Modeling Soil Organic Carbon Using Remotely-Sensed Predictors: A Case Study from Fuzhou City, China

Terefe Hanchiso Sodango
Fujian Normal University https://orcid.org/0000-0002-9899-7763
Jinming Sha
Fujian Normal University Cangshan Campus: Fujian Normal University
Xiaomei Li (lixiaomei@fjnu.edu.cn)
Fujian Normal University Cangshan Campus: Fujian Normal University
Jiali Shang
Ottawa Research and Development Centre
Zhongcong Bao
Fujian Normal University Cangshan Campus: Fujian Normal University

Research

Keywords: Soil organic Carbon, Remotely-sensed predictors, Land-use, Landform, Lithology

Posted Date: September 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-68374/v1

Modeling soil organic carbon using remotely-sensed predictors: a case study from Fuzhou city, China

Terefe Hanchiso Sodango[a,b], Jinming Sha [a,b,c], Xiaomei Li [d,] Jiali Shang[e], Zhongcong Bao [a,b]*

[a]State Key Laboratory for Subtropical Mountain Ecology of the Ministry of Science and Technology and Fujian Province, Fujian Normal University, Fuzhou, China, [b]School of Geographical Sciences, Fujian Normal University, Fuzhou, China, [c]China-Europe Center for Environment and Landscape Management, Fuzhou, China, [d]College of Environmental Science & Engineering, Fujian Normal University, China

[e]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, 960 Carling Ave, Ottawa, ON K1A 0C6, Canada

Abstract

Background: Assessing the spatial dynamics of soil organic carbon (SOC) is essential for carbon monitoring. Since, variability of SOC is mainly attributed to biophysical land surface variables, integrating a compressive set of such indices may support the pursuit for optimum set of predictor variables. Therefore, this study was aimed at predicting the spatial distribution of SOC in relation to remotely-sensed variables and other covariates. Hence, the land surface variables were combined from remote sensing, topographic, and soil spectral sources. Moreover, the most influential variables for prediction were selected using the RF and Classification and Regression Tree (CART).

Results: The results indicated that the RF model has good prediction performance with corresponding R^2 and RMSE values of 0.96, and 0.91 mg/g, respectively. The distribution of SOC content showed variability across landforms (CV=78.67%), land-use (CV=93%), and lithology (CV=64.67%). Forestland had the highest SOC (13.60 mg/g) followed by agriculture (10.43 mg/g), urban (9.74 mg/g), and water body (4.55 mg/g) land-uses. Furthermore, bauxite and laterite lithology had the highest SOC content (14.69 mg/g) followed by fluvial (14.52 mg/g) and shale (13.57 mg/g), whereas the lowest was predicted in sandstone (5.53mg/g). The mean SOC concentration was 11.70 mg/g, where the majority of area was classified as humous and organo-humus, distributing in the mountainous regions.

The biophysical land surface indices, brightness removed vegetation indices, topographic indices (, and soil spectral bands, respectively were the most influential predictors of SOC.

Conclusion: The spatial variability of SOC may be influenced by landform, land-use, and lithology of the study area. Remotely-sensed predictors including land moisture, land surface temperature and built-up indices added valu-

able information for prediction of SOC. Hence, the land surface indices may provide new insights into SOC modeling in complex landscapes of warm sub-tropical urban regions.

Keywords: Soil organic Carbon, Remotely-sensed predictors, Land-use, Landform, Lithology

1. Background

Soil organic carbon (SOC) is essential for the normal functioning of the ecosystems [1]. It also plays a significant role in global C dynamics and climate change study as it stores the largest total carbon pool of terrestrial ecosystems [2]. In the urban context, the existence of an optimum SOC content is a critical factor for greening projects and is also a good indicator of the state of urban ecosystems and soil quality [3–5]. However, its spatial distribution is influenced by landscape, lithologic and land use factors. Particularly, land use change due to intensive human activities including urbanization and industrialization processes hugely impacts its spatial distribution [6, 7].

In line to this, estimation of the spatial variability of SOC in urban areas has attracted the interests of many researchers. For instance, Chen et al. (2015) studied the SOC densities in urban built-up areas of 35 Chinese cities. The study reported the carbon storage of 35 cities in china including the Fuzhou city (i.e., the current study area)[8]. Similarly, Wang et al. (2017) determined the spatial variation of SOC on a hilly coastal landscape of Wafangdian, Liaoning Province and reported higher contents towards the mountainous areas. Additionally, they confirmed strong influence of land-use on the spatial variation of SOC[9]. Raciti et al. (2011) also made a comprehensive assessment of SOC contents of residential areas and stated that soils in residential areas with past agricultural use had a higher capability to sequester carbon [3, 10]. Likewise, Xia et al. (2017) studied the spatial variations of SOC in relation to land-use change in eastern regions of China and confirmed that land-use change from and into a paddy field had a high impact on SOC variability[11].

Even though, many previous researches estimated SOC in urban areas globally and nationally, the current study area has limited similar studies. Moreover, there is a need to find the optimal set of suitable environmental predictors that may influence prediction of SOC distribution. As the spatial distribution of SOC is highly influenced by various environmental variables, assessing suitability of remotely-sensed predictors and other environmental covariates for mapping of the spatial variability of topsoil SOC in complex urban environment is imperative. To that end, McBratney et al. (2003) also highly recommended the need of new researches so as to select suitable environmental covariates for digital soil mapping [12]. Furthermore, even though the previous researches used multisource datasets, they ignored essential biophysical land surface variables that may add information to SOC distribution. For instance,

soil moisture, land surface temperature and built-up index were omitted. Since, SOC is highly influenced by soil temperature and moisture than any other factors, integrating them may provide new insight for SOC mapping [13, 14]. The previous SOC prediction studies overlooked soil moisture and temperature indices perhaps due to a shortage of high-resolution soil moisture and temperature data. Especially, the traditional ground-based soil-moisture observation networks produce sparse soil-moisture data for smaller regions. Meanwhile, some studies used atmospheric temperature and precipitation as a proxy to measure soil temperature and moisture [15, 16]. However, they also have very coarse resolution for smaller geographic scales. In the meantime, the remote sensing products of soil moisture such as Advanced Scatterometer (ASCAT) and Soil Moisture Ocean Salinity (SMOS) have coarse resolutions (i.e., in tens of kilometers) [17]. But, optical/thermal infrared (TIR) sensor products have higher spatial resolutions (meters to kilometers) and could be good solution for regional and local applications[18]. Variables derived from optical sensor products including vegetation temperature condition index (VTCI) and land surface temperature (LST) can be better choice to obtain the soil moisture and temperature information.

Therefore, the main aim of this study was to identify the role of remotely-sensed variables for SOC prediction and to understand the contribution of the landform, land-use, and lithology on the spatial variation of SOC in the coastal city of Fuzhou, China. To that end, land surface variables such as vegetation temperature condition index (VTCI), land surface temperature (LST), and normalized difference built-up index (NDBI) were integrated with other environmental covariates obtained from multiple sources including remote sensing, topography, and proximal sensing. Meantime, the most important environmental variables were selected and used to estimate the spatial distribution of SOC using a random forest and CART model.

2. Materials and Methods

2.1. Description of the Study Area and Sampling Locations

Fuzhou district is located in the southeastern coastal area of China in the estuary of the Minjiang River. It is the capital city of the Fujian province (Figure 1) and serves as a central city for the Western Taiwan Straits Economic Zone with a total area of 11,462.41 km^2. It is geographically located at 118°08'E to 120°31' East Longitude and 25° 15'N to 26° 29'N North Latitude in the southeast of China. It neighbors with Ningde and Nanping to the north, Quanzhou, and Putian to the south, and Sanming to the west. The city has 13 administrative regions comprising of 6 districts, one county-level city, and six counties. The current study site includes all regions except Pingtan County.

The climatic condition of the area belongs to the humid subtropical maritime climate, with an annual average temperature of 16-20℃ and the annual average precipitation of 900-1200 millimeters.

It is also covered by acidic volcanic rocks and Cretaceous sandstones from Jurassic Period [19, 20]. The soil of the upper part is dominated by red soil [21, 22], whereas mountainous regions have mainly red and laterite soil.

Additionally, the area is characterized by complex topographic features. The northern, western, and southern parts of the study area are dominated by mountains, whereas the eastern part is mainly plain landform [23].

2.2. Soil Sampling and Laboratory Analysis

Topsoil samples (0-20 cm depth) were collected from all administrative counties of Fuzhou City except Pingtan Island during the February and March months. The soil samples were collected using purposively distributed sample points that represent the dominant land cover, soil types, and landform of the study area. To that end, the land use map, the topographic map and lithology map were superimposed using ArcGIS 10.3 to identify sample points. A total of 244 sample sites were selected as predetermined sampling points. However, due to complex topography, land-use, lack of budget, and accessibility problems 121 samples were collected. Hand-held Global Position System (GPS) receiver was used to identify sample points and to capture the location information. Then, the samples were transported to the laboratory, air-dried, and sieved. The samples were separated for spectroscopic measurements and chemical analysis. The SOC was determined using a dry combustion method using a CNS elemental analyzer (Flash EA 1112 NC-Soil, Thermo Fisher Scientific, Pittsburgh, PA). Spectral soil data was measured using an indoor spectral measurement in the wavelength ranges of 350nm to 2500nm using the FieldSpec-3 spectroradiometer. The sampling intervals were set to be 1nm, and as a result, a total of 2151 bands were produced.

2.3. Description and Pre-processing of Landsat-8 Images, and Soil Spectral Data Transformations

Landsat images are commonly used for SOC predictions due to their high spectral resolution[24]. This study also used Landsat-8 data of Operational Land Imager (OLI) Level 1, Surface Reflectance Level 2, and Thermal Infrared Sensor (TIRS). The image acquired in December 2013 having less cloud cover (i.e., <10%) with path and raw numbers of 119 and 42 were downloaded from United States Geological Survey archives (https://earthexplorer.usgs.gov/). The radiometric calibration was performed using ENVI. Radiometric calibration was done by converting the images into the top of the atmosphere (TOA) reflectance using radiometric rescaling coefficients provided in the product metadata (MTL) file[25].

Moreover, bandwidth of the spectral soil data was ranging from 350nm to 2500nm. Different spectral transformations such as multiplicative scatter correction (MSC), first derivative, second derivative, normalization, and continuum removal were applied to remove background noise [26]. Moreover, smoothing approaches of the Savitzky-Golay filtering algorithm (SG) with a second-order polynomial and averaging was performed across a 10-band window to remove the complexity of bandwidths by eliminating redundancy between adjacent bands and compressing band data without losing information [27].

2.4. Data Sources, Software, and Extraction of Environmental Covariates

The study integrated covariates obtained from multisource including remote sensing, Digital Elevation Models (DEM), soil maps, soil spectral data, and other sources (e.g. OSM layers). The environmental covariates were extracted using the equations provided in Table 2. To that end, R, ArcGIS 10.3, ENVI, QGIS, social sciences (SPSS) version 25.0, and SAGA GIS software were used.

The LST of the study area was calculated using Landsat 8 TIR bands, while the NDBI was used to extract built-up areas [28]. A threshold value of 0.038 was used to extract the built-up area. Additionally, the spectral bands of the Landsat-8 image were used along with soil spectral data. Additionally, the DEM data with a resolution of 10m was obtained from aerial imageries and used to extract landform, hydrological, and spatial indices. The topography was classified using the approaches described in FAO guidelines for soil description [62]. Additional landform characteristics of the study area was described using different landscape metrics including slope, relief, curvature, TWI, TPI, and others. Furthermore, proximity to essential features such as industries, landfill sites, water-points, and port points was calculated as the Euclidean distance from sampling points [29].

All environmental covariate maps and point datasets were converted into similar formats and stacked using ArcGIS 10.3 software. All datasets were resampled into the 30m pixel using 'bilinear resampling' for continuous data and 'nearest' for categorical data and converted into the same coordinate and projection systems in the R environment[30].

The land-use /cover of the study area was classified using a supervised image classification technique. Representative ground control points (i.e., 50 points for each class) were captured to compare the class signatures. Out of 50 ground control points, 40 were used for classification calibration, whereas the remaining ten were used for validation. The land-use was classified into four predominant (i.e., urban, vegetation (forest), agriculture, and water bod-

ies) covers, as can been seen in Figure 8. Finally, classification results and accuracy of classification were generated (Table 3).

2.5. Statistical Analysis, Spatial Modeling and Validation

The summary statistics were calculated using standard descriptive statistics for the variates. The coefficient of variation (CV) and the Pearson correlation coefficients were used to measure the spread of the mean and the linear dependence between SOC values and environmental covariates, respectively. Similarly, Wilding (1985) and Nielsen and Bouma (1985) used CV values to characterize the variability into classes of low (0-15), moderate (16-35%) and high (> 36%) SOC variability [31, 32].

A combination of random forest and CART were used for prediction, performance evaluation, and selection of the essential variables. Random forest avoids over-fitting and provides reliable error estimates of out-of-bag (OOB), avoiding the need for an independent validation dataset [33].

The 75 % of soil samples were used for the training of a model, and identifying essential variables for prediction, and then the model was applied to the validation set. Additionally, CART, was implemented using the R package "rpart" to improve interpretability[34]. The generated regression forest with a minimum split factor of 4, is seen in figure 12. Similarly, Wiesmeier et al. (2011) used a combination of CART and RF for SOM determination using limited samples in semi-arid steppes in Northern China [35].

The model parameters (i.e., ntree, mtry, and nodesize) were optimized using both tuning and manual adjustment, where tune random forest and manual adjustments were done iteratively on training datasets in R [36]. Liaw and Wiener (2002) confirmed that applying the tuning function can improve the results of the model [37]. Hence, based on the previous literature, the tuning function was applied to training data and used to select the number of mtry and number of the trees grown. Finally, the mtry and the total number of trees selected by tune function to grow forest was crosschecked with manual mtry entries. The trained and developed model was first applied to the training dataset and used to identify essential variables for the prediction. The variable importance curve was implemented recursively for 15 times, and the highly influential variables that frequently appeared on the top rank were selected accordingly [38, 39].

A tree was built from a bootstrap sample of the original dataset, which allows for robust error estimation with the remaining test of Out-Of-Bag (OOB) samples. The OOB samples were predicted from the bootstrap samples, and

the mean square error (MSE_{OOB}) that depends on the samples that are omitted from the bootstrapped samples of OOB was computed as stated in equation 4.

$$MSE_{OOB} = n^{-1} \sum_{i=1}^{n} (z_i - \hat{z}_i^{OOB})^2 \quad \text{.....(4)}$$

Where n-is number of observations, z_i-is average prediction of the i^{th} observation, z_i^{OOB}- is the average prediction for the i^{th} observation from all trees for which the observation was OOB.

Additionally, the percentage of explained variance (Var_{ex}) was calculated as:

$$Var_{ex} = 1 - \frac{MSE_{OOB}}{Var_z} \quad \text{..... (5)}$$

Where, Var_z-is the total variance of the response variable.

The ntree parameter (the number of trees in the forest) was adjusted using the mean squared error (MSE values) as a measure of the prediction accuracy of the RF model (figure 13). Similar to this study, the MSE error estimate was used in the validation procedure [40]. Wiesmeier et al. (2011) also used MSE_{oob} for validation of the RF model in the prediction of soil organic matter [35]. OOB estimate was used to evaluate prediction performance as it solves the problem of collecting an independent validation dataset [41]. Many previous studies that used RF for SOC prediction have rarely used cross-validation since RF internally estimates errors during the running of the model. For instance, Minasny et al. (2013a) reviewed studies that focused on SOC prediction and reported that more than half of the studies do not show validation [42].

Ntree and mtry were selected to be 500 and 52, respectively, as a large number of trees is recommended for datasets with sophisticated features and when the emphasis is given for identifying essential variables [40] (Figure 15).

The statistical indices of root mean square error (RMSE) and mean error (ME) were used, as stated in equations 7 and 8 to evaluate the performances of the model.

$$ME = \frac{1}{n} \sum_{i=1}^{n} (\hat{x}\imath - yi) \quad \text{.....(7)}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} ((\hat{x}\imath - yi)^2} \quad \text{.....(8)}$$

Where n: number of data points, x_i measured values, and y_i: predicted values, $\overline{x_\imath}$ mean measured value, $\overline{y_\imath}$, mean predicted values

3. Results

3.1. Spatial prediction of SOC and Model Validation

The result of prediction was evaluated on the training and test datasets. The statistics of validation were provided in Table 8. The ME of training set varied from 0.05 mg/g to 0.06 mg/g with the mean value of 0.04 mg/g, while the RMSE values ranged from 1.35 mg/g to 1.38 mg/g with the mean value of 1.37 mg/g. The ME of the test set varied from 0.3 mg/g to 0.43 mg/g with the mean value of 0.4 mg/g, while the RMSE values ranged from 0.94 mg/g to 0.97 mg/g with the mean value of 0.96 mg/g. In addition, the tune random forest model applied to select optimum number of mtry and number of trees was verified. Accordingly, the Out-of-bag (OOB) errors was 0.44, 0.22, 0.11 and 0.09 for mtry values of 9, 17, 34, and 52, respectively, implying using more number of predictor variables had lower OOB error (figure 14) and better accuracy for prediction of SOC in the study area. The results indicate that the RF had an excellent performance for SOC prediction. The SOC variability map predicted by RF model using selected important predictors was shown in Figure 16.

The mean concentration of SOC was 11.70 mg/g, with a minimum value of 0.7 mg/g and the maximum value of 45.80 mg/g. It was highly variable with the coefficient of variation (CV) of 81% and a standard deviation of 8.92. The SOC concentration was grouped into six levels (i.e., low, moderate, high, humus, and Organo-humus classes.) as mentioned in Zhang et al. (2011) [43]. As can be portrayed from the reclassification map (Figure 17) of SOC levels distribution, a shallow level of SOC content was concentrated in the middle of the study area where the highly urbanized downtown with high NDBI values is located comprising about 2.01 % of the total area. Moderate SOC level was predicted in the surroundings of the urbanized area, whereas the high SOC level was predicted in forested (vegetated) areas. Based on the reclassification results, 0.01 % of the study area had organic SOC levels. The remaining area was covered by moderate, high, humus, and Organo-humus SOC levels, with a proportion of 5.11%, 2.17%, 47.51%, and 43.19%, respectively.

Humous SOC level was covering about half of the study area (47.51 %), spreading towards the east to west through the middle of the study area. Organo-humus SOC level also took almost the remaining half of the study area (43.19%), mainly concentrating in mountainous areas of the southern and northern parts, whereas moderate SOC level was located in the central parts.

3.2. Importance of Environmental Variables

Based on the IncNodePurity index model, the top fifteen ranking important variables for prediction of SOC distribution in the study area were VTCI, ASTAVI, EVI, Lithology, NDBI, TPI, Slope, X2224, LST, Band4, X424., RVI, TWI, Band7, and NDVI (Figure 18).

Based on the result, the VTCI index was the most crucial variable for the prediction of SOC from all 52 predictors, followed by VIs, lithology, and NDBI. The remaining variables, such as TPI, Slope, X2224, LST, Band4, X424, RVI, TWI, Band7, and NDVI, occupied the remaining top fifteen positions as influential variables. NDBI and LST were among the influential variables with rankings of fifth and ninth places.

Similarly, topographic variables of TPI and Slope had superior impact ranking in sixth and seventh positions, respectively, while TWI was thirteenth.

3.3. Influence of Topographic and Vegetation Indices

The topographic and hydrological factors such as slope, curvature, aspect, TPI, MBI, and TWI of the study area had high variability with increased CV (see Table 4). Slope, aspect, and MBI were more variable with > 50 % CV than curvature, TPI, and TWI, which had <50 % CV values. Additionally, the Pearson correlation values of the slope, curvature, aspect, TPI, MBI, and TWI were variable. Slope, curvature, and MBI were negatively correlated to SOC, while aspect, TPI, and TWI were positively correlated (Table 4). However, only MBI and Curvature were significantly correlated with SOC at ($p<0.05$). Moreover, the LS factor was highly variable, with a CV of 68 %. On the other hand, all VIs had a very low CV except for RVI and BI indices that had slight variability. Additionally, the Pearson correlation values showed that NDVI, ASTAVI, GI, EVI, SAVI, CTVI, TVI, NRVI, RVI, and TSAVI_91 had a significant positive correlation with SOC concentration at $p<0.05$ (see Table 4).

3.4. SOC Distribution Across Landform, Land-Use, and Lithology

The distribution of SOC content across landforms, land-use, and lithology was examined. The result shows that its distribution was highly variable across landforms with a mean CV of 78.67 %. The highest SOC variation was recorded within medium-gradient Mountains(SM) with a CV of 122.08% (Table 5), followed by high-gradient mountains (TM) (12.64 mg/g). Considerably high SOC contents were predicted in plain, medium gradient hills (SH) and medium-gradient mountains (SM) with mean values of 11.94 mg/g, 11.88 mg/g, and 10.90 mg/g, respectively. On the contrary, a lower proportion of SOC was predicted in high-gradient hills (TH) and shoreline (WR) with mean values of 7.47 mg/g and 1.67 mg/g, respectively.

Land-use wise, forestland had the highest mean SOC contents (13.60 mg/g) than other land uses followed by agricultural (10.43 mg/g), urban (9.74 mg/g), and water body (4.55 mg/g) (Table 6).

Furthermore, the highest amount of SOC was recorded in weathered residuum (dominated by bauxite and laterite) with a mean SOC content of 14.69 mg/g. Similarly, an increased amount of SOC was predicted on fluvial (14.52 mg/g) and shale lithology (13.57 mg/g). However, unconsolidated marine rock, granite, gneiss (migmatite), and pyroclastic (ignimbrite) lithology had lower contents of SOC with proportions of 10.99mg/g, 11.61mg/g, 11.91mg/g, and 12.04mg/g, respectively (Table 7). The lowest SOC content was predicted in sandstone (greywacke) or arkose lithology (5.53mg/g), inland water (lakes) (7.47 mg/g), and siltstone, mudstone, clay stone (7.92 mg/g).

4. Discussions

4.1. Spatial variability of SOC

The results of this study indicated that the RF model has good prediction performance with corresponding R^2 and RMSE values of 0.96, and 0.91 mg/g, respectively. The result is consistent with previous studies, which reported RF as an accurate model for prediction of SOC [44, 45]. Moreover, the spatial distribution map of SOC had a similar trend with distribution of selected environmental factors including soil moisture, soil temperature, and the extent of impervious surface (see Figures 9, 10, and 11) suggesting their strong relationships. The built-up area had the lowest SOC content but the forested and mountainous regions had the highest SOC content suggesting the high influence of impervious surface for such disparities.

4.2. Importance of Environmental Variables

VTCI was the most essential variable for the prediction of SOC from 52 environmental predictors. The main reason for the superior influence of VTCI on the spatial prediction of SOC distribution may be attributed to its precision to measure the crop water status and the subsequent impacts of soil moisture on the aboveground biomass. Moreover, since VTCI is derived based on the relationship between land surface temperature (LST) and vegetation index (NDVI), it could provide more information for the spatial prediction of SOC. Alvarez and Lavado (1998) also reported that the SOC contents of the topsoil was highly correlated to moisture to temperature ratios[46]. However, the Wetness Index (WI), derived as tasseled cap transformation (TCT) provided little information for prediction of SOC distribution in the study area. This result delivered understanding about the soil moisture derived in combination with vegetation and temperature ratio (VTCI) may be a better predictor for SOC mapping in a similar environment.

The reason for high importance of VTCI for SOC mapping can be due to its control on the extent of vegetation cover (i.e., quantity and quality of OM enters into the soil), the rate of mineralization, and litter decomposition [47, 48]. Moreover, since soil moisture is an essential element for microbial growth, it may facilitate the degradation of plant and animal residues that improves SOC contents [13, 49].

ASTAVI and EVI were the second and third influential variables for the prediction of SOC distribution. The reason for their influence may be due to their ability to provide proxies for measuring aboveground biomass that may influence SOC contents stored to the soil as a litter [50, 51]. Additionally, the characteristics of ASTAVI (i.e., low sensitivity to soil backgrounds) may contribute its share to derive more information related to SOC contents than other vegetation indexes [52]. Similarly, enhanced vegetation index (EVI) has low soil, and atmospheric effects than other VIs used in this study. The exclusively stronger influences of the ASTAVI and EVI than other vegetation indices suggest that VIs with minimized brightness-related soil effects (i.e., ASTAVI, and EVI) may perform better than RVI, NDVI, and others for SOC prediction in the complex landscape such as Fuzhou city. However, the previous studies have not separately used VIs based on their strength of reducing background effects in predictions of SOC rather they used a mixture of both. For instance, Peng et al. (2015) also confirmed that EVI was one of the top predictors [53]. Compared to remote sensing raw-bands, vegetation indices performed better perhaps due to their ability to accurately inferring crop/vegetation/bare soil characteristics. This result suggested that VIs were good predictors, but VIs that significantly remove soil color effects were better predictors for SOC contents.

NDBI was one of the top predictors for SOC distribution since impervious surface may impact the spatial distribution of SOC. This result shows that NDBI index can be used for SOC prediction at the city scale in a complex urban landscape where the land-use change into built-up areas is prominent. Similarly, previous studies reported that land-use change associated to urbanization processes was profoundly influencing total carbon fluxes. For instance, Raciti et al. (2011) compared the carbon (C) pools in residential areas with similar soil type to forested reference sites and reported substantial variability[3]. Hence, the selection of the NDBI variable among the top predictors suggests that land surface variables maybe among highly influential predictors for soil properties modeling in complex urban landscape. This result also suggests that residential and non-residential areas may have a diverse pattern of SOC distribution.

Additionally, a significant positive correlation between the NDBI and SOC ($r = 0.26$, $p < 0.05$) also explained strong relationships. The high extent of impervious surface may impact SOC contents since increased human dis-

298 turbances and a mix with artificial materials from buildings causing variability. Similarly, Yan et al. (2015) con-
299 firmed the importance of impervious surface on SOC quantification and stated that SOC from impervious surface
300 accounted for over half of the city's SOC stock in their study[54].
301 Likewise, the LST was correspondingly among the influential variables where SOC shows decreasing trends when
302 LST increased. The result is in agreement with the previous research that confirmed temperature was an important
303 variable for SOC contents[55].
304 The soil spectral data with 2224nm and 424nm wavelength along with LANDSAT-8 row spectral data (red and
305 shortwave infrared 2) were among influential predictors. Similarly, other studies confirmed hyperspectral remote
306 sensing data as key predictors for SOC [56, 57]. Peng et al. (2015) also reported that Landsat bands combined with
307 VIS-NIR were efficient for the prediction of SOC in the topsoil [58].
308 Generally, biophysical land surface indices of VTCI, LST, and NDBI, brightness removed vegetation indices (i.e.,
309 ASTAVI and EVI), topographic indices (i.e., TPI and Slope), soil spectral bands (i.e., 424nm and 2224 nm), respec-
310 tively were the most influential variables for SOC prediction in the study area suggesting their potential importance
311 in similar complex urban landscape.

312 **4.3. Influence of Topographic and Vegetation Indices**

313 MBI ($r=-0.18$, $p < 0.05$) and curvature ($r=-0.19$, $p<0.05$) were significantly negatively correlated with SOC contents
314 emphasizing the influence of the prominent variability of the topographic and hydrological factors of the study area.
315 Other studies also confirmed the high influence of landscape and hydrological variables on the SOC patchiness [59,
316 60]. Furthermore, there was a negative correlation between slope and SOC that can be explained by erosion and
317 deposition processes [61]. Similarly, Stevens (2014) reported a positive correlations between SOC contents and as-
318 pect, TPI, and TWI [62]. The positive relationship between aspect and SOC contents is explained by the direction of
319 slopes to the sun that may cause variation in temperature that leads to decomposition. The slope directing to shady
320 slopes may have the lower decomposition due to low soil temperature. Low soil wetness results in a decrease in
321 SOC since it affects microbial activity [62]. Variability of the LS factor may also have impacted the amount of dep-
322 ositions [63, 64].
323 Compared with topographic variables, vegetation indices were better correlated with SOC contents.

4.4. SOC Distribution Across the Landform, Land-Use, and Lithology

Prediction of increased SOC contents in high-gradient mountains (TM), medium-gradient hills (SH), and medium-gradient Mountains (SM) may be due to the availability of large share of vegetation covers in these landforms. Additionally, landform-related lithological, moisture, and temperature variations may have influenced the SOC distribution along the altitudinal gradient [65, 66]. The reason for decent SOC contents in plain landform may be associated to the abundance of fluvial materials. Additionally, the dominant soil types were paddy-soil, with a small proportion of red soil and plaster fields. The paddy field of the area was undergone through intensive agricultural practices where application of organic fertilizers may have influenced the SOC content of the landform[11].

Low SOC contents in the high-gradient hills (TH) landform may be attributed to the impact of complex land-use in the area and disturbance posed by intensive human activities. Kemen port, transportation hubs, and large development projects in proximity may have contributed to the decrease of SOC contents[67]. Additionally, the Luoyuan Bay located in this area may also have contributed for low SOC distribution in this landform. Wu et al. (2013) reported the soil sediments in Luoyuan Bay had a high level of eutrophication [68], suggesting the SOC contents of this area may be washed into the surrounding water body. Another reason could be increased use of nitrogen fertilizers and fossil fuels in the surrounding areas may have stimulated the loss of organic carbon from terrestrial soils into the surrounding water bodies through erosion [68].

In general, the landform variations may contribute to changes in soil properties (clay content), human activities (i.e., land-use), and vegetation quality and quantity, alteration of climatic variables (temperature and precipitations) that may affect the SOC distribution. Similarly, previous studies reported that landform elements play a significant role in the variability of SOC[69, 70]. Even though the variability of SOC depends on geologic and climatic, topographic, vegetation[32], and land-use variables [71], landform plays a crucial role in modifying all these factors [72].

Additionally, the SOC distribution was highly variable across the land-use. The study area was characterized by highly urbanized downtown, which was mainly covered by impervious surfaces to a hilly and mountainous area which is dominated by plantation forests. Therefore, these land-use/cover dynamics might contribute to SOC variations in the area. Similarly, Chuai et al. (2012) reported a higher SOC density in towns, woodland, paddy land, and shallow water areas due to industrial and human influence [73].

Additionally, weathered residuum had the highest content of SOC. It may be due to possession of aluminosilicate red soil (acid red soil), Fe, and permeates paddy soil. The Fe leaching, coupled with the high activities of microor-

ganisms, may lead to the increased content of SOC in the weathered residuum, bauxite, and laterite lithology. The previous study reported increased SOC in bauxite lithology due to increased pyrite owing to the redox reactions[74]. Moreover, a considerable amount of the area dominated by laterite lithology was used for agriculture, where there was the cultivation of paddy rice and long-term application of feedlot manure to cropland. Therefore, the farm management systems may have affected the SOC contents. Similar results were reported by Lui et al. (2016) that stated long-term fertilization practices profoundly influenced the SOC content of red soil of southern China [75]. However, the lowest content of SOC in the sandstone dominated area may be attributed to the least weathering of the sandstone[76, 77]. Similarly, a high concentration of SOC in siliceous red soil, red clay, and yellow-red soils can be related to the high binding capability to organic matter. Additionally, the high content of silicon material may have influenced the chemical and physical properties since the soils of this area were originated from the arenaceous rock. The fluvial lithology was mainly located in the eastern parts of the Mingjian river banks of the study area, where the landform was dominated by plain (LP) surface. The reason for lower concentration of SOC contents in acid red soil, silicon aluminum red soil, and red sand soils may be related to the higher content of sand and low decomposition rate. The result is consistent with Zhang et al. (2010) that reported similar results in the hilly red soil region of South China[78].

Similarly, previous studies reported consistent results about the contribution of the landform, land-use, and lithology on the spatial variability of SOC [79–81].

5. Conclusion

This study was aimed at predicting the spatial distribution of SOC in relation to environmental covariates, including land temperature, soil moisture, and extent of urbanizations using the RF and CART models in the coastal city of Fuzhou city, China. To that end, a compressive set of biophysical land surface variables such as LST, VTCI, and NDBI were combined with other environmental covariates. The environmental covariates extracted from remote sensing, topography, and soil spectral sources were used to predict the SOC distribution and to select the most influential variables for the spatial prediction.

The results indicated that the RF had an excellent performance for SOC prediction. The SOC content of the study area was highly variable owing to the heterogeneity of the landform, land use, lithology, surface temperature, soil moisture contents, and the rate of built-up. Biophysical variables including soil moisture status index (VTCI), adjusted transformed soil-adjusted vegetation index (ASTAVI), enhanced vegetation index (EVI), lithology, and built-

up index (NDBI) were the five most influential predictors by hugely contributing for the prediction of SOC in the study area. The results suggested that biophysical land variables of VTCI, LST, and NDBI were good predictors. Additionally, the selection of NDBI as one of the essential predictors may provide an insight to predict SOC in residential areas.

The current study has derived biophysical land variables such as soil moisture, land surface temperature, and human influence using substantially improved indices, which were often ignored in the prediction of SOC in previous studies. However, the indices were among the most influential variables for the prediction of the spatial distribution of SOC in a complex coastal urban environment of Fuzhou city. Even though, the variables were derived from high-resolution Landsat-8 images, the result might be further improved in the future studies by using better resolution images such as Sentinel products.

This result shows that similar approaches and biophysical land variables can be used in other regional and local level SOC prediction studies in similar sub-tropical coastal environments.

Declaration

Data Availability Statement

The data used for this study is part of an interdisciplinary project on Land Degradation (IUCLAND). Data supporting the conclusions of this manuscript will be made available by the corresponding author. All supporting data will be publicly available.

Competing Interests

The authors declare no competing interests.

Funding

This research was supported by Innovation on Remote Sensing Education and Learning (IRSEL) grant number (IRSEL 586037-EPP-1-2017-1-HU-EPPKA2-CBHE-JP (2017 - 2980 / 001 - 001)), Interdisciplinary Project on Land Degradation (IUCLAND), and International Cooperation Project on “Integrated geospatial information technology and its application to resource and the Chinese Ministry of Science and Technology (MOST).

Contributions

THS, XL, JSha, JS, and ZB conceived the study and collected field soil samples, performed analysis, and validation. THS wrote the draft paper. JS edited the paper. XL and JSha procured funding and supervised. All authors read and approved the final manuscript.

Corresponding author: Xiaomei Li; lixiaomei@fjnu.edu.cn

Acknowledgements

We are grateful to the editor and anonymous reviewers who have contributed immensely to the improvement of this article.

6. Reference

1. Tiessen H, Cuevas E, Chacon P (1994) The role of soil organic matter in sustaining soil fertility. Nature 371:783–785
2. Oyonarte C, Aranda V, Durante P (2008) Soil surface properties in Mediterranean mountain ecosystems: Effects of environmental factors and implications of management. For Ecol Manage 254:156–165
3. Raciti SM, Groffman PM, Jenkins JC, Pouyat R V., Fahey TJ, Pickett STA, Cadenasso ML (2011) Accumulation of Carbon and Nitrogen in Residential Soils with Different Land-Use Histories. Ecosystems. https://doi.org/10.1007/s10021-010-9409-3
4. Yan Y, Kuang W, Zhang C, Chen C (2015) Impacts of impervious surface expansion on soil organic carbon--a spatially explicit study. Sci Rep 5:17905
5. Edmondson JL, O'Sullivan OS, Inger R, Potter J, McHugh N, Gaston KJ, Leake JR (2014) Urban Tree Effects on Soil Organic Carbon. PLoS One 9:e101872
6. Pouyat R, Groffman P, Yesilonis I, Hernandez L (2002) Soil carbon pools and fluxes in urban ecosystems. Environ Pollut 116:S107–S118
7. Hall SJ, Baker MA, Jones SB, Stark JM, Bowling DR, Ecosyst U Contrasting soil nitrogen dynamics across a montane meadow and urban lawn in a semi-arid watershed. https://doi.org/10.1007/s11252-016-0538-0
8. Chen WY (2015) The role of urban green infrastructure in offsetting carbon emissions in 35 major Chinese cities: A nationwide estimate. Cities 44:112–120
9. Wang S, Zhuang Q, Jia S, Jin X, Wang Q (2017) Spatial variations of soil organic carbon stocks in a coastal hilly area of China. Geoderma 314:8–19
10. Raciti SM, Hutyra LR, Rao P, Finzi AC (2012) Inconsistent definitions of "urban" result in different conclusions about the size of urban carbon and nitrogen stocks. Ecol Appl. https://doi.org/10.1890/11-1250.1
11. Xia X, Yang Z, Xue Y, Shao X, Yu T, Hou Q (2017) Spatial analysis of land use change effect on soil organic carbon stocks in the eastern regions of China between 1980 and 2000. Geosci Front 8:597–603
12. McBratney A., Mendonça Santos M., Minasny B (2003) On digital soil mapping. Geoderma 117:3–52

13. Curtin D, Beare MH, Hernandez-Ramirez G (2012) Temperature and Moisture Effects on Microbial Biomass and Soil Organic Matter Mineralization. Soil Sci Soc Am J. https://doi.org/10.2136/sssaj2012.0011
14. Shi T, Wang J, Chen Y, Wu G (2016) Improving the prediction of arsenic contents in agricultural soils by combining the reflectance spectroscopy of soils and rice plants. Int J Appl Earth Obs Geoinf 52:95–103
15. WANG D, HE N, WANG Q, LÜ Y, WANG Q, XU Z, ZHU J (2016) Effects of Temperature and Moisture on Soil Organic Matter Decomposition Along Elevation Gradients on the Changbai Mountains, Northeast China. Pedosphere 26:399–407
16. Schrumpf M, Kaiser K, Schulze E-D (2014) Soil Organic Carbon and Total Nitrogen Gains in an Old Growth Deciduous Forest in Germany. PLoS One 9:e89364
17. Bartalis Z, Wagner W, Naeimi V, Hasenauer S, Scipal K, Bonekamp H, Figa J, Anderson C (2007) Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). Geophys Res Lett. https://doi.org/10.1029/2007GL031088
18. Zhang D, Tang R, Zhao W, Tang B, Wu H, Shao K, Li ZL (2014) Surface soil water content estimation from thermal remote sensing based on the temporal variation of land surface temperature. Remote Sens. https://doi.org/10.3390/rs6043170
19. Zhuo L, Tao H, Wei H, Chengzhen W (2016) Plantation in Fujian Province. China PLoS ONE 11:151527
20. Zhang Jinzhang. (1984) BOREHOLE GEOLOGY AND HYDROTHERMAL ALTERATION IN FUZHOU GEOTHERMAL FIELD, FUJIAN PROVINCE, CHINA.
21. Ni J, Li X, Guo J, Wang J, Yang H, Wei R (2011) Distribution Pattern, Sources and Potential Risks of Polycyclic Aromatic Hydrocarbons in Urban Soils of Fuzhou City, China. Mol Environ Soil Sci Interfaces Earth's Crit Zo. https://doi.org/10.1007/978-3-642-05297-2_67
22. Liu H, Lin X, Xie T (2014) Chinese Academy of Surveying and Mapping, Beijing 100830, (CHINA) 3 Administrative board, Guang DongShun DeQing Yuan(Ying De) Economic Cooperation Zone.
23. Liu Y, Su C, Zhang H, Li X, Pei J (2014) Interaction of Soil Heavy Metal Pollution with Industrialisation and the Landscape Pattern in Taiyuan City, China. PLoS One 9:e105798
24. Roy DP, Wulder MA, Loveland TR, et al (2014) Landsat-8: Science and product vision for terrestrial global change research. Remote Sens Environ 145:154–172
25. U.S. Department of Interior (2013) Using the USGS Landsat 8 Product. U.S. Dep. Inter.

26. Cui S, Zhou K, Ding R, Zhao J, Du X (2018) Comparing the effects of different spectral transformations on the estimation of the copper content of Seriphidium terrae-albae. J Appl Remote Sens 12:1
27. Savitzky A, E MJ (1951) Smoothing and Differentiation of Data by Simplified Least Squares Procedures.
28. Grover A, Singh R (2016) Monitoring Spatial Patterns of Land Surface Temperature and Urban Heat Island for Sustainable Megacity. Environ Urban ASIA 7:38–54
29. Peuquet DJ (1992) An algorithm for calculating minimum Euclidean distance between two geographic features. Comput Geosci 18:989–1001
30. Hengl T, Heuvelink GBM, Kempen B, et al (2015) Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. PLoS One. https://doi.org/10.1371/journal.pone.0125814
31. Nielsen DR (ed. ., Bouma J (ed. . (1985) Soil spatial variability; proceedings of a workshop of the ISSS (Int. Society of Soil Science) and the SSSA (Soil Science Society of America), Las Vegas (USA), 30 Nov - 1 Dec 1984.
32. Wilding LP, Drees LR (1983) Spatial variability and pedology. Pedogenes. soil Taxon. / Ed. by L.P. Wilding, N.E. Smeck, G.F. Hall
33. Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146:102–113
34. Breiman L (1984) (34) Classification and regression trees Regression trees. Encycl Ecol. https://doi.org/10.1007/s00038-011-0315-z
35. Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I (2011) Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil 340:7–24
36. Sorenson PT, Small C, Tappert MC, Quideau SA, Drozdowski B, Underwood A, Janz A (2017) Monitoring organic carbon, total nitrogen, and pH for reclaimed soils using field reflectance spectroscopy. Can J Soil Sci 97:241–248
37. Liaw A, Wiener M (2002) Classification and Regression by randomForest.
38. Forkuor G, Hounkpatin OKL, Welp G, Thiel M (2017) High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. PLoS One 12:e0170478

494 39. Vaysse K, Lagacherie P (2017) Using quantile regression forest to estimate uncertainty of digital soil
495 mapping products. Geoderma. https://doi.org/10.1016/j.geoderma.2016.12.017
496 40. Probst P, Boulesteix AL (2018) To tune or not to tune the number of trees in random forest. J. Mach. Learn.
497 Res.
498 41. Ridgeway G (2013) gbm: Generalized Boosted Regression Models. R Packag. version 1.6-3.1
499 42. Minasny B, McBratney AB, Malone BP, Wheeler I (2013) Chapter One – Digital Mapping of Soil Carbon.
500 Adv Agron. https://doi.org/10.1016/B978-0-12-405942-9.00001-3
501 43. Zhang L, Liu Y, Li X, Huang L, Yu D, Shi X, Chen H, Xing S (2018) Effects of soil map scales on
502 simulating soil organic carbon changes of upland soils in Eastern China. Geoderma 312:159–169
503 44. Stum AK, Boettinger JL, White MA, Ramsey RD (2010) Random Forests Applied as a Soil Spatial
504 Predictive Model in Arid Utah. In: Digit. Soil Mapp. Springer, Dordrecht, pp 179–190
505 45. Li D, Wen L, Yang L, Luo P, Xiao K, Chen H, Zhang W, He X, Chen H, Wang K (2017) Dynamics of soil
506 organic carbon and nitrogen following agricultural abandonment in a karst region. J Geophys Res
507 Biogeosciences 122:230–242
508 46. Alvarez R, Lavado RS (1998) Climate, organic matter and clay content relationships in the Pampa and
509 Chaco soils, Argentina. Geoderma. https://doi.org/10.1016/S0016-7061(97)00141-9
510 47. Jenny H, Amundson R (1994) FACTORS OF SOIL FORMATION A System of Quantitative Pedology.
511 48. Wall DH, Bardgett RD, Behan-Pelletier V, Herrick JE, Jones TH, Ritz K, Six J, Strong DR, van der Putten
512 WH (eds) (2012) Soil Ecology and Ecosystem Services.
513 https://doi.org/10.1093/acprof:oso/9780199575923.001.0001
514 49. Nocita M, Stevens A, Noon C, van Wesemael B (2013) Prediction of soil organic carbon for different levels
515 of soil moisture using Vis-NIR spectroscopy. Geoderma 199:37–42
516 50. Sun W, Li X, Niu B (2018) Prediction of soil organic carbon in a coal mining area by Vis-NIR
517 spectroscopy. PLoS One 13:e0196198
518 51. Qi J, Chehbouni A, Huete AR, Kerr YH, Sorooshian S (1994) A modified soil adjusted vegetation index.
519 Remote Sens Environ 48:119–126
520 52. Purevdorj TS, Tateishi R, Ishiyama T, Honda Y (1998) Relationships between percent vegetation cover and
521 vegetation indices. Int J Remote Sens. https://doi.org/10.1080/014311698213795

53. Peng Y, Xiong X, Adhikari K, Knadel M, Grunwald S, Greve MH (2015) Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. PLoS One 10:e0142295
54. Yan Y, Kuang W, Zhang C, Chen C (2015) Impacts of impervious surface expansion on soil organic carbon – a spatially explicit study. Sci Rep 5:17905
55. Gray JM, Bishop TFA, Wilson BR (2015) Factors Controlling Soil Organic Carbon Stocks with Depth in Eastern Australia. Soil Sci Soc Am J 79:1741
56. Vohland M, Ludwig M, Thiele-Bruhn S, Ludwig B, Vohland M, Ludwig M, Thiele-Bruhn S, Ludwig B (2017) Quantification of Soil Properties with Hyperspectral Data: Selecting Spectral Variables with Different Methods to Improve Accuracies and Analyze Prediction Mechanisms. Remote Sens 9:1103
57. Gabriel JL, Zarco-Tejada PJ, López-Herrera PJ, Pérez-Martín E, Alonso-Ayuso M, Quemada M (2017) Airborne and ground level sensors for monitoring nitrogen status in a maize crop. Biosyst Eng 160:124–133
58. Peng Y, Xiong X, Adhikari K, Knadel M, Grunwald S, Greve MH (2015) Modeling Soil Organic Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. https://doi.org/10.1371/journal.pone.0142295
59. Cheng S, Fang H, Zhu T, Zheng J, Yang X, Zhang X, Yu G (2010) Effects of soil erosion and deposition on soil organic carbon dynamics at a sloping field in Black Soil region, Northeast China. Soil Sci Plant Nutr 56:521–529
60. Martinez C, Hancock GR, Kalma JD (2010) Relationships between 137Cs and soil organic carbon (SOC) in cultivated and never-cultivated soils: An Australian example. Geoderma 158:137–147
61. Rhanor T (2013) TOPOGRAPHIC POSITION AND LAND COVER EFFECTS ON SOIL ORGANIC CARBON DISTRIBUTION OF LOESS-VENEERED HILLSLOPES IN THE CENTRAL UNITED STATES.
62. Stevens F, Bogaert P, Van Oost K, Doetterl S, Van Wesemael B (2014) Regional-scale characterization of the geomorphic control of the spatial distribution of soil organic carbon in cropland. Eur J Soil Sci 65:539–552
63. McBride MB (1994) Environmental chemistry of soils. Environ. Chem. soils.
64. Yazidhi B (2003) A comparative study of soil erosion modelling in Lom Kao-Phetchabun, Thailand.
65. Schuler U, Herrmann L, Ingwersen J, Erbe P, Stahr K (2010) Comparing mapping approaches at

subcatchment scale in northern Thailand with emphasis on the Maximum Likelihood approach. CATENA 81:137–171

66. Bui LV, Stahr K, Clemens G (2017) A fuzzy logic slope-form system for predictive soil mapping of a landscape-scale area with strong relief conditions. CATENA 155:135–146

67. Peterson BJ, Melillo JM (2017) The potential storage of carbon caused by eutrophication of the biosphere. 37:117–127

68. Wu H-Y, Chen K-L, Zhang P, Fu S-F, Hou J-P, Chen Q-H (2013) [Eco-environmental quality assessment of Luoyuan Bay, Fujian province of East China based on biotic indices]. Ying yong sheng tai xue bao = J Appl Ecol 24:825–31

69. GESSLER PE, MOORE ID, McKENZIE NJ, RYAN PJ (1995) Soil-landscape modelling and spatial prediction of soil attributes. Int J Geogr Inf Syst 9:421–432

70. Sherwood WC, Hartshorn AS, Eaton LS (2010) Soils, geomorphology, landscape evolution, and land use in the Virginia Piedmont and Blue Ridge. In: Mid-Atlantic Shore to Appalach. Highl. F. Trip Guideb. 2010 Jt. Meet. Northeast. Southeast. GSA Sect. Geological Society of America, pp 31–50

71. Easter M, Paustian K, Killian K, Coleman K, Bernoux M, Falloon P, Batjes N, Cerri CC (2007) Simulating SOC changes in 11 land use change chronosequences from the Brazilian Amazon with RothC and Century models. Agric Ecosyst Environ 122:46–57

72. Florinsky I., Eilers R., Manning G., Fuller L. (2002) Prediction of soil properties by digital terrain modelling. Environ Model Softw 17:295–311

73. CHUAI X-W, HUANG X-J, WANG W-J, ZHANG M, LAI L, LIAO Q-L (2012) Spatial Variability of Soil Organic Carbon and Related Factors in Jiangsu Province, China. Pedosphere 22:404–414

74. Ellahi SS, Taghipour B, Nejadhadad M, Salamab Ellahi S, Taghipour B, Nejadhadad M (2017) The Role of Organic Matter in the Formation of High-Grade Al Deposits of the Dopolan Karst Type Bauxite, Iran: Mineralogy, Geochemistry, and Sulfur Isotope Data. Minerals 7:97

75. Liu Y, Gao P, Zhang L, Niu X, Wang B (2016) Spatial heterogeneity distribution of soil total nitrogen and total phosphorus in the Yaoxiang watershed in a hilly area of northern China based on geographic information system and geostatistics. Ecol Evol 6:6807–6816

76. Ahn C, Jones S (2013) Assessing Organic Matter and Organic Carbon Contents in Soils of Created

578 Mitigation Wetlands in Virginia. Environ Eng Res 18:151–156
579 77. Kidd D, Malone B, McBratney A, Minasny B, Webb M (2015) Operational sampling challenges to digital
580 soil mapping in Tasmania, Australia. Geoderma Reg. https://doi.org/10.1016/j.geodrs.2014.11.002
581 78. Zhang Y, Yu Q, Ma W, Chen L (2010) Atmospheric deposition of inorganic nitrogen to the eastern China
582 seas and its implications to marine biogeochemistry. J Geophys Res 115:D00K10
583 79. Forkuor G, Hounkpatin OKL, Welp G, Thiel M (2017) High Resolution Mapping of Soil Properties Using
584 Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and
585 Multiple Linear Regression Models. PLoS One 12:e0170478
586 80. Grimm R, Behrens T, Märker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on
587 Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146:102–113
588 81. Samuel-Rosa A, Heuvelink GBM, Vasques GM, Anjos LHC (2015) Do more detailed environmental
589 covariates deliver more accurate soil maps? Geoderma 243–244:214–227
590 82. Zha Y, Gao J, Ni S (2003) Use of normalized difference built-up index in automatically mapping urban
591 areas from TM imagery. Int J Remote Sens 24:583–594
592 83. Zhou Y, Yang G, Wang S, Wang L, Wang F, Liu X (2014) A new index for mapping built-up and bare land
593 areas from Landsat-8 OLI data. Remote Sens Lett 5:862–871
594 84. Weiss A (2001) Topographic position and landforms analysis. 200:
595 85. JENNESS, J (2006) Topographic position index (tpi_jen.avx_extension for Arcview 3.x, v.1.3a, Jenness
596 Enterprises [EB/OL]. http://www.jennessent.com/arcview/tpi.htm
597 86. Rasul A, Balzter H, Smith C (2015) Spatial variation of the daytime Surface Urban Cool Island during the
598 dry season in Erbil, Iraqi Kurdistan, from Landsat 8. Urban Clim 14:176–186
599 87. Thompson JA, Bell JC, Butler CA (2001) Digital elevation model resolution: effects on terrain attribute
600 calculation and quantitative soil-landscape modeling. Geoderma 100:67–89
601 88. Wan Z, Wang P, Li X (2004) Using MODIS Land Surface Temperature and Normalized Difference
602 Vegetation Index products for monitoring drought in the southern Great Plains, USA. Int J Remote Sens
603 25:61–72
604 89. Avdan U, Jovanovska G (2016) Algorithm for Automated Mapping of Land Surface Temperature Using
605 LANDSAT 8 Satellite Data. J Sensors 2016:1–8

606 90. IDB - Index DataBase. https://www.indexdatabase.de/. Accessed 26 Oct 2018
607 91. Rondeaux G, Steven M, Baret F (1996) Optimization of soil-adjusted vegetation indices. Remote Sens
608 Environ 55:95–107
609 92. Ford WM, Menzel MA, Odom RH (2002) Elevation, aspect, and cove size effects on southern Appalachian
610 salamanders. Southeast Nat 1:315–324
611 93. Weier J, Herring D (2000) Measuring Vegetation (NDVI and EVI). NASA Earth Obs.
612 94. Chejarla VR, Maheshuni PK, Mandla VR (2016) Quantification of LST and CO_2 levels using Landsat-8
613 thermal bands on urban environment. Geocarto Int 31:913–926
614 95. Pennock DJ, Zebarth BJ, De Jong E (1987) Landform classification and soil distribution in hummocky
615 terrain, Saskatchewan, Canada. Geoderma 40:297–315
616 96. Sørensen R (2006) Topographical influence on soil chemistry.
617 97. Jones KH (1998) A comparison of algorithms used to compute hill slope as a property of the DEM. Comput
618 Geosci 24:315–323
619 98. MacMillan RA, Shary PA (2009) Chapter 9 Landforms and Landform Elements in Geomorphometry. Dev
620 Soil Sci 33:227–254
621 99. Taghizadeh-Mehrjardi R, Nabiollahi K, Kerry R (2016) Digital mapping of soil organic carbon at multiple
622 depths using different data mining techniques in Baneh region, Iran. Geoderma.
623 https://doi.org/10.1016/j.geoderma.2015.12.003
624 100. Möller M, Volk M, Friedrich K, Lymburner L (2008) Placing soil-genesis and transport processes into a
625 landscape context: A multiscale terrain-analysis approach. J Plant Nutr Soil Sci 171:419–430
626 101. Möller M, Volk M (2015) Effective map scales for soil transport processes and related process domains —
627 Statistical and spatial characterization of their scale-specific inaccuracies. Geoderma 247:151–160
628 102. Möller M, Koschitzki T, Hartmann K-J, Jahn R (2012) Plausibility test of conceptual soil maps using relief
629 parameters. CATENA 88:57–67
630 103. Li X, McCarty GW, Karlen DL, Cambardella CA (2018) Topographic metric predictions of soil
631 redistribution and organic carbon in Iowa cropland fields. CATENA 160:222–232
632 104. Qin C, Zhu A, Yang L, Li B, Workshop TP 7th I, 2006 undefined Topographic wetness index computed
633 using multiple flow direction algorithm and local maximum downslope gradient. researchgate.net

105. (2018) LS factor and soil chemistry - Google Search.
106. Mróz M, Sciences AS-T, 2004 undefined Comparison of several vegetation indices calculated on the basis of a seasonal SPOT XS time series, and their suitability for land cover and agricultural crop. uwm.edu.pl
107. Mróz M, Sobieraj A (2004) 39 Comparison of Several Vegetation Indices Calculated on the Basis COMPARISON OF SEVERAL VEGETATION INDICES CALCULATED ON THE BASIS OF A SEASONAL SPOT XS TIME SERIES, AND THEIR SUITABILITY FOR LAND COVER AND AGRICULTURAL CROP IDENTIFICATION.
108. Thiam A, Eastman JR Chapter Eighteen: Vegetation Indices.
109. Van Khoa P, Hoa NH, Tuan DA USING REMOTE SENSING INDICES TO REDUCE EFFECTS OF HILLSHADE ON LANDSAT 8 IMAGERY.
110. Payero JO, Neale CMU, Wright JL COMPARISON OF ELEVEN VEGETATION INDICES FOR ESTIMATING PLANT HEIGHT OF ALFALFA AND GRASS.
111. Patel NR, Parida BR, Venus V, Saha SK, Dadhwal VK (2012) Analysis of agricultural drought using vegetation temperature condition index (VTCI) from Terra/MODIS satellite data. Environ Monit Assess 184:7153–7163
112. Ahmad F (2012) Spectral vegetation indices performance evaluated for Cholistan Desert. J Geogr Reg Plan 5:165–172
113. Richardson A, remote CW-P engineering and, 1977 undefined Distinguishing vegetation from soil background information. asprs.org
114. Kandwal R, Jeganathan C, Tolpekin V, Kushwaha SPS (2009) Discriminating the invasive species, 'Lantana' using vegetation indices. J Indian Soc Remote Sens 37:275–290
115. Barsi J, Lee K, Kvaran G, Markham B, Pedelty J (2014) The Spectral Response of the Landsat-8 Operational Land Imager. Remote Sens 6:10232–10251
116. Erbek FS, Özkan C, Taberner M (2004) Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. Int J Remote Sens 25:1733–1748
117. LIEROP W van, MacKENZIE AF (1977) SOIL pH MEASUREMENT AND ITS APPLICATION TO ORGANIC SOILS. Can J Soil Sci 57:55–64
118. Nachtergaele F, Van Velthuizen H, Verelst L, et al (2009) Harmonized World Soil Database - Version 1.1.

119. Ben Dor E SOIL SPECTRAL IMAGING: MOVING FROM PROXIMAL SENSING TO SPATIAL QUANTITATIVE DOMAIN.

Figures

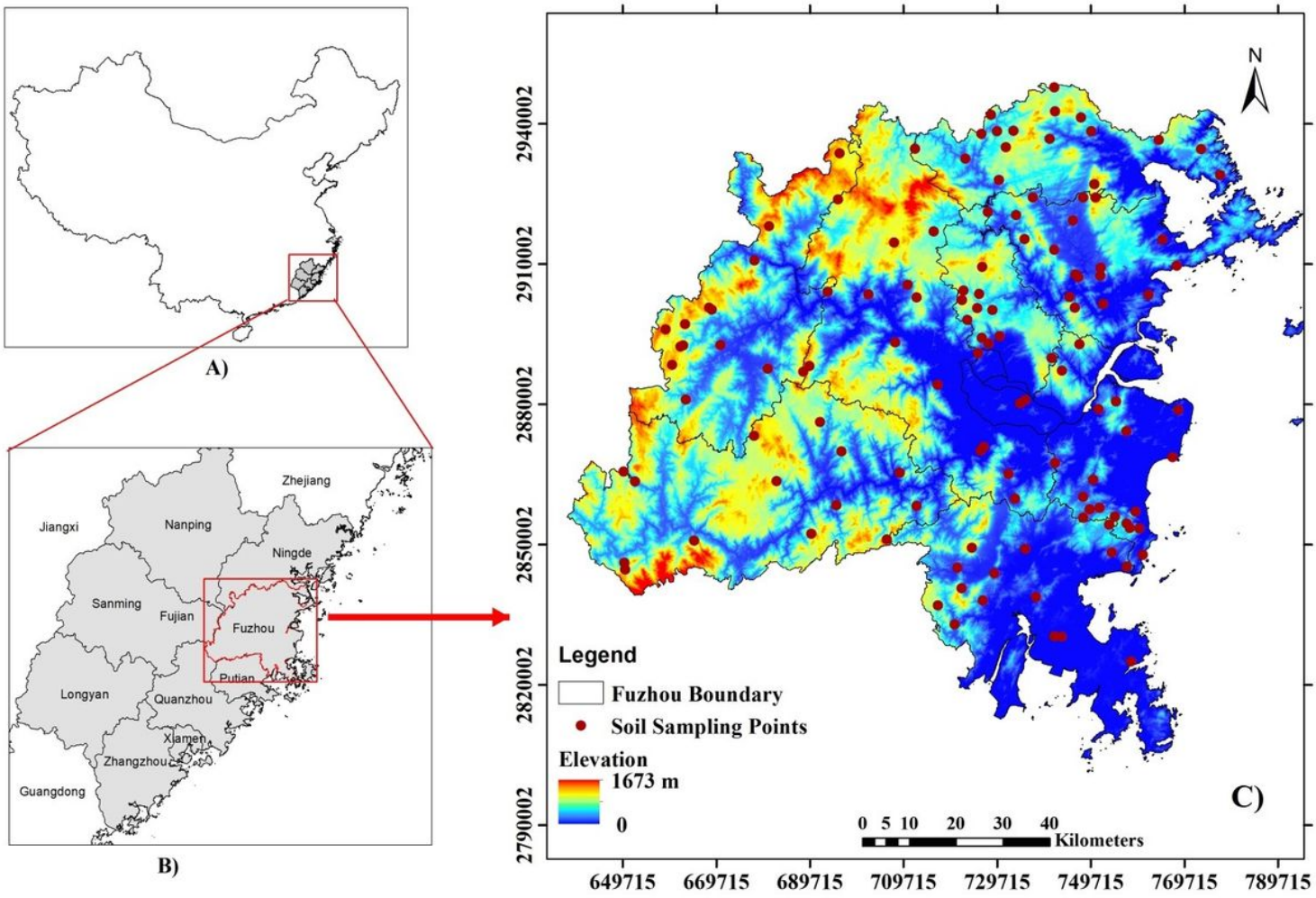

Figure 1

The Study Area Map: A) Location of Fujian Province, China B) Location of Fuzhou C) The study area (sample points) Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

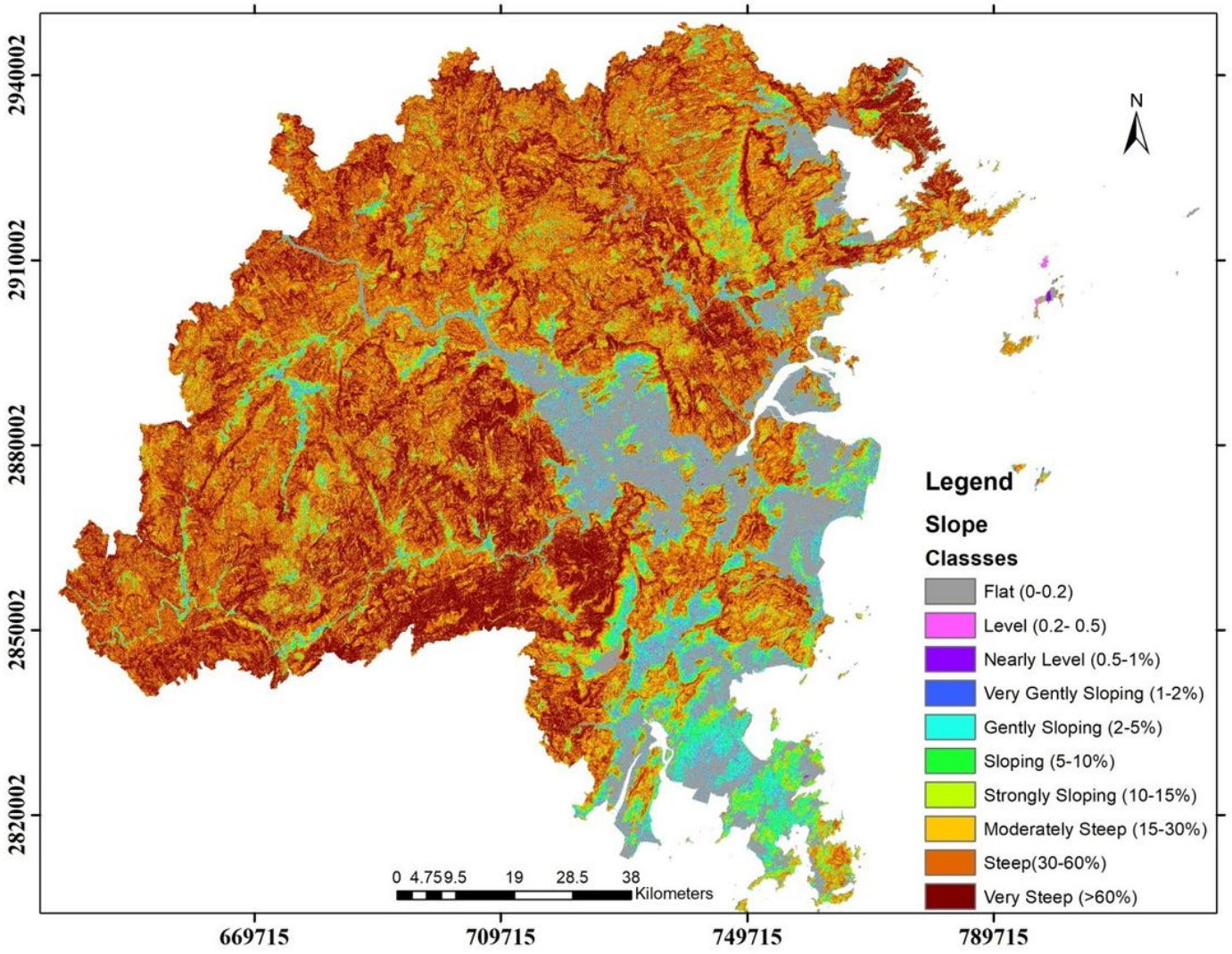

Figure 2

The topographic description map of the study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

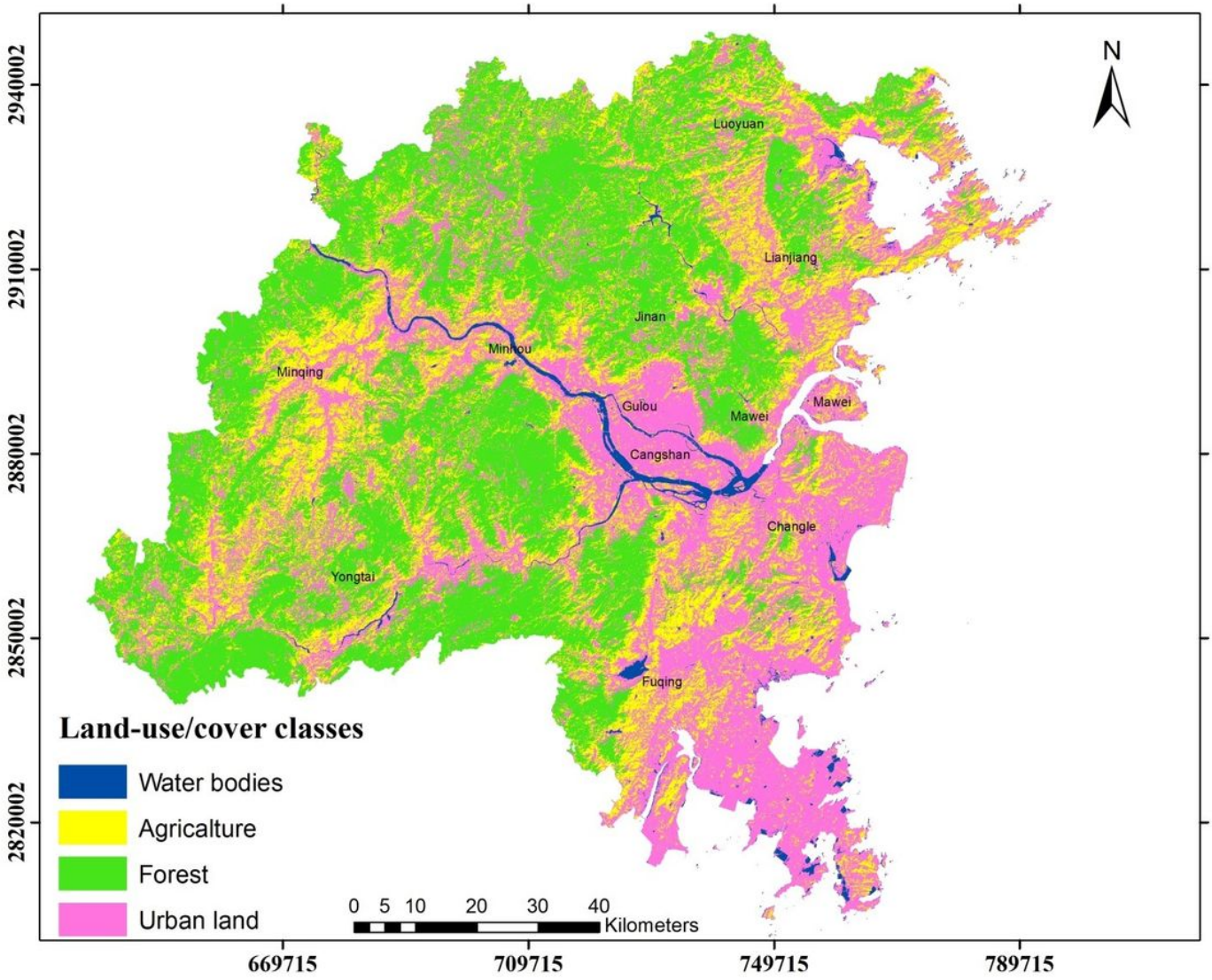

Figure 3

The Land Use/Cover map of the study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

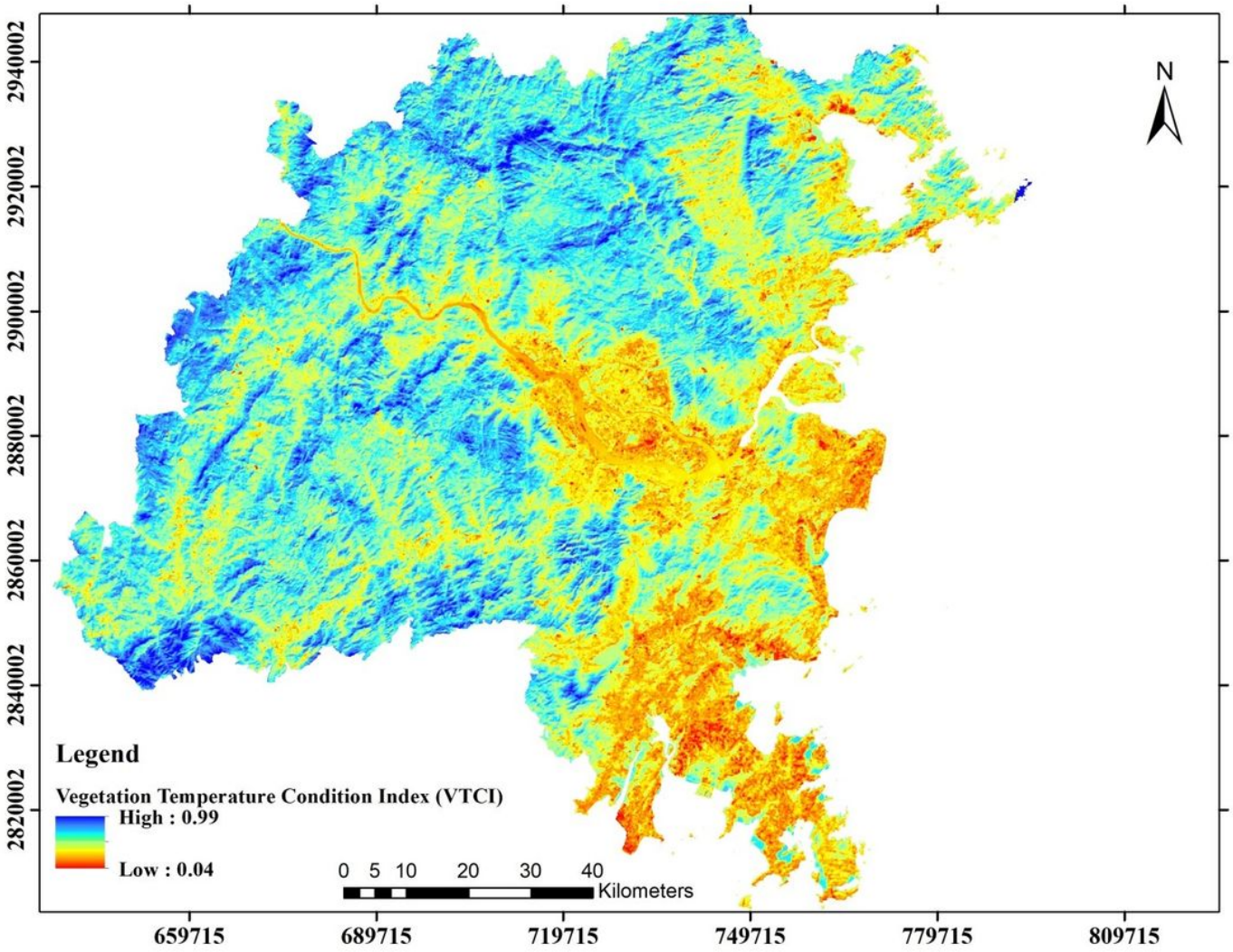

Figure 4

Vegetation Temperature Condition Index (VTCI) map of the study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

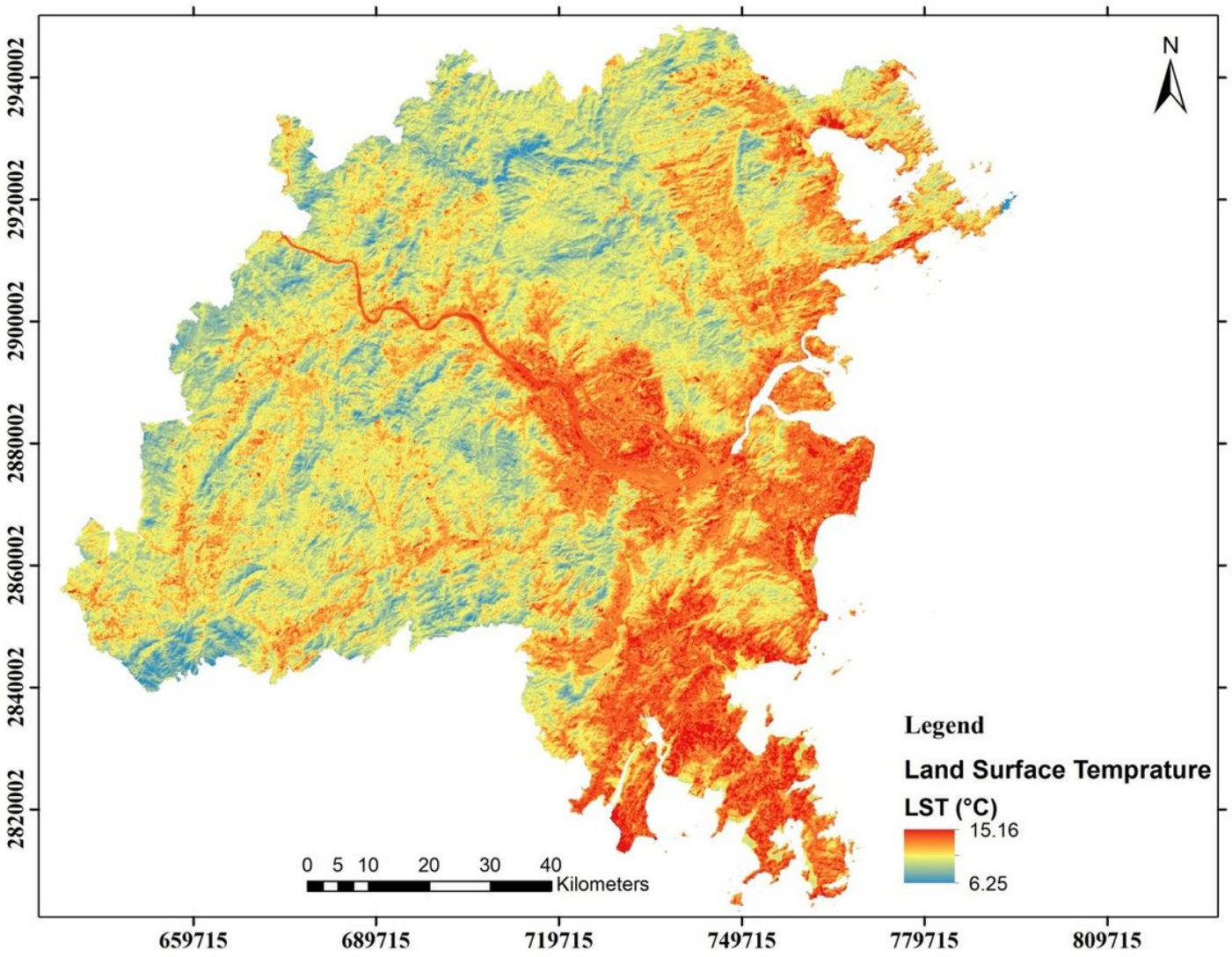

Figure 5

LST map of the study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

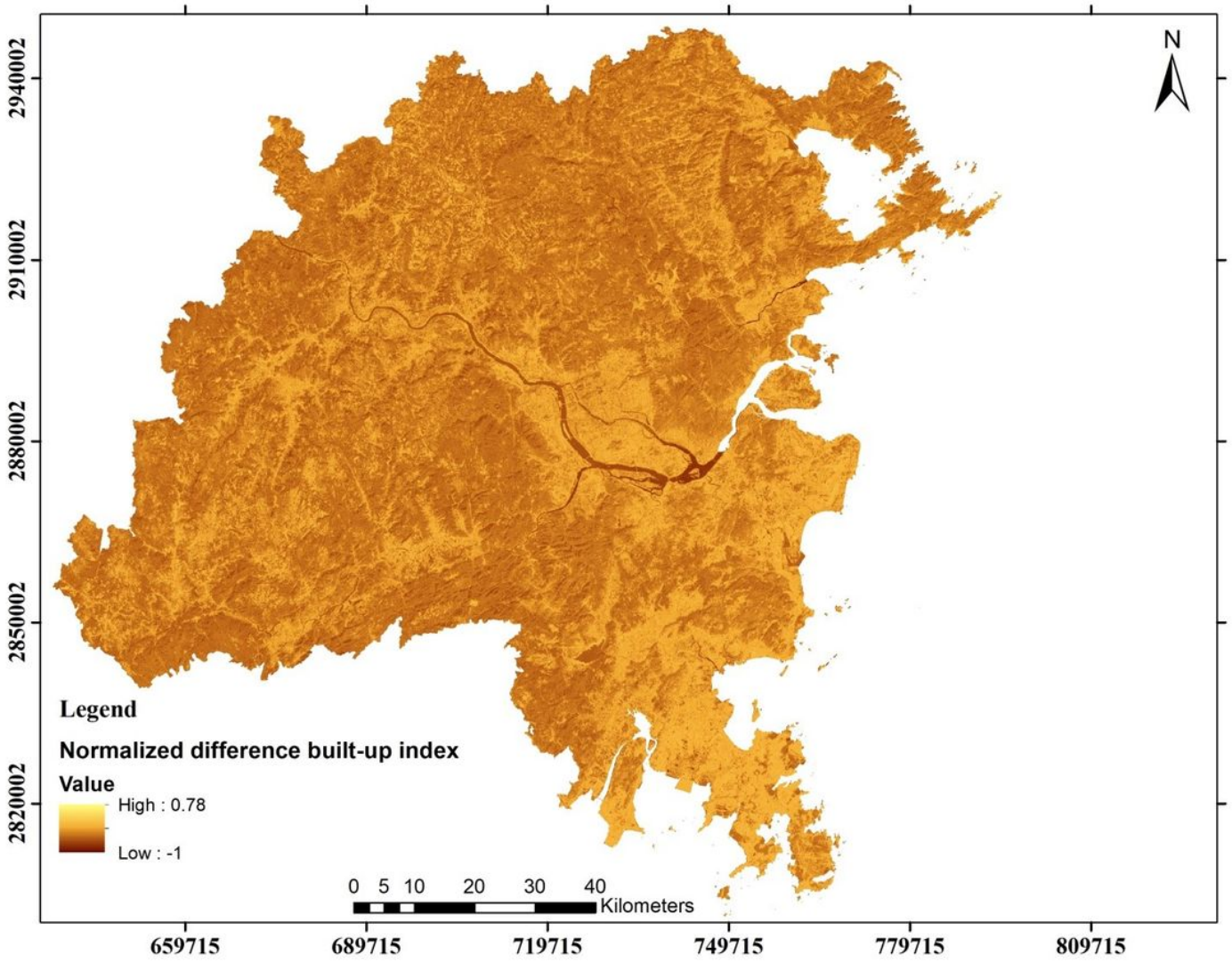

Figure 6

The built-up index (NDBI) map of the study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

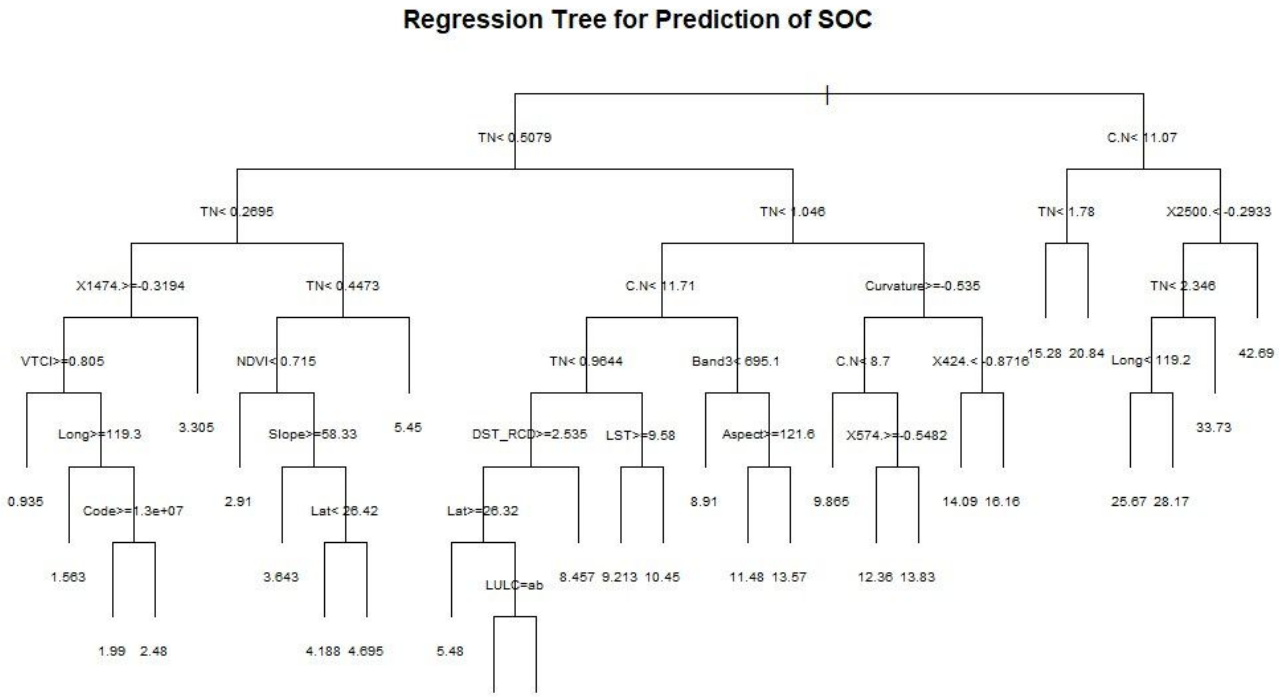

Figure 7

CART tree of bootstrapped samples of data variables used for the SOC predictions

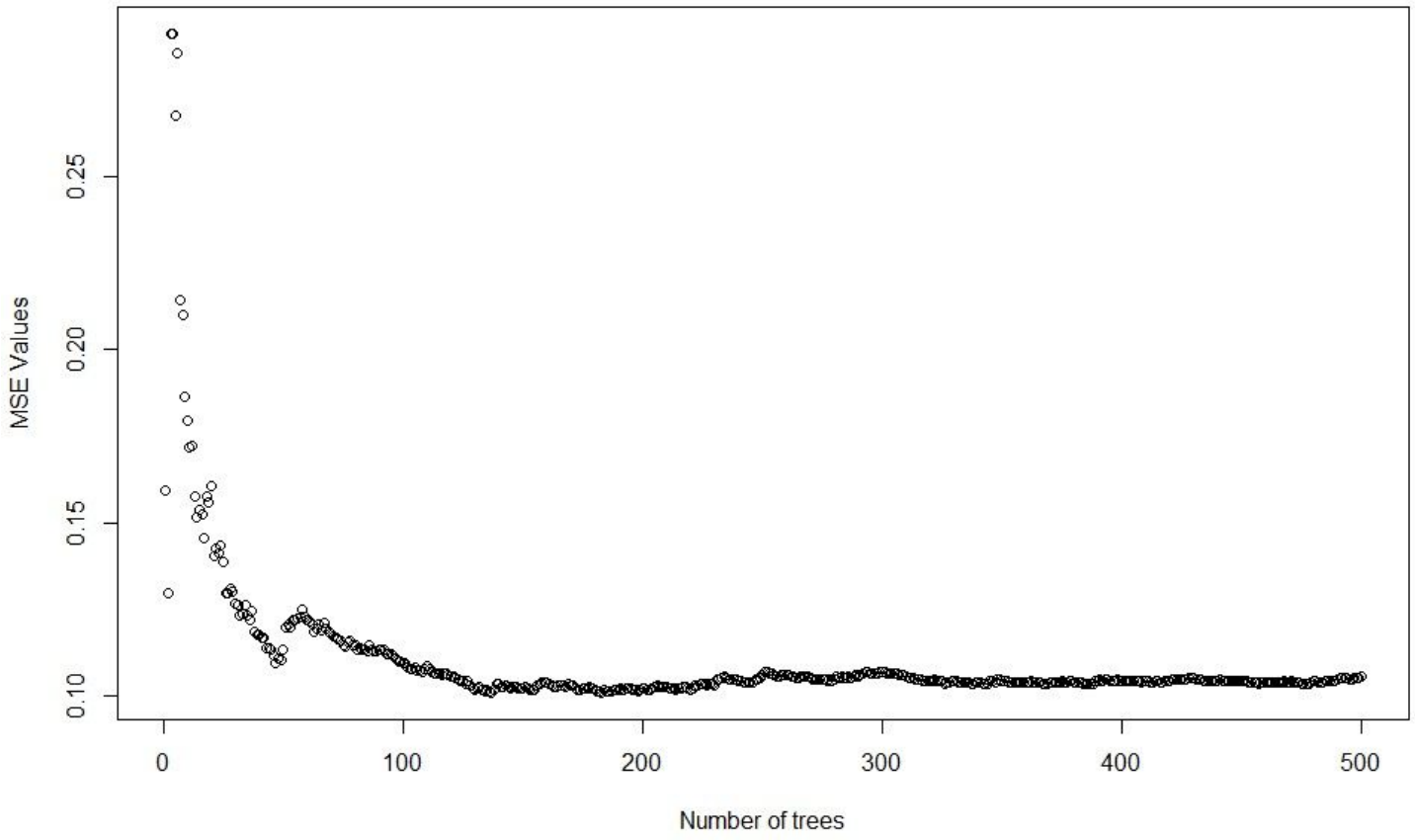

Figure 8

MSE error vs number of trees

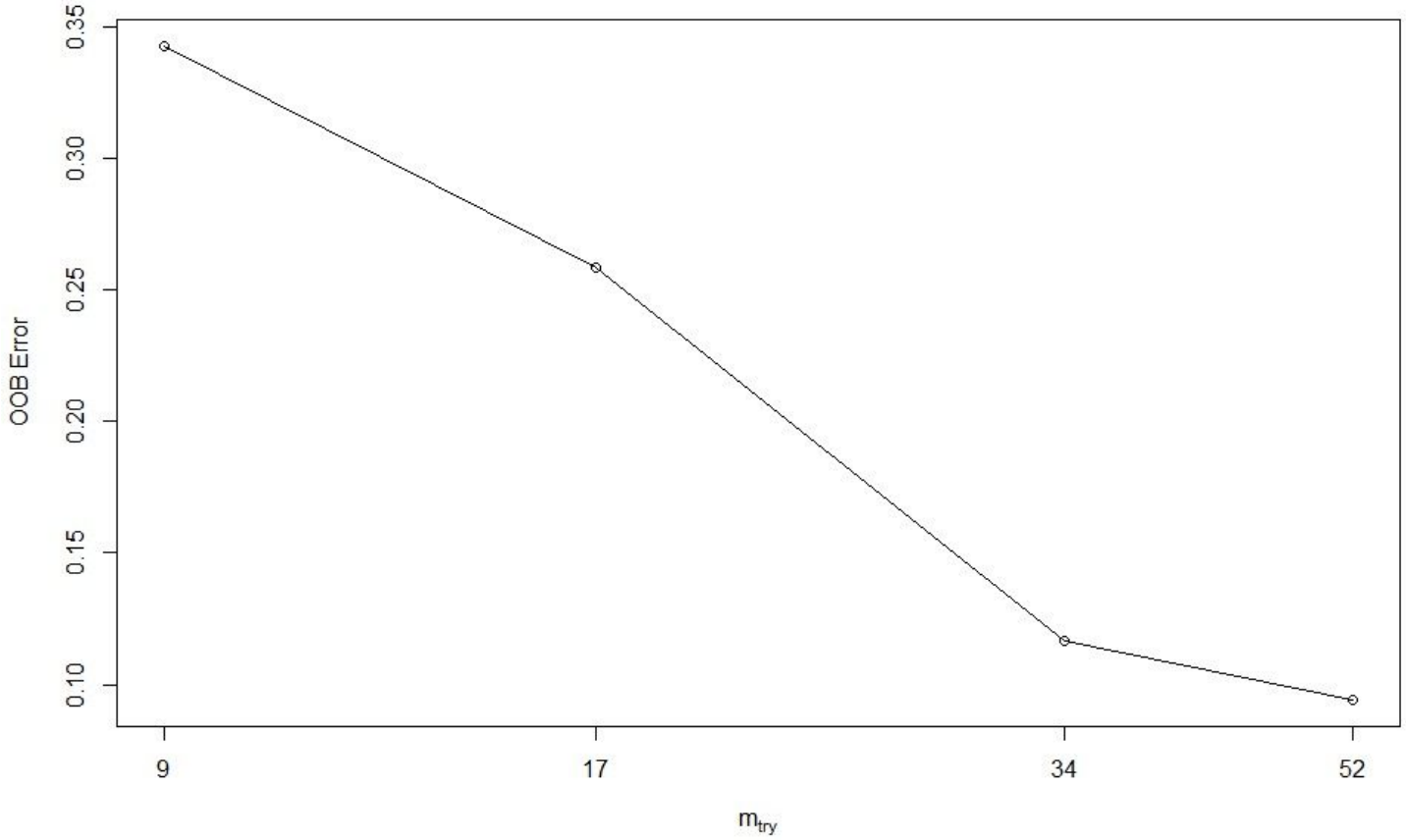

Figure 9

Tuning function as applied on the model for adjusting the mtry parameter

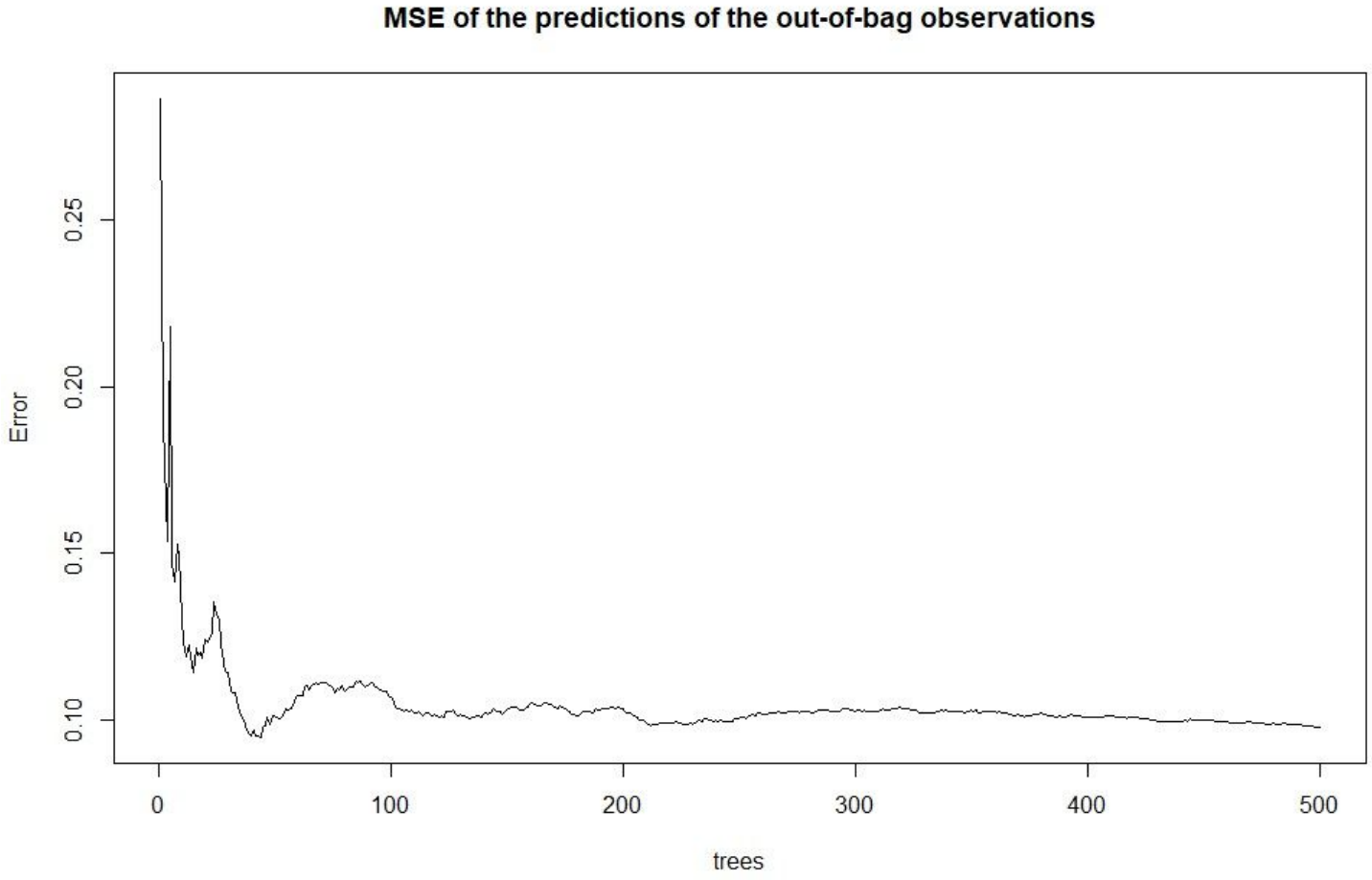

Figure 10

Number of trees adjusted using tuning parameters

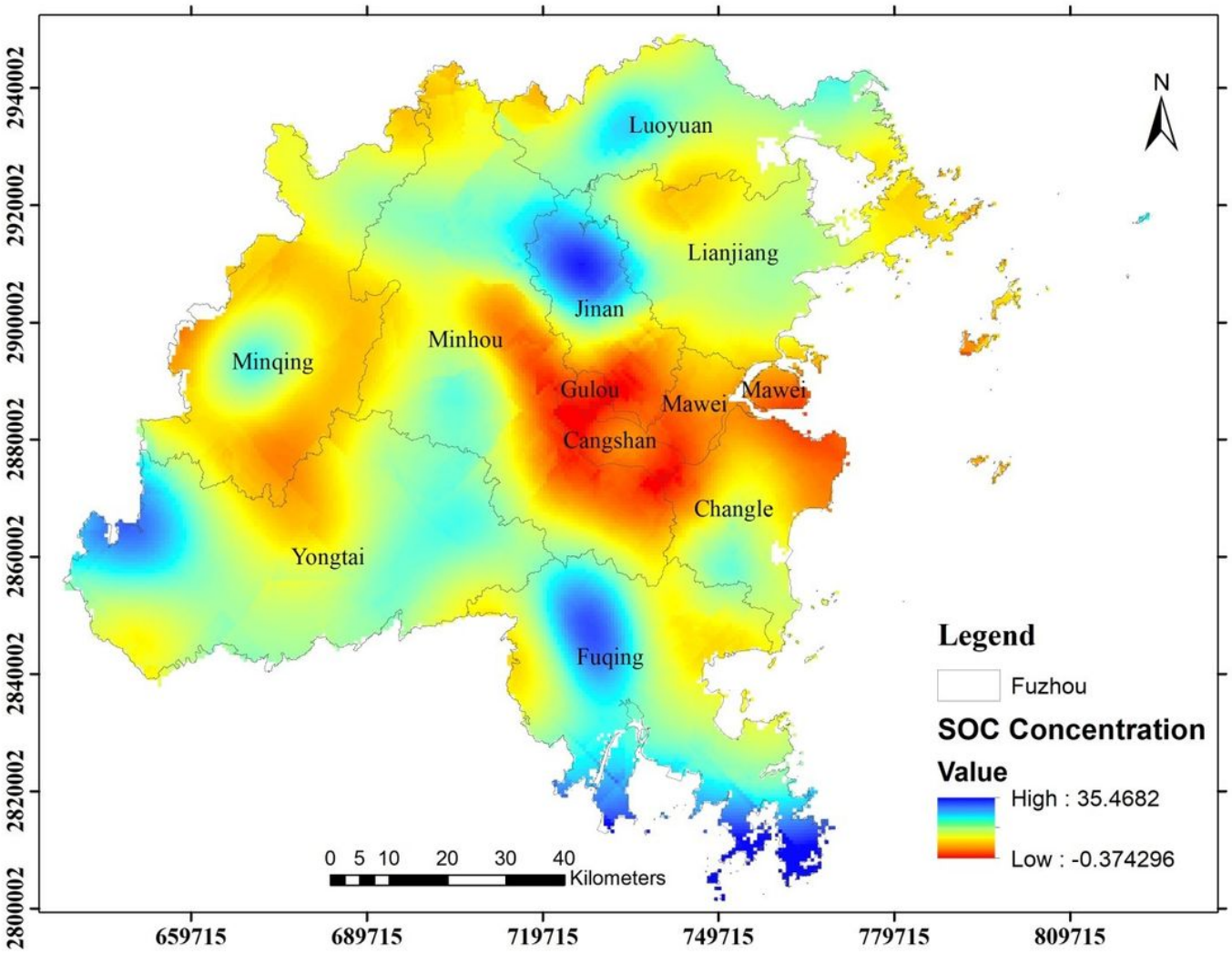

Figure 11

SOC variability map predicted using RF. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

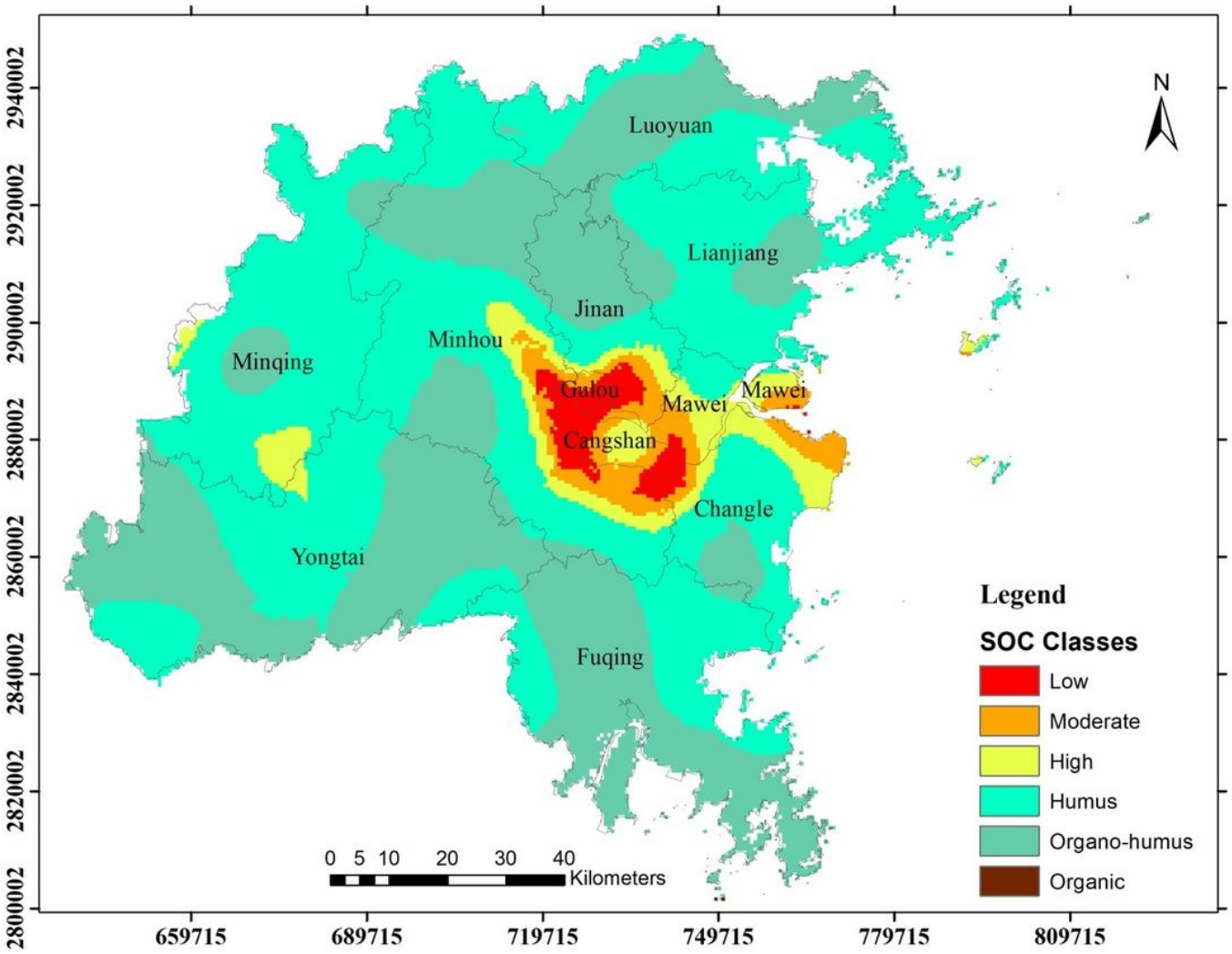

Figure 12

The classes or levels of SOC distribution in the study area. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

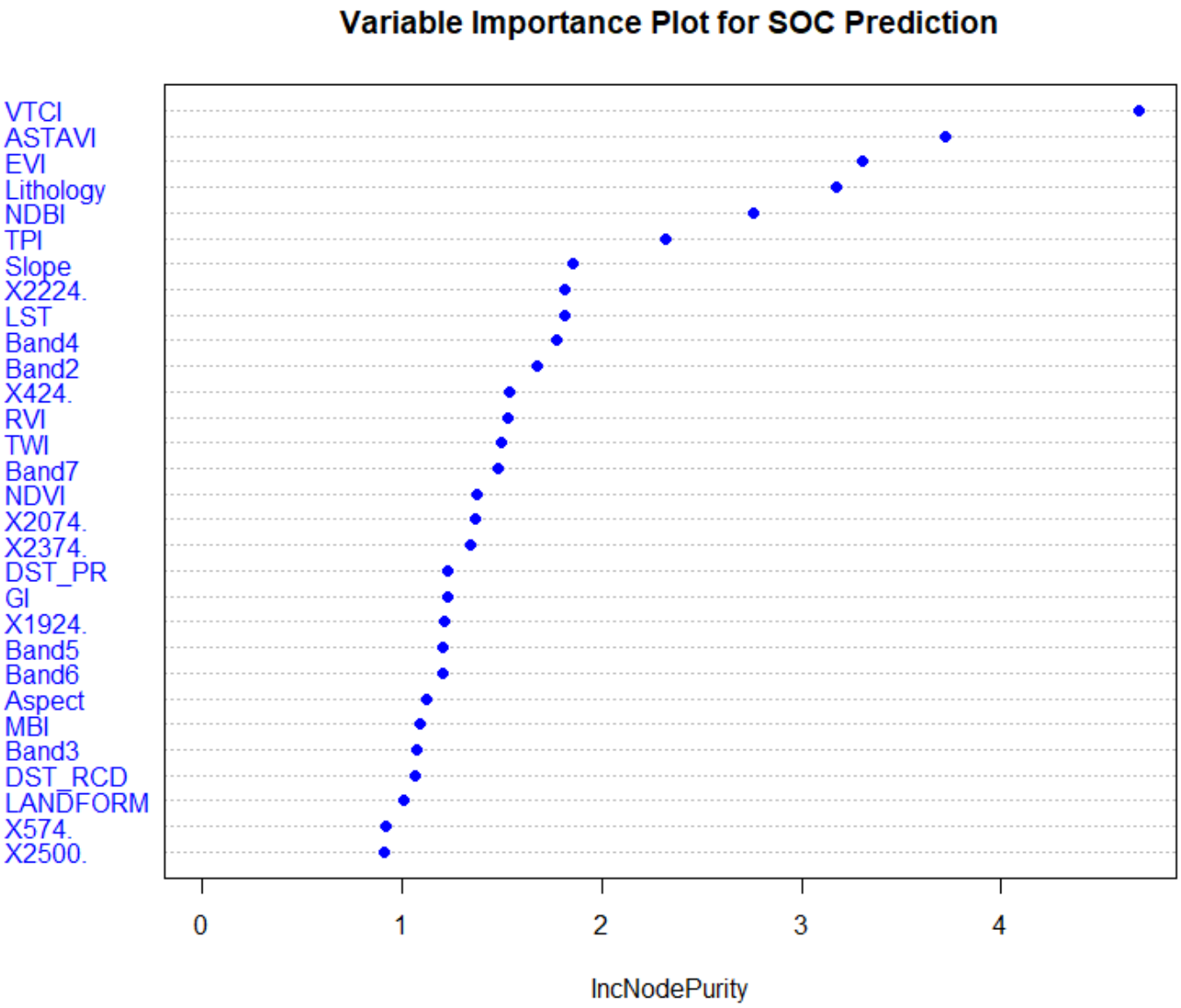

Figure 13

Variable Importance measured as IncNodePurity index

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.pdf