883, Dokuchaev introduced the concept of the soil factors function by

studying calcareous black soils [2]. Milne [3,4] proposed the theory of soil catena, which relates the development of soils on slopes to hydrological and erosional processes, and found that topography and hydrology are important factors affecting soil development. Jenny [5] proposed a concept and model based on five major soil-forming factors, which laid the foundation for the development of digital soil mapping. Wood classified slopes into four distinct types: "convex slopes", "straight slopes", "finely fragmented slopes" and "triangular slopes", as well as "straight slopes", "fine slopes", and "triangular slopes" [6]. Ruhe further developed Wood's model by classifying slopes into five slope types: "crest", "shoulder", "backslope", "foot", and "foot of slope" [7]. These landform types play an important role in soil development and can be used as important environmental variables for soil prediction.

Topography is the dominant factor of regional soil development and distribution, and topographic types or elements play an important role in indicating the soil spatial distribution and variability. There are two main categories for landform classification: One is based on rules and expert knowledge [8–10], and the other is based on classification or clustering by algorithms such as machine learning [11]. Many scholars have carried out related research and developed some integrated modules or individual software [10,12,13]. At present, there are two widely used landform classification methods in soil mapping: Landforms and Geomorphons. Landforms is a landform classification method based on topographic position index (TPI). By classifying the TPI, 10 basic landform types are finally obtained. Guisan proposed the calculation method of TPI, which is the same as the mean calculation (residual analysis) method proposed by Wilson and Gallan [14,15]. Tagil and Jenness further developed this method, and divided the landform into 4 slope positions and 10 geomorphological types [9]. Geomorphons is a novel landform classification method [8], and has been widely used in soil mapping [16,17]. This method uses local ternary patterns (LTP) and the line-of-sight principle to identify specific landform forms, which greatly improves computational efficiency [8].

The landform type is based on the analysis of topographic variables or elements, and finding the suitable topographic variables or elements to characterize and classify landforms is the key to improving the classification accuracy [18]. Terrain variables can be divided into local variables and regional variables. Local variables reflect the micro-morphological characteristics of the slope, such as slope, aspect, slope shape, profile curvature, TPI, and other micro-topographic factors. Regional variables are macro-topographic factors that can reflect the overall characteristics of geomorphic units, such as topographic relief, ground roughness, relative slope position, and topographic humidity index [10].

However, the local variables are easily limited by grid resolution and analysis scale, and the variable values are not stable. Some scholars have proved that local terrain attributes such as slope, aspect, and curvature are very sensitive to grid resolution, and the second derivative (curvature) is more sensitive than the first derivative (slope and aspect) [19]. Local variables are easily limited by the neighborhood analysis window and ignore the overall characteristics of the geomorphic unit. On the other hand, the relative slope position (RSP), as a regional topographic variable proposed by Skidmore, is not limited by the scale of analysis, which can greatly reduce the uncertainty of analysis [20]. RSP is the relative position obtained by the distance from a point to the ridge line and the valley line. The ridge line and valley line are typical linear topographic features, and play an important role in revealing the regional landform features [21,22]. RSP can reflect the macro position of a point in a regional landform unit, so as to have better consistency with the elevation and soil spatial distribution.

Except for the topography factor, in agricultural soil areas greatly affected by humans, land use type also has an important impact on soil distribution, and human activities may fundamentally change the process of soil development. The key to soil survey and mapping is to construct landform types and soil–landscape units that conform to the soil spatial distribution. At present, there are a lot of technologies and products of land use classifica-

tion [23–27], but the landform type has not yet formed a unified classification system, and there is no landform classification method, especially for soil surveys [17,28–30].

Many previous studies on digital soil mapping have focused on simple applications of topography or terrain attributes, but have overlooked the spatial adaptability of topography types and soil [16,28,31]. Therefore, we designed a landform classification method based on RSP and elevation—Terrainforms (TF), and combined the landform type with land use type to construct the soil–landscape units for soil type and attribute prediction. This research may improve the accuracy and efficiency of soil prediction, and provide some reference for soil survey and landform classification.

2. Materials and Methods

2.1. Study Area and Data Sources

The study area is located in the northern part of Jurong City, Jiangsu Province of China, which is a hilly agricultural region with a total area of about 2150 ha (shown in Figure 1). The study area has a north subtropical continental monsoon climate, with an average annual temperature of 15.2 °C and an annual precipitation of 1060 mm. Forest and cultivated land are the main land use types, and paddy fields and irrigable land are the main types of cultivated land.

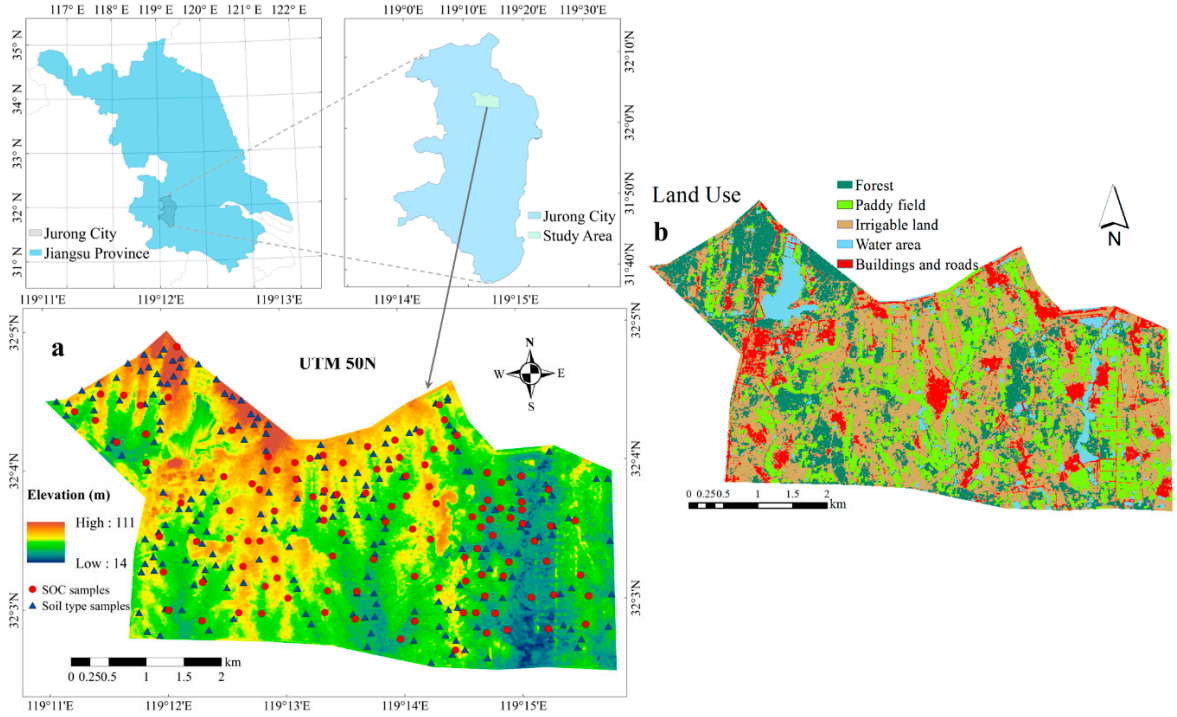


Figure 1. Location of the study area. (**a**) Elevation and soil samples, (**b**) land use type.

The DEM with 2.1 m resolution was generated based on ZY3-02 image stereo pairs with a resolution of 2.1 and 2.5 m, and after resampling to 30 m to reduce the landscape fragmentation caused by high resolution, the topographic indices and geomorphological types were generated in SAGA GIS (System for Automated Geoscientific Analysis, http://www.saga-gis.org, accessed on 15 June 2022). By using object-oriented methodology, the land use type with a resolution of 1.0 m was generated from GF2 imagery which has three visible bands and one near-infrared band [32]. Then, the results of the study were validated based on the topsoil SOC samples and soil type (soil species) samples (shown in Figure 1a). We obtained the SOC values of 115 topsoil (0~20 cm) samples and the soil types of 200 samples. SOC data are normally distributed, and the soil type represents the soil species level. The statistical analysis of SOC and soil type samples has been described in previous studies, and more descriptions can be obtained in references [32,33].

In addition, we also selected five topographic indices [34,35] and three vegetation indices [36–38] as environmental covariates, including the convergence index (CI), plan curvature (PlC), profile curvature (PrC), slope, topographic wetness index (TWI), soil adjusted vegetation index (SAVI), soil adjusted total vegetation index (SATVI), and soil red index (SRI). All of the study areas had a warm temperate continental monsoon climate. The spatial variability in temperature and precipitation was weak, and parent material has a low correlation with soil, so these factors were not selected to be used as predictor variables.

2.2. Research Methodology

2.2.1. Geomorphons Landform Classification Method

Geomorphons (GM) was a novel geomorphological classification method proposed by Jasiewicz, which uses computer vision tools rather than differential geometry tools to classify terrain elements [8]. This method makes use of adaptive techniques using line-of-sight principles, which can generate neighborhoods with different sizes and shapes to adapt to the local terrain [39]. In addition, the method utilizes the concept of local ternary patterns (LTP) to identify local terrain elements. The set of all possible LTPs is 498 different types, which are finally reclassified into the 10 most common landform types. Since no differential geometry algorithm is used, it can reduce a large amount of computation and can significantly improve computational efficiency [40].

Geomorphons has two main parameters: The search radius L of the retrieval window and the flatness threshold t, where t is regarded as the size of the slope threshold of the horizontal region. Due to the adaptive technique of the line-of-sight principle, the value of the flatness threshold t can have a great impact on the Geomorphons classification results. The Geomorphons method is invoked in SAGA to automatically classify 10 landform types based on the DEM, which are pit/depression, valley, footslope, hollow, slope, spur, shoulder, ridge, peak, and flat.

2.2.2. TPI-Based Landform Classification Method—Landforms

Landforms (LF) is a method for determining landform types by grading the topographic position index (TPI) [9,14,15]. The TPI is a measure of the distance between the elevation of a center point within a given area (z_0) and the mean elevation of the surrounding radius (r) neighborhood ($\overline{z}$), which can be expressed by the Equations:

$$\mathrm{TPI} = z_0 - \overline{z} \tag{1}$$

$$\overline{z} = \frac{1}{n_R}\sum_{i \in R} z_i \tag{2}$$

The TPI is calculated from the elevation of the neighboring area, with positive TPI values indicating that the centroid is located above the surrounding average (e.g., ridge) and negative TPI values indicating that the centroid is located below the surrounding average (e.g., valley). Areas with TPI values close to zero are either flat (slope close to zero) or have a constant slope (slope at the point is significantly greater than zero). In general, the range of TPI values increases with scale due to the tendency for elevation to be spatially autocorrelated, and the range of TPI values depends not only on elevation differences, but also on r. Larger values of r show major landscape unit, while smaller values of r highlight minor features such as secondary valleys and ridges [31].

The Landforms (LF) method in SAGA allows for the classification of landforms into 10 types: high ridges, midslope ridges, local ridges, upper slopes, open slopes, plains, valleys, upland drainages, midslope drainages, and rivers, as well as open slopes, plains, valleys, upland drainages, midslope drainages, and streams.

2.2.3. Landform Classification Method Based on RSP and Elevation

RSP is a kind of regional topographic variable that is not easily affected by the analysis scale, and elevation is the basic topographic variable of digital terrain analysis (DTA), which

has an important impact on landform classification. In this study, we designed a landform classification method based on RSP and elevation—Terrainforms (TF). The specific steps are as follows:

a. Calculate the RSP

RSP is a regional topographic attribute obtained by calculating the distance from a point on slope to ridge line and valley line, and can also be referred to as the relative position index (RPI) [20]. Firstly, the ridge line and the valley line are extracted according to the DEM, and then the Euclidean distance algorithm is used to interpolate between the ridge line and the valley line, and the Euclidean distance of the ridge line and the valley line closest to each pixel is obtained. The ratio of the Euclidean distance from the pixel to the nearest valley line, to the sum of the Euclidean distance from the pixel to the nearest ridge line and valley line is the RSP, which can be expressed as follows:

$$\mathrm{RSP} = d_0/(d_1 + d_2) \quad (3)$$

b. Landform classification-based RSP

In ArcGIS 10.8, RSP was divided into four categories according to the Jenkes natural discontinuity method, and the breakpoint values were 0.20, 0.46, and 0.73, respectively. Finally, the landforms were divided into four categories: flat-valley, foot-slope, mid-slope, and top-slope.

The breakpoint value of the classification is changeable, and the cut-off points between categories can be manually selected to optimize the classification for specific scenarios and problems [20]. The classification interval can be determined according to the relationship between the geomorphological characteristics of the field and the research object or purpose.

c. Landform classification-based RSP and elevation

In slope position classification, additional topographic indicators (e.g., elevation) may help to more accurately depict terrain and extract different types of features. Under certain conditions, the landform classification based on RSP alone cannot distinguish landform types in undulating units of different heights. That is to say, in the terrain units with different undulating heights (such as mountain hills and loess mounds), there will be different large terrain types but with the same RSP type. However, the distribution patterns of soil types or properties in hills and mounds (loess mounds) are often different. For example, organic matter decreases from top to bottom in natural hilly geomorphic units, but increases from top to bottom in mounds. Therefore, on the basis of RSP classification, it is necessary to distinguish the geomorphic units with different heights in combination with the elevation.

On the basis of RSP classification (step b), the landform was further subdivided by superimposing elevation. Through the comparison of DEM and field landscape characteristics, it is found that the highest elevation of mounds was lower than 70 m, and the hills were higher than 70 m. Most of the area was top-slope type, and the difference in soil type was mainly distributed in the top-slope area. Therefore, 70 m was determined as the boundary between the high top-slope and the low top-slope. The landform classification system and schematic diagram based on RSP and elevation are shown in Table 1 and Figure 2.

Table 1. Landform classification hierarchy based on RSP and elevation.

Class	Description	Breakpoints
1	Flat-valley	RSP < 0.20
2	Foot-slope	RSP $\geq$ 0.20, RSP < 0.46
3	Mid-slope	RSP $\geq$ 0.46, RSP < 0.73
4	Low top-slope	RSP $\geq$ 0.73, Elevation < 70 m
5	High top-slope	RSP $\geq$ 0.73, Elevation $\geq$ 70 m

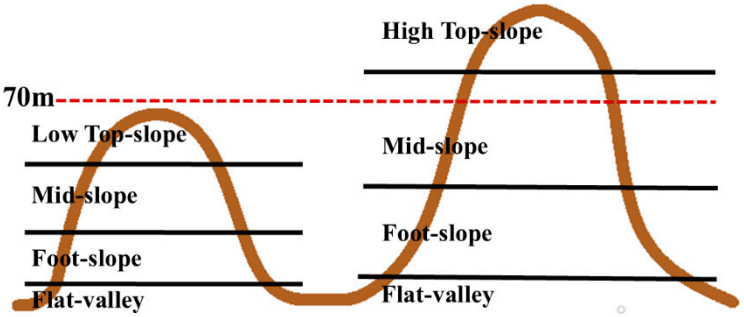


Figure 2. Schematic diagram of landform classification based on RSP and elevation.

2.2.4. Soil–Landscape Units Construction

Soil and landscape have geographical similarities, and the construction of landscape units that conform to the spatial distribution of soil type or attribute is very helpful to improve soil prediction accuracy. In this paper, we use soil–landscape units (SL) to represent the landscape units oriented to soil survey and mapping. Topography and land cover are two dominant factors for regional soil formation and development, and landform type and land use type are the key indicators for constructing soil–landscape units. Due to the small study area, climate and parent material factors have little effect on soil distribution. Soil is mainly affected by topography and biological and human factors, and land use type is intuitive and easy to obtain to quantify the biological and human factors. Therefore, the influence of environmental factors such as temperature, precipitation, and parent material on soil can be ignored, and only two environmental factors, landform type and land use type, are selected to construct soil–landscape units. The flowchart for constructing soil–landscape units is shown in Figure 3.

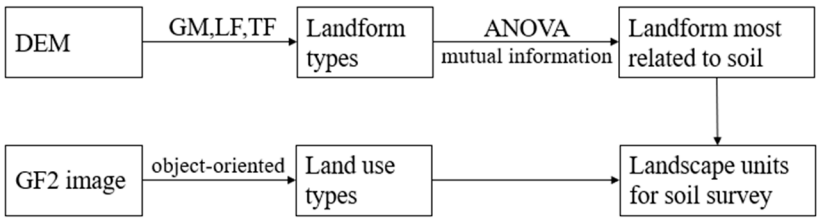


Figure 3. Flowchart for constructing landscape units for soil survey.

2.2.5. Modeling and Validation Methods

In this section, we first introduced random forest (RF) as the data mining method for modeling. RF is an ensemble model combined with bagging algorithm. The decision tree is used as the base classifier, and the Bootstrap method is used for sampling with putback. Multiple base learners are trained to effectively avoid the problem of overfitting of a single model [41].

In this study, three evaluation metrics were calculated to analyze the model performance of SOC: *RMSE*, R^2, and *MAE* [42,43]. The R^2 was used to verify the stability of the model, and the larger the R^2, the more stable the model. These validation metrics are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}, \quad (4)$$

$$R^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y_i})^2 / \sum_{i=1}^{n}(y_i - \overline{y_i})^2, \quad (5)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|, \quad (6)$$

where n represents the number of sample points, and y_i and $\hat{y}_i$ represent the observed and predicted values of SOC content of site i, respectively.

Two metrics were calculated to evaluate the model performance of soil type: *overall accuracy* and *Kappa index*. The *overall accuracy* is the sum of the main diagonal components of the confusion matrix divided by the total number of samples. The *Kappa index* is a consistency measure that combines the total number of samples, the number of soil types, and the correctly classified samples [44].

$$Overall\ Accuracy = \frac{\sum_{j=1}^{k} x_{jj}}{N} \tag{7}$$

$$Kappa = \frac{N\sum_{j=1}^{k} x_{jj} - \sum_{j=1}^{k} x_{ij}x_{ji}}{N^2 - \sum_{j=1}^{k} x_{ij}x_{ji}} \tag{8}$$

where k is the number of classifications, N is the overall sample size, x_{jj} is the number of correctly classified samples, and x_{ij} and x_{ji} are the numbers of misclassified samples in row i and column j, respectively.

3. Results and Analysis

3.1. Landform Classification Based on Different Methods

The Geomorphons (GM) method were more affected by the flatness threshold (t) and less affected by the search radius (L). In SAGA v.8.1.3, the parameter L was set to the default value of 10,000 to eliminate the influence of search radius, and the parameter t was set to 1°, 3°, and 5° to perform landform classification, respectively. The Landforms (LF) method was mainly affected by the search radius (r), and the parameter r was set to 100, 300, and 500 pixels to perform landform classification, respectively. Then, according to the principle of near spatial distance and similar geomorphological characteristics of landform types, GM and LF were reclassified into four landform types (shown in Figure 4).

According to Figure 4, regarding to the GM method, when t = 1°, the area of foot-slope is very small; when t = 5°, the landform is mainly foot-slope; and when t = 3°, the distribution of the four landform types is more in line with the actual landscape characteristics. This is due to the fact that the study area is mainly undulating landforms, and when the flatness threshold is small, the landform classification is more sensitive to slope, which makes the area of flat-valley areas smaller. Meanwhile, when the flatness threshold is large, the landform classification results are not sensitive to slope, which makes the area of flat-valley areas larger. Overall, either too large or too small a value of t resulted in classification results that differed significantly from the field landscape characteristics. Landform classification results with t = 3° best matched the field landscape characteristics.

According to the LF method, it can be obtained that the differences in the landform classification results for different r are mainly in the two landform types of top-slope and mid-slope, while the changes in the foot-slope and the flat-valley areas are smaller. As the value of r increases, the area at the top-slope gradually increases and the area in the mid-slope gradually decreases. However, the area of the foot-slope is larger in the classification result, which is different from the actual landscape. In addition, there is a large portion of paddy fields misclassified as lowland, which makes it easy to confuse paddy fields with rivers and lakes. This is because TPI is a neighborhood attribute. When the search radius is smaller than the width of rivers and lakes, the TPI value is 0, which will be classified as foot-slope; while when the search radius is larger than the width of paddy fields, the TPI value is not equal to 0, which will be classified as another non-foot-slope type.

The RSP was generated in SAGA GIS, and the landform classification was carried out using the RSP and elevation according to the method in Section 2.2.3. Then, we generated the classification maps of the four landform types (TFs4) based on the RSP and the five landform types (TF) based on the RSP and elevation (shown in Figure 5). It can be seen that the four RSP-based landform types have clear boundaries with obvious spatial distribution patterns. The addition of elevation to the landform types produces two top-slope types

in hills and mounds with higher elevation, and this subdivision helps to distinguish the spatial distribution patterns of soil type or properties.

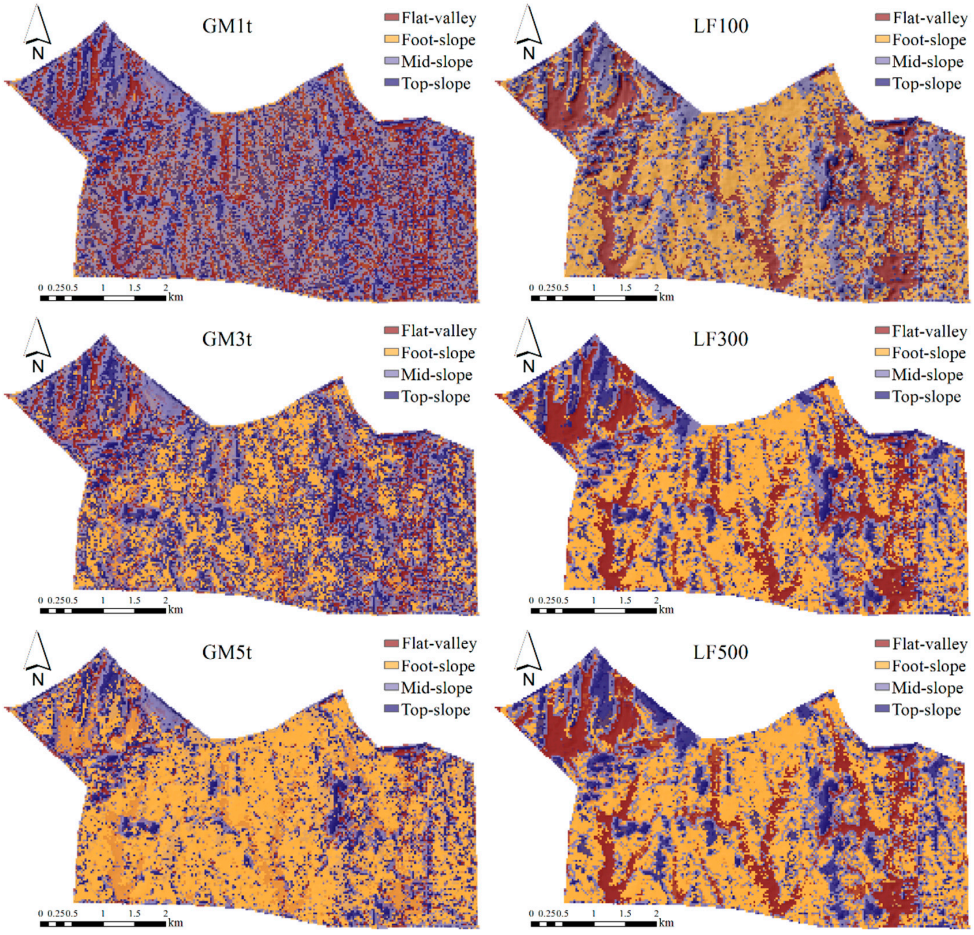


Figure 4. Landform classification maps based on different parameters of GM and LF methods. Notes: GM1t, GM3t, GM5t stand for t = 1°, 3°, 5°, respectively; LF100, LF300, LF500 stand for r = 100, 300, 500, respectively.

The GM method is prone to misclassification of rivers and lakes, which is due to the fact that GM is based on the visual line principle. As a result, GM is more dependent on slope, but ignores the important influence of elevation or relative undulation. The LF method has a more obvious landform spatial distribution pattern, but is prone to misclassification in the mounds area, especially when distinguishing between foot-slope and flat-valley types. Since LF is based on the neighborhood attribute—TPI for classification, different lengths of search radius may lead to completely different classification results. Although it is possible to integrate the results of multi-scale analysis to form nested terrain, it is difficult to perform quantitative analysis and is not conducive to use by non-professionals. On the other hand, RSP is a regional topographic attribute derived from ridgelines and valley lines, which is not affected by the analysis scale, making it easy to analyze and compare the results of landform classifications in different regions. The TF landforms based on RSP and elevation can ensure the integrity of landform distribution in the same undulating landform unit, and can subdivide the landform types with different undulating degrees. The TF landforms are more in line with the characteristics of the field landscape and are consistent with the spatial distribution of the soil.

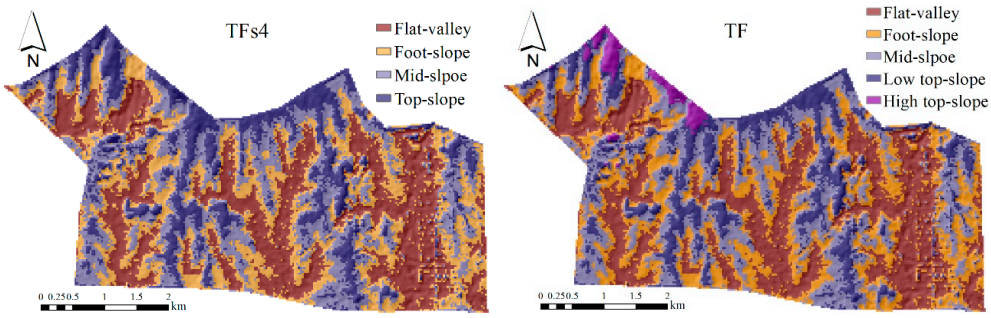
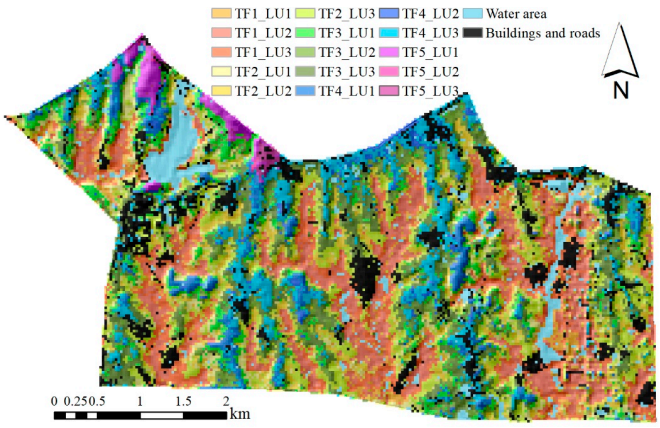


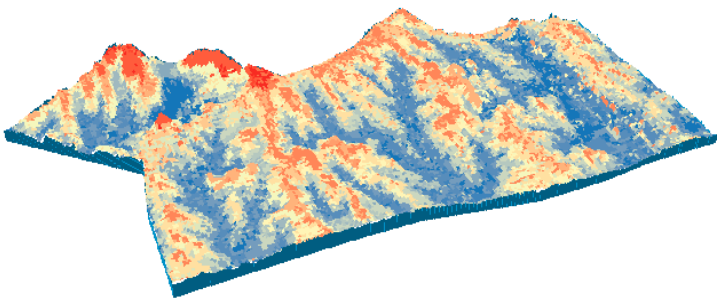
Figure 5. Landform classification maps based on RSP and elevation. Notes: TFs4 and TF represent landform types based on RSP, RSP, and elevation, respectively.

3.2. Soil–Landscape Units Distribution Maps

The soil–landscape units (SL) were constructed using the landform types (TF) and land use types (LU) according to the method mentioned above. By comparing the spatial distribution characteristics of soil and landform types, we selected TF as the landform type, which has the highest consistency with the soil spatial distribution, and then the TF and LU types were combined to construct soil–landscape units (shown in Figure 6). Based on the distribution map of soil–landscape, the soil–landscape units that combine TF and LU have a high correlation with the elevation, and have different landscape distribution characteristics in hills and mounds. The soil–landscape units have a good spatial correlation with soil distribution because soil types and properties also have different spatial distribution patterns in hills and mounds. We can find that the spatial distribution of soil–landscape units is consistent with the landscape characteristics in the field and has a high degree of consistency with the soil spatial distribution.



(**a**) rendering map



(**b**) 3D map

Figure 6. The spatial distribution maps of soil–landscape units. (**a**) Rendering map, (**b**) 3D map. Notes: TF1~5 represent the flat-valley, foot-slope, mid-slope, low top-slope, and high top-slope, respectively; and LU1~3 represent forest, paddy field, and irrigated land, respectively.

3.3. Correlation Analysis of Different Landform Classification Methods with Soil

The topsoil SOC samples and soil type samples were selected to validate the results of landform classification and soil–landscape classification based on ANOVA and mutual information values. According to the mutual information between the three landform types and soil type (shown in Table 2), it can be obtained that the mutual information between GM and soil type is lower than 0.15, the mutual information between LF and soil type is around 0.20, and the mutual information between TFs4 and soil type is 0.27. After adding elevation to the landform classification, the mutual information between TF and soil type reaches 0.35, which indicates that the RSP and elevation-based landform classification are more in line with the distribution pattern of soil type. What is more, after we constructed the soil–landscape by TF and LU, the mutual information value between SL and soil type reached 0.82, so the soil–landscape type combined with landform type and LU has a better correlation with soil type.

Table 2. The mutual information values between soil type and landform/landscape types.

Landform/Landscape	Mutual Information
GM1t	0.083
GM3t	0.142
GM5t	0.119
LF100	0.196
LF300	0.204
LF500	0.197
TFs4	0.273
TF	0.347
LU	0.683
SL	0.822

Notes: GM1t, GM3t, GM5t stand for t = 1°, 3°, 5°, respectively; LF100, LF300, and LF500 stand for r = 100, 300, and 500, respectively. TFs4, landform type based on RSP; TF, landform type based on RSP and elevation; LU, land use type; SL, soil–landscape type.

Based on the variance analysis (ANOVA) of the three landform types with SOC (shown in Table 3), we can find that the correlation between GM and SOC is the lowest, and only when t = 1°, does it have a weak correlation. The correlation between LF and SOC is high, and there is a significant correlation when r = 100 and r = 300, and a weak correlation when r = 500. TFs4 has a weak correlation with SOC, because according to RSP, the same landform types will be obtained in hills and mounds, while the distribution of SOC in hills and mounds is different. Meanwhile, TF and SOC have a completely significant correlation. The addition of elevation can distinguish hilly and mounds. The landform type is more in line with the SOC spatial distribution, thus significantly improving the correlation between landform type and SOC. On the other hand, the correlation between SL and SOC is higher than that of TF and lower than that of LU. This is because SOC is more affected by human activities (e.g., LU) in agricultural areas and has a relatively weak correlation with landforms. Soil–landscape type can integrate multiple key factors that play a leading role in soil distribution, such as landform and LU, and soil–landscape is still an effective indicator for soil prediction or sampling.

Table 3. The variance analysis of SOC and landform/landscape types.

Landform/Landscape	Analysis of Variance	
	F	***p***
GM1t	3.05	0.0512
GM3t	1.15	0.332
GM5t	1.91	0.132
LF100	7.71	0.0002

Table 3. *Cont.*

Landform/Landscape	Analysis of Variance	
	F	***p***
LF300	6.84	0.0003
LF500	3.26	0.0242
TFs4	3.03	0.0326
TF	6.84	0.00006
LU	20.61	2.40×10^{-8}
SL	6.05	6.79×10^{-8}

Notes: GM1t, GM3t, GM5t stand for t = 1°, 3°, 5°, respectively; LF100, LF300, and LF500 stand for r = 100, 300, and 500, respectively. TFs4, landform type based on RSP; TF, landform type based on RSP and elevation; LU, land use type; SL, soil–landscape type.

3.4. SOC and Soil Type Modeling Analysis

Based on SOC and soil type samples, we used the RF model to predict soil type and SOC to verify the results of soil–landscape classification and three landform classification methods. For the GM and LF methods, we selected the classification results of the parameters with the highest correlation with soil type/SOC. As a result, the landform types corresponding to soil type prediction are GM3t and LF300, and the landform types corresponding to SOC are GM1t and LF100, respectively. Then, feature selection is performed on the predictive variables. In order to reduce the impact of the original variables on landform and soil–landscape types, we removed variables such as elevation, RSP, and LU that participate in landform/landscape classification. At the same time, the vegetation indexes with high correlation with TF and SL were removed to reduce the influence of variable collinearity on the prediction results.

The RF model was invoked in R to model and validate the soil type and SOC according to different landforms or soil–landscape types, respectively, and the results are shown in Tables 4 and 5. We can find that the prediction accuracy of GM and LF in both the calibration and validation sets of soil type is relatively high, which is related to the fact that soil type is strongly influenced by geomorphology. However, the prediction accuracy of GM and LF in the SOC validation set is lower, especially when the R^2 of GM is close to 0. This is because the landform spatial distribution of GM and LF does not correspond to the SOC spatial distribution, which results in a low correlation between landform type and SOC.

Table 4. The prediction accuracy of soil type in different landform/landscape types.

Landform/Landscape	Calibration Set		Validation Set	
	Accuracy	***Kappa***	***Accuracy***	***Kappa***
TFs4	0.66	0.45	0.62	0.42
TF	***0.68***	***0.49***	***0.66***	***0.48***
GM3t	0.66	0.46	0.56	0.33
LF300	0.66	0.46	***0.66***	***0.48***
SL	**0.78**	**0.66**	**0.78**	**0.66**

Notes: The model with the highest accuracy is marked in bold, and the model with the second highest accuracy is marked in bold italic. GM3t stands for t = 3°; LF300 stands for r = 300. TFs4, landform type based on RSP; TF, landform type based on RSP and elevation; SL, soil–landscape type.

On the other hand, SL has the highest prediction accuracy in both soil type and SOC models, and TF has the second-highest prediction accuracy. It shows that SL and TF can be used as important input variables for soil type and attribute prediction, especially soil–landscape type composed of landform and LU can greatly improve soil prediction accuracy. In terms of soil type prediction, SL and TF have achieved high accuracy in both calibration and validation sets. For SOC prediction, the R^2 of SL in calibration and the validation sets is high, but the R^2 of TF in the validation set is lower than that in the calibration set. We can conclude that SL has good stability in soil type and attribute (e.g., SOC) prediction,

and TF has good stability in soil type prediction, but has relatively low stability in soil attribute prediction.

Table 5. The prediction accuracy of SOC in different landform/landscape types.

Landform/Landscape	Calibration Set			Validation Set		
	RMSE	***R^2***	***MAE***	***RMSE***	***R^2***	***MAE***
TFs4	2.72	0.47	2.20	3.23	0.01	2.55
TF	***2.62***	***0.49***	***2.16***	***2.81***	0.23	***2.33***
GM1t	2.94	0.41	2.34	3.20	0.01	2.59
LF100	2.93	0.36	2.32	2.95	***0.24***	2.45
SL	**2.50**	**0.49**	**2.02**	**2.30**	**0.50**	**1.99**

Notes: The model with the highest accuracy is marked in bold, and the model with the second highest accuracy is marked in bold italic. GM1t stands for $t = 1°$; LF100 stands for $r = 100$. TFs4, landform type based on RSP; TF, landform type based on RSP and elevation; SL, soil–landscape type.

4. Discussion

Based on the comparison of the three landform classification methods, it can be concluded that the designed TF method has the highest prediction accuracy and correlation with soil and topographic attributes, followed by LF, and GM has the lowest correlation and prediction accuracy. As GM is limited by the flatness threshold (t) [8], it leads to the fact that slope can have a great influence on its classification. Although slope has an important influence on geomorphology, elevation is often the decisive and fundamental topographic attribute of geomorphological types, and slope is mainly used for micro-geomorphological subdivision within a single geomorphological unit. LF performs better in landform classification, but it is also affected by the scale of analysis, which makes its classification accuracy not stable enough. Since TPI is a local topographic attribute based on elevation within a neighborhood [9], it causes LF to rely too much on local small-scale elevation and neglects the relative position of the whole geomorphic units. In addition, the LF method requires selecting the optimal analysis scale based on professional knowledge, which limits its application in cross-disciplinary fields [17]. The choice of analysis scale can also lead to unstable classification results, limiting the applicability of LF in other regions.

In real field landscapes, a specific area often has different landform types or different undulation conditions (e.g., hills and mounds exist simultaneously), and the distribution range of different landform units is not consistent [13]. Therefore, a fixed analysis scale will affect the classification results. As GM and LF methods are constrained by slope and analysis scale. Their classification accuracy is not high in complex terrain regions. However, the TF method integrates the regional geomorphological attribute RSP and the basic geomorphological attribute elevation, which can well reflect the overall geomorphological features and micro-geomorphological features (relative slope) of topographically complex regions.

Based on the analysis of correlation and modeling, it can be concluded that TF and SL largely reflect the relationship corresponding to landforms and soils. TF and SL have a high correlation with vegetation indices, which reflects that TF and SL are in good agreement with the field landscape features. In addition, the accuracy of adding the geomorphology/landscape type to the SOC prediction model is lower than that of adding vegetation indices, while the accuracy of adding geomorphology/landscape type to the soil type prediction model is higher than that of adding vegetation indices. This is due to the fact that SOC is highly influenced by anthropogenic use patterns, and thus SOC is more highly correlated with land use cover than landforms [33]. On the other hand, soil type is mainly determined by the core soil layer, and soil types other than paddy fields in the study area are less affected by human utilization. Moreover, the topography is the dominant factor in the regional soil type formation [45,46], and the spatial distribution of paddy fields and non-paddy fields have a high correlation with topography, so the addition of geomorphology/landscape type can greatly improve the accuracy of soil type prediction.

Due to the weak correlation between soil with climate, parent material, and other soil-forming factors in the study area, we utilized only two landscape factors, landform, and land use type, to construct the soil–landscape units. When constructing soil–landscape units in other areas, climate factors also need to be taken into account on a large scale, and the influence of parent material factors also needs to be considered in mountain regions where soil-parent material has a high correlation with soil [45]. When constructing soil–landscape units based on a larger number of landscape factors, the number of types increases exponentially. If we use soil–landscape types as a predictor variable, each soil–landscape type requires at least one or more soil samples, and the number of field soil samples will greatly increase accordingly. Thus, it is not appropriate to use soil–landscape types directly as a predictor variable, but the soil–landscape types can be used as a variable after reclassification or clustering. In addition, we can also extract the typical samples based on soil–landscape types without being used as a predictor variable for modeling, which can make full use of the correlation between soil–landscape and soil.

5. Conclusions

In this study, we designed a landform classification method based on RSP and elevation (Terrainforms, TF), and constructed the soil–landscape units by landform land use type for the soil survey and mapping. The constructed landform and soil–landscape types were well consistent with the soil spatial distribution. Two commonly used landform classification methods (Geomorphons and Landforms) were also used to compare with this experimental method, and the results of landform classification and soil–landscape classification were applied to the spatial prediction of soil type and attribute (SOC). It was found that the landform types based on RSP and elevation had a higher correlation with topography, vegetation index, and soil than that of the other two methods; the highest and second highest prediction accuracies were obtained in both soil type and SOC prediction models based on soil–landscape and TF landform types.

RSP is a regional topographic attribute derived based on ridgelines and valley lines, which can better reflect the overall characteristics of geomorphological units and the morphological features of microtopography than slope and topographic position indices. Elevation is the most basic topographic attribute in digital terrain analysis. The results showed that the landform classification based on RSP and elevation is an efficient and accurate landform classification method, and the soil–landscape units constructed by combining geomorphological type and land use type can be used as an important environmental variable in soil prediction and sampling, which can provide some guidance and reference for landform classification and digital soil mapping.

Author Contributions: Data curation, C.Z. and F.Z.; funding acquisition, J.P.; methodology, C.Z.; software, C.Z.; writing—original draft preparation, C.Z.; writing—review and editing, C.Z. and J.P.; supervision, C.L., Z.L. and J.P.; validation, C.Z., W.L. and Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This project has been funded by the National Natural Science Foundation of China (41971057), and Jiangsu Postgraduate Research and Innovation Program (KYCX21_0615).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in article.

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

References

1. Staff, S.S.D. *Soil Survey Manual*; United States Department of Agriculture: Washington, DC, USA, 1993.
2. ДОКУЧАЕВ, В.В. *Русский чернозем*; Императорское Вольное экон. о-во: С. Петербургъ, Russia, 1883; pp. 1–640.
3. Milne, G. Some suggested units of classification and mapping, particularly for East African soils. *Soil Res.* **1935**, *4*, 183–198.
4. Milne, G. Normal erosion as a factor in soil profile development. *Nature* **1936**, *138*, 548–549. [CrossRef]

5. Jenny, H. Factors of soil formation: A system of quantitative pedology/Hans Jenny. *Soil Sci.* **1941**, *52*, 415. [CrossRef]
6. Wood, A. The development of hillside slopes. *Geol. Assoc. Proc.* **1942**, *33*, 128–138. [CrossRef]
7. Ruhe, R.V. Transactions of the 7th International Congress of Soil Science. *Elem. Soil Landsc.* **1960**, 165–169.
8. Jasiewicz, J.; Stepinski, T.F. Geomorphons—A pattern recognition approach to classification and mapping of landforms. *Geomorphology* **2013**, *182*, 147–156. [CrossRef]
9. Tagil, S.; Jenness, J. GIS-based automated landform classification and topographic, landcover and geologic attributes of landforms around the Yazoren Polje, Turkey. *J. Appl. Sci.* **2008**, *8*, 910–921. [CrossRef]
10. Hengl, T.; Maathuis, B.H.P.; Wang, L. Chapter 13 Geomorphometry in ILWIS. In *Geomorphometry-Concepts, Software, Applications*; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands, 2009; pp. 309–331.
11. Drăguţ, L.; Blaschke, T. Automated classification of landform elements using object-based image analysis. *Geomorphology* **2006**, *81*, 330–344. [CrossRef]
12. Hofierka, J.; Mitášová, H.; Neteler, M. Chapter 17 Geomorphometry in GRASS GIS. In *Geomorphometry-Concepts, Software, Applications*; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands, 2009; pp. 387–410.
13. Olaya, V.; Conrad, O. Chapter 12 Geomorphometry in SAGA. In *Geomorphometry-Concepts, Software, Applications*; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands, 2009; pp. 293–308.
14. Guisan, A.; Weiss, S.B.; Weiss, A.D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* **1999**, *143*, 107–122. [CrossRef]
15. Wilson, J.P.; Gallant, J.C. *Topographic Analysis: Principles and Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2000.
16. Libohova, Z.; Winzeler, H.E.; Lee, B.; Schoeneberger, P.J.; Datta, J.; Owens, P.R. Geomorphons: Landform and property predictions in a glacial moraine in Indiana landscapes. *Catena* **2016**, *142*, 66–76. [CrossRef]
17. Ngunjiri, M.W.; Libohova, Z.; Owens, P.R.; Schulze, D.G. Landform pattern recognition and classification for predicting soil types of the Uasin Gishu Plateau, Kenya. *Catena* **2020**, *188*, 104390. [CrossRef]
18. Cheng, W.; Zhou, C.; Chai, H.; Zhao, S.; Liu, H.; Zhou, Z. Research and compilation of the geomorphologic atlas of the People's Republic of China (1:1,000,000). *J. Geogr. Sci.* **2011**, *21*, 89–100. [CrossRef]
19. MacMillan, R.A.; Shary, P.A. Chapter 9 Landforms and Landform Elements in Geomorphometry. In *Geomorphometry-Concepts, Software, Applications*; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands, 2009; pp. 227–254.
20. Skidmore, A.K. Terrain position as mapped from a gridded digital elevation model. *Int. J. Geogr. Inf. Syst.* **2007**, *4*, 33–49. [CrossRef]
21. Atkinson, J.; de Clercq, W.; Rozanov, A. Multi-resolution soil-landscape characterisation in KwaZulu Natal: Using geomorphons to classify local soilscapes for improved digital geomorphological modelling. *Geoderma Reg.* **2020**, *22*, e00291. [CrossRef]
22. Veronesi, F.; Hurni, L. Random Forest with semantic tie points for classifying landforms and creating rigorous shaded relief representations. *Geomorphology* **2014**, *224*, 152–160. [CrossRef]
23. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [CrossRef]
24. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [CrossRef]
25. Seydi, S.T.; Ahmadi, S.A.; Ghorbanian, A.; Amani, M. Land Cover Mapping in a Mangrove Ecosystem Using Hybrid Selective Kernel-Based Convolutional Neural Networks and Multi-Temporal Sentinel-2 Imagery. *Remote Sens.* **2024**, *16*, 2849. [CrossRef]
26. Zhu, Q.; Lei, Y.; Sun, X.; Guan, Q.; Zhong, Y.; Zhang, L.; Li, D. Knowledge-guided land pattern depiction for urban land use mapping: A case study of Chinese cities. *Remote Sens. Environ.* **2022**, *272*, 112916. [CrossRef]
27. He, H.; Liang, T.; Hu, D.; Yu, X. Remote sensing clustering analysis based on object-based interval modeling. *Comput. Geosci.* **2016**, *94*, 131–139. [CrossRef]
28. Gruber, F.E.; Baruck, J.; Geitner, C. Algorithms vs. surveyors: A comparison of automated landform delineations and surveyed topographic positions from soil mapping in an Alpine environment. *Geoderma* **2017**, *308*, 9–25. [CrossRef]
29. Lindsay, J.B.; Cockburn, J.M.H.; Russell, H.A.J. An integral image approach to performing multi-scale topographic position analysis. *Geomorphology* **2015**, *245*, 51–61. [CrossRef]
30. 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]
31. De Reu, J.; Bourgeois, J.; Bats, M.; Zwertvaegher, A.; Gelorini, V.; De Smedt, P.; Chu, W.; Antrop, M.; De Maeyer, P.; Finke, P.; et al. Application of the topographic position index to heterogeneous landscapes. *Geomorphology* **2013**, *186*, 39–49. [CrossRef]
32. Zhu, C.; Zhu, F.; Li, C.; Yan, Y.; Lu, W.; Fang, Z.; Li, Z.; Pan, J. Extracting Typical Samples Based on Image Environmental Factors to Obtain an Accurate and High-Resolution Soil Type Map. *Remote Sens.* **2024**, *16*, 1128. [CrossRef]
33. Zhu, C.; Wei, Y.; Zhu, F.; Lu, W.; Fang, Z.; Li, Z.; Pan, J. Digital Mapping of Soil Organic Carbon Based on Machine Learning and Regression Kriging. *Sensors* **2022**, *22*, 8997. [CrossRef]
34. 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 Tillage Res.* **2019**, *194*, 104289. [CrossRef]

35. Long, J.; Li, J.; Huang, Q.; Qiu, L.X.; Lu, L.M.; Bian, A.A.; Zhu, L.X.; Li, H.H.; Qian, X.J.; Xing, S.H.; et al. Effects of raster resolution on quantifying farmland soil organic carbon stock in various landforms of a complex topography, China. *Geoderma Reg.* **2023**, *34*, e00668. [CrossRef]
36. Flynn, T.; de Clercq, W.; Rozanov, A.; Clarke, C. High-resolution digital soil mapping of multiple soil properties: An alternative to the traditional field survey? *S. Afr. J. Plant Soil* **2019**, *36*, 237–247. [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. 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] [PubMed]
39. Gawrysiak, L.; Kociuba, W. Application of geomorphons for analysing changes in the morphology of a proglacial valley (case study: The Scott River, SW Svalbard). *Geomorphology* **2020**, *371*, 107449. [CrossRef]
40. Flynn, T.; Rozanov, A.; Ellis, F.; de Clercq, W.; Clarke, C. Farm-scale soil patterns derived from automated terrain classification. *Catena* **2020**, *185*, 104311. [CrossRef]
41. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]
42. Andrade, R.; Silva, S.H.G.; Faria, W.M.; Poggere, G.C.; Barbosa, J.Z.; Guilherme, L.R.G.; Curi, N. Proximal sensing applied to soil texture prediction and mapping in Brazil. *Geoderma Reg.* **2020**, *23*, e00321. [CrossRef]
43. Wu, C.; Zheng, Y.; Yang, H.; Yang, Y.; Wu, Z. Effects of different particle sizes on the spectral prediction of soil organic matter. *Catena* **2021**, *196*, 104933. [CrossRef]
44. Machado, D.F.T.; de Menezes, M.D.; Silva, S.H.G.; Curi, N. Transferability, accuracy, and uncertainty assessment of different knowledge-based approaches for soil types mapping. *Catena* **2019**, *182*, 104134. [CrossRef]
45. Dobos, E.; Hengl, T. Chapter 20 Soil Mapping Applications. In *Geomorphometry-Concepts, Software, Applications*; Developments in Soil Science; Elsevier: Amsterdam, The Netherlands, 2009; pp. 461–479.
46. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotic, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [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.