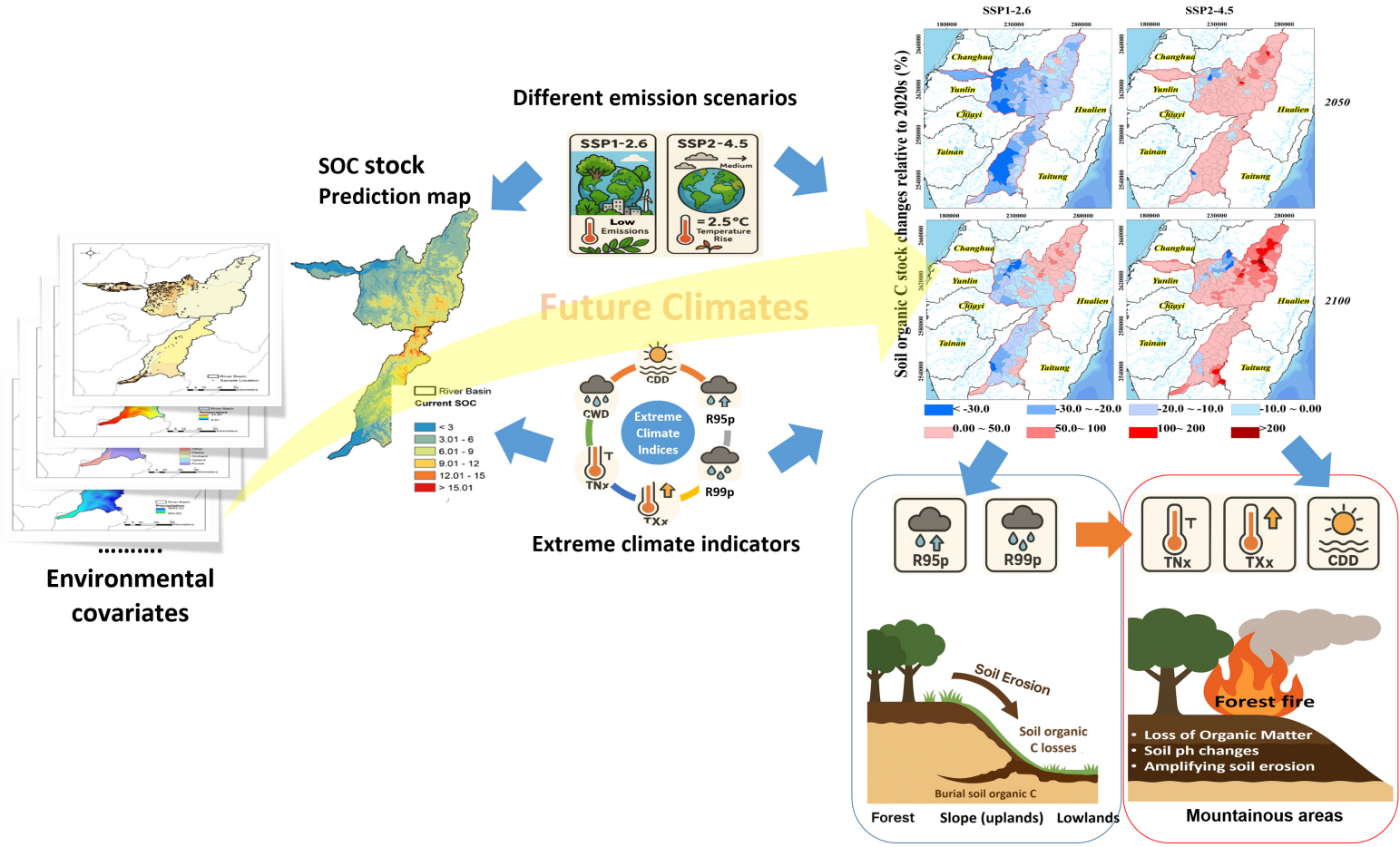
Soil Organic Carbon Projections and Climate Adaptation Strategies across Pacific Rim Agro-ecosystems

Chien-Hui Syu[1], Chun-Chien Yen[1,2], Selly Maisyarah[2], Bo-Jiun Yang[1], Yu-Min Tzou[2], and Shih-Hao Jien[2*]

[1] Agricultural Chemistry Division, Taiwan Agricultural Research Institute, Taichung City 40227, Taiwan, ROC

[2] Department of Soil and Environmental Sciences, National Chung Hsing University, Taichung 40227, Taiwan, ROC

Corresponding author: Shih-Hao Jien; E-mail: shjien@nchu.edu.tw

Abstract

In Pacific Rim regions highly exposed to climate variability, accurate projections of soil organic carbon (SOC) are critical for future effective land management and climate adaptation strategies. This study integrated digital soil mapping with CMIP6-based climate projections to estimate the spatiotemporal distribution of SOC stocks in subtropical (Zhuoshui River) and tropical (Laonong River) watersheds in Taiwan. We collected 901 soil samples and data on 18 environmental covariates and modeled SOC stocks at a 20-m resolution through the Cubist and random forest algorithms, which were also combined with regression kriging. The Cubist-based kriging model was discovered to achieve the highest performance in SOC stock prediction. Mountainous areas were found to contain >80% of SOC stocks, and tropical zones were discovered to store substantially less carbon than subtropical zones. Future emission scenarios revealed spatial heterogeneity in SOC stock dynamics. Under scenario SSP1-2.6, SOC stocks were projected to decline by up to 20.9%, especially in uplands, due to erosion driven by extreme rainfall (R95p, R99p). In contrast, scenario SSP2-4.5 predicted SOC stock increases of 7.9%, particularly in mountainous areas where higher TNx and TXx enhanced productivity. Partial least squares path modeling revealed a climate–topography interaction and explicitly quantified their contributions to SOC stocks, dominated by topography and followed by prolonged dry spells (CDD). This interaction is more pronounced in uplands than in mountainous areas, where topography mitigates temperature extremes and their effects on SOC stocks. Extended CDD may decrease organic inputs by reducing vegetation growth and soil moisture, thereby enhancing carbon losses. Examining the interactions between climatic extremes, landscape types, and SOC stocks is essential for enhancing soil resilience and ensuring stable SOC stocks in the future.

Keywords: soil organic carbon, stock, digital soil mapping, climatic extreme, emission scenario, land type

1. Introduction

Soil organic carbon (SOC) is one of the largest carbon pools in the global carbon cycle (Grace, 2004) and is a key concern of agricultural and environmental policies (Johnston et al., 2004). SOC also has a crucial influence on the carbon cycle at the local and global levels (Singh et al., 2018). Multiple studies have examined whether carbon storage in agricultural soils can offset global warming, and various frameworks have been developed for evaluating the dynamics of SOC. Among these, the landscape scale has enabled researchers to consider the interplay between natural processes, human patterns, and SOC dynamics (Viaud et al., 2010). Therefore, this scale is the most appropriate for assessing environmental and agricultural ecosystems (Li et al., 2021).

Changes in climatic conditions such as temperature, carbon dioxide concentration, and precipitation may influence the dynamics of SOC by affecting the rates of soil processes such as mineralization, decomposition, leaching, and total carbon loss. In areas prone to climatic extremes—such as floods, droughts, and heat waves—these conditions may further affect the dynamics of SOC (Li et al., 2021; Chalchissa et al., 2022). In addition, extreme climate events may strongly affect the content of SOC, with subsequent effects on agricultural productivity and ecosystem services. Therefore, before the dynamics of SOC can be evaluated at the landscape scale in response to climate change, a spatiotemporal technique is required. Zhu and Lin (2010) argued that in areas with major terrain variation and low sampling density, utilizing a non-geostatistical approach or a combination of geostatistical and non-geostatistical approaches can improve prediction ability.

In digital soil mapping (DSM), soil properties in unsampled areas are predicted using statistical or machine learning models that relate soil observations to environmental factors (Grunwald, 2009). Spatial variability is key to model construction (Zhu and Lin, 2010). There are two major approaches including: (1) non-geostatistical methods based on the SCORPAN model (Jenny, 1941; McBratney et al., 2003), such as multiple linear regression (MLR), generalized additive models, Cubist, and random forest (RF) (Quinlan, 1992; Breiman, 2001); and (2) geostatistical methods accounting for spatial autocorrelation, including ordinary, simple, and universal kriging. Machine learning algorithms (e.g., MLR, RF, Cubist) are widely applied for mapping SOC content and stocks (Lamichhane et al., 2019; Siewert, 2018; Rudiyanto et al., 2018). Model performance depends on spatial scale, observation density), and terrain (Zhu and Lin, 2010; Tsui et al., 2016; Keskin and Grunwald, 2018). This study used regression kriging, a hybrid approach combining regression models (e.g., Cubist, RF) with spatial interpolation of

residuals (Ma et al., 2017). Although regression kriging often improves RF and Cubist predictions, its advantage is not universal, highlighting the need for meta-analytical evaluations (Vaysse and Lagacherie, 2015; Ma et al., 2017; Lamichhane et al., 2019).

Taiwan is located on the frontline of the Pacific Rim and is highly prone to the combined effects of climate change and El Niño–Southern Oscillation events. The frequency and spatial variability of extreme climate events in this region are expected to dramatically increase in the future. Therefore, understanding the combined effects of extreme climate variability and long-term climate change on regional climate and SOC variation is essential for evaluating the vulnerability of regional agriculture, water resources, and ecosystems. In Taiwan, the Zhuoshui River Watershed (ZRW) and Laonong River Watershed (LRW) are the two largest and most crucial agricultural eco-watersheds. The ZRW includes fluvial plains and is one of the most essential agricultural areas in Taiwan. In the ZRW, rice, vegetables, and other crops are extensively cultivated. This high agricultural activity underscores the importance of SOC in sustaining soil fertility and agricultural production. In the LRW, SOC plays an essential role in supporting the limited amount of agriculture that is practiced. These two watersheds are located in different climatic zones, which may affect their SOC dynamics. The ZRW is located in central Taiwan and has a subtropical climate, whereas the LRW is located in southern Taiwan and has a tropical monsoon climate. Future climate-change-related changes in temperature and precipitation may substantially affect the content and total stocks of SOC in these two regions, altering their agricultural production and land use patterns.

Therefore, this study applied digital soil mapping approaches to generate high-resolution maps of SOC stock distribution in the surface layer (0–30 cm) of the ZRW and LRW in Taiwan, with the aim of better understanding the spatiotemporal dynamics of SOC under different emission scenarios (SSP1-2.6 and SSP2-4.5) and various extreme climate indicators projected for 2050 and 2100.

2. Materials and Methods

2.1 Research area

The Zhuoshui River watershed (ZRW) is located in central Taiwan. Its basin covers Changhua County, Nantou County, Yunlin County, and Chiayi County. The watershed has an area of 3156.9 km^2 and is located in a subtropical climate zone (Fig. 1a and 1b). This area has an average annual temperature of 8.6–23.6 °C and receives annual cumulative rainfall of 834.6–3693.4 mm. The elevation of the study area ranges from 0 to 3844.2 m above sea level,

108   and it has diverse topography consisting of mountains (2060.7 km^2, 65.2%), hills (843.3 km^2,

109   26.6%), and plains (261.7 km^2, 8.2%). In terms of soil classification, the upstream areas are

110   primarily characterized by stony soils, whereas the western plains are predominantly

111   characterized by silty alluvial soils. The Laonong River watershed (LRW) is located in southern

112   Taiwan, representing an upper mainstream area of the Gaoping River. Its basin covers Nantou

113   County, Kaohsiung City, Pingtung County, and Taitung County. The watershed has an area of

114   2038 km^2 and is located in a tropical climate zone. This area has an average annual temperature

115   of 19.5 °C, and it received annual cumulative rainfall of 3222.6 mm during the period 2011–

116   2020. According to the 2015 Land Cover Survey, the upper reaches of the basin are

117   predominantly mountainous areas (73.3%), whereas the downstream gentle slopes and plains

118   are predominantly agricultural areas (10.6%; Fig. S1).

119

120   ***2.2 Soil samples and analyses***

121   Soil survey data for the period 2012–2020 were obtained from the Taiwan Agricultural

122   Research Institute. A total of 901 topsoil samples (0–30 cm, based on IPCC (2006, 2019a)

123   recommendation) were obtained. Each sample's location was recorded using a handheld global

124   positioning system device. After the samples had been air-dried at room temperature, they were

125   sieved through a 35-mesh screen (0.5 mm sieve opening) and stored in plastic containers. They

126   were then analyzed using the loss-on-ignition (LOI) method (Nelson and Sommers, 1996).

127   Because the LOI method typically overestimates SOC (Li et al., 2021), a correction function

128   was applied to adjust SOC content from LOI values to those obtained using a total organic

129   carbon (TOC) analyzer (solid TOC cube, Elementar). The correction equation is as follows:

130

131   TOC $=0.7084 * $ LOI $- 0.0986$ (R^2 $=0.94$; P <0.001)   [1]

132

133   where TOC and LOI refer to the SOC contents (%) determined by the TOC analyzer and

134   the LOI method, respectively. A sample's total bulk density was determined using the clod

135   method or soil core method (Blake and Hartge 1986). Finally, the soil organic carbon stock

136   (SOC$_{stock}$, kg m^{-2}) was calculated using the following equation:

137

$$SOC_{stock} = TOC * \rho * D/10 \qquad\qquad [2]$$

138

139   where TOC is the SOC content (%), ρ is the bulk density of soil (g cm^{-3}), and D is the soil

depth (cm). Owing to the substantial variability in coarse fragment content, this study excluded them from the calculation of SOC stocks.

2.3 Environmental covariates

Environmental covariates were categorized on the basis of factors pertaining to soil formation, including topographical data such as digital elevation models (DEM), satellite remote sensing imagery, meteorological data, land use survey data, and soil order (Table 1). All environmental covariates were resampled at a spatial resolution of 20 m by using R software version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria).

The DEM was derived from a 20-m grid numerical terrain model established by the Taiwanese Ministry of the Interior. To create an elevation map, the "Fill Sinks" function of SagaGIS 8.0.1 was used to smooth the discontinuities in the model. These elevation data were employed to generate relevant topographical attributes. These attributes included the slope, aspect, terrain ruggedness index (TRI), terrain position index (TPI), topographic wetness index (TWI), multiresolution index of valley bottom flatness (MrVBF), multiresolution ridge top flatness (MrRTF), curvature, flow accumulation, and stream power index (SPI). Ma et al. (2017) argued that topographical parameters can serve as environmental covariates in organic carbon prediction models.

The normalized difference vegetation index (NDVI) was calculated using infrared (b4) and near-infrared (b8) satellite imagery data (Sentinel 2) for the period 2016–2020 to determine the proportion of space covered by vegetation, which was produced by the Google Earth Engine at a resolution of 20 m. Climate is one of the key soil-forming factors and, according to Wiesmeier et al. (2019), a major driver influencing SOC storage. This study used climatic data from 2011 to 2020, including mean annual temperature (MAT) and total annual precipitation (TAP), obtained from Taiwan's Central Weather Bureau. The original spatial resolution of these data was 1 km. To match the spatial scale of other covariates, these raster layers were resampled to 20 m resolution using the resample function from the raster package in R (Hijmans, 2022), with bilinear interpolation (method = "bilinear"). In addition to these factors, land cover type also influences SOC storage (Edmondson et al., 2014). Therefore, this study used a 2015 land cover map produced by TARI, classified into five categories: (1) paddy fields; (2) upland farming (including miscellaneous grains, tea trees, betel nuts, and bamboo); (3) orchards; (4) forests (including plantations, primary forests, and high-mountain arrow bamboo forests); and (5) others (miscellaneous and riverine lands). The flow chart of the methodology is illustrated

in Fig. S2.

2.4 Predictive models

Because of research advancements in the field, the techniques used in DSM have evolved from simple linear models to comprehensive machine learning technologies (Minasny and McBratney, 2016). In this study, two widely used data mining models, namely Cubist and RF models, were employed. Both models were further combined with regression kriging to account for both geographical and non-geographical effects, resulting in the Regression Kriging with Cubist and Regression Kriging with Random Forest models. Their capabilities in predicting the spatial distribution of SOC were compared.

The Cubist model is a rule-based classification algorithm proposed by Quinlan (1992). It was developed on the basis of the M5 tree model. The Cubist model segregates data into several subsets on the basis of "if–then" patterns and identifies linear relationships between the target variables and environmental covariates in each subset. In the present study, we used the "Cubist" package in R software version 4.0.5 for model development (Kuhn and Quinlan, 2022). We adopted the following parameters: (1) rules for data classification based on rule count, (2) extrapolation for determining the model's degree of extrapolation, and (3) committees that generate multiple committee models on the basis of the number of samples to be processed in the model (i.e., the amount of data) to refine the previous prediction and output collective results. We used the "caret" package (Kuhn, 2008) to perform hyperparameter tuning for the Cubist model, testing committee values of 1, 5, 10, 15, 20, and 25. Using 5-fold cross-validation, caret automatically evaluated the performance of each hyperparameter setting based on the root mean square error (RMSE). The tuning results indicated that setting committees = 20 produced the lowest cross-validation RMSE, suggesting that this configuration achieved the highest predictive accuracy. Therefore, committees = 20 was selected as the final model parameter.

The RF model is an ensemble learning algorithm introduced by Breiman (2001). This model reconstructs a data set into multiple new sets with identical sample size through random resampling during model training. For each data set, environmental covariates are randomly selected for constructing classification or regression trees. In the case of continuous variables, the model's predicted value is the average output of all regression trees. In this study, we used the "randomForest" package of R software version 4.0.5 for model development (Liaw and Wiener, 2002). We adopted the following parameters: (1) mtry, which determines the number

of environmental covariates extracted for each new data set in regression tree construction, and (2) ntree, which determines the number of regression trees in the RF. We used the "caret" package to perform hyperparameter tuning for the Random Forest model. The parameter *mtry* was tested with values ranging from 2 to 9. Using 5-fold cross-validation, *caret* automatically evaluated the performance of each hyperparameter setting based on the root mean square error (RMSE). The tuning results indicated that the model achieved the lowest cross-validation RMSE when *mtry* = 7. The number of trees (*ntree*) was kept at the default value of 500, which is generally sufficient to ensure model stability (Peng et al., 2025). Therefore, the final Random Forest model was trained using *mtry* = 7 and *ntree* = 500.

2.5 Model training and validation

Before model development, we employed the "rpart" package in R software version 4.0.5 (Therneau and Atkinson, 2022) to extract 70% of the data as the training data set (calibration set), totaling 600 samples. We used the remaining 301 samples (30%) as the validation data set (validation set) and used it to determine the model's predictive performance. Model performance was evaluated by comparing the predicted values with the observed values in the validation group. Root mean square error (RMSE) and coefficient of determination (R^2) values were used as assessment indicators and calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - p_i)^2}, \qquad [3]$$

$$R^2 = \frac{\sum_{i=1}^{n}(p_i - \mu_o)^2}{\sum_{i=1}^{n}(o_i - \mu_o)^2}, \qquad [4]$$

where p_i and o_i represent the predicted and observed values, respectively, and μ_o represents the mean observed value. Because of the spatial variability of soil properties, effectively quantifying uncertainty in predictive map outputs is crucial. Therefore, in this study, the prediction interval was calculated using 90% quantiles of upper and lower limits of prediction through the bootstrap method (Malone et al., 2014).

2.6 Geographic information and data analyses

In this study, maps were created and spatial statistical analyses were conducted using the geographic information system (GIS) software ArcMap 10.8. Data processing and statistical analyses were conducted using Microsoft Excel 2016 (Microsoft, Redmond, WA, USA) and R

236   software version 4.0.5.

237   Redundancy analysis is a multivariate or multiresponse technique similar to regression. In

238   this study, redundancy analysis was conducted to determine whether extreme climatic indices

239   and SOC stock changes were associated with the matched grid cells (different land types)

240   evaluated in our previous study (Jien et al., 2025). All statistical analyses, unless indicated

241   otherwise, were conducted using SPSS version 18.0 (SPSS, Chicago, IL, USA). A p-value of

242   <0.05 was considered statistically significant. Partial least squares path modeling (PLS-PM)

243   was employed to identify the pathways underlying the study variables, including emission

244   scenarios, extreme climate indices, and SOC stock and land types. A PLS-PM model was

245   constructed using the "innerplot" function of the "plspm" package (Sanchez et al., 2017). The

246   model's quality and performance were evaluated using the goodness-of-fit (GOF) test. Finally,

247   the "ggplot2" package in R software (Wickham, 2016) was used for redundancy analysis and

248   plot generation (Villanueva and Chen, 2019).

249

2.7 Climate data in various emission scenarios and with extreme climate indices

251   Future climate predictors were obtained from CMIP6-based global climate models,

252   including shared socioeconomic pathways (SSPs) established by the Intergovernmental Panel

253   on Climate Change (IPCC). Model MIROC6—developed by the Japan Agency for Marine-

254   Earth Science and Technology, Atmosphere and Ocean Research Institute, University of

255   Tokyo—was selected for predicting SOC stocks from future climate data. The historical and

256   projected extreme climate indices of CMIP6 were employed for different socioeconomic

257   pathways, specifically for scenarios SSP1-2.6 (sustainable development) and SSP2-4.5 (middle

258   of the road). These data were used to examine climate patterns and trends and establish models

259   for predicting the impact of future climate change on various environments.

260   In this study, we identified the following six extreme climatic indices: consecutive dry

261   days (CDD), consecutive wet days (CWD), very wet day precipitation (R95P), extremely wet

262   day precipitation (R99P), minimum value of daily maximum temperature (TNx), and

263   maximum value of daily maximum temperature (TXx). These indicators provided valuable

264   insights into the effects of extreme weather events across the study area. According to

265   Chalchissa and Kuris (2024), the correlation between these indicators and soil health factors

266   may offer a comprehensive understanding of soil health and the potential for carbon

267   sequestration in agricultural systems.

268

3. Results

3.1 Statistical description of SOC stock

The sampling sites of this study are presented in Fig. S1(a). The average topsoil SOC stock across all sampling points was 4.36 kg m^{-2}. In the ZRW, the topsoil SOC stock ranged from 0.19 to 31.8 kg m^{-2}, with an average of 4.51 kg m^{-2}. In the LRW, the topsoil SOC stock ranged from 0.41 to 14.4 kg m^{-2}, with an average of 3.80 kg m^{-2}. Although the data for the LRW were more concentrated, the overall data exhibited positive skewness, with a skewness value of 0.15. Therefore, a natural logarithm transformation was applied to improve model performance by approximating a normal distribution.

3.2 Model performance in SOC stock prediction

This study constructed SOC stock predictive models using the Cubist, RF, and regression kriging with the training data set and environmental covariates. The performance of these models was evaluated using R^2 and RMSE values. Among the evaluated models, the RF model demonstrated the highest predictive performance in the training data set. The distribution of the training data set (calibration set) and validation data set is depicted in Fig. 2. Therefore, we focused on model performance in the prediction of validation data set prior to model selection. In this respect, the performance indicators of the Cubist model were $R^2 = 0.43$ and RMSE = 0.45 kg m^{-2}, while those of the RF model were $R^2 = 0.46$ and RMSE = 0.43. After incorporating regression kriging, the indicators improved to $R^2 = 0.48$ and RMSE = 0.50 for the Cubist model, and remained at $R^2 = 0.46$ and RMSE = 0.43 for the RF model (Fig. 2).

3.3 Importance analysis of environmental covariate

For the variable importance analysis in the RF model, the increase in mean squared error (MSE) was calculated when each covariate was excluded during the random selection process. In the Cubist model, the usage ratios of various environmental covariates were computed. These indicators revealed the key role of environmental covariates in the prediction of SOC stocks (Fig. 3a). In the RF model, covariates that led to an increase in mean squared error (IncMSE) greater than 15% included elevation (23%), soil order (20%), annual mean temperature (19%), and precipitation (15%). In the Cubist model, the primary classification factors were annual mean temperature (45%), soil order (18%), and elevation (17%). The results of the Cubist model indicated that the importance of aspect, curvature, and flow accumulation was relatively low, thus they exhibited either no usage or very low usage

frequency (Fig. 3a). Among the covariates included, more than half of the data incorporated covariates such as elevation (98%), annual mean temperature (62%), NDVI (59%), TRI (54%), K-value (53%), and slope (52%). These results indicated that climatic and topographic factors strongly contributed to model performances. In summary, the RF and Cubist models identified soil order, elevation, and annual mean temperature as the factors representing the influence of soil, topography, and climate, respectively, on the SOC stock in the study areas.

3.4 Predicted map of SOC stock

The predicted spatial distribution of SOC stock is presented in Fig. 3b. According to the statistical analysis of the prediction map, the mean SOC stock in the ZRW and LRW was 5.51 and 6.38 kg m^{-2}, respectively. The first quartile, median, and third quartile were 3.85, 5.31, and 6.88 kg m^{-2}, respectively, for the ZRW and 3.80, 5.96, and 8.85 kg m^{-2}, respectively, for the LRW. A reduction in SOC stock from mountainous areas to lowland areas (plain areas) was found for both watersheds. In addition, high SOC stock (approximately 15 kg m^{-2}) was discovered in southeastern areas in the ZRW and in northeastern areas in the LRW. Lower storage values of <3 kg m^{-2} were found in downstream areas in the LRW and near the estuary in the ZRW. The areas located along the downstream plains of ZRW exhibited low SOC stock (<2.5 kg m^{-2}) near the river, with SOC stock higher farther from the river.

3.5 Uncertainty analysis for predictive models

For each sampling point in the training data set, prediction residuals were established through leave-one-out cross-validation for the regression kriging and Cubist models. The study area was then classified by landscape in accordance with the classification rules of the Cubist model (Table 2). Fig. 4 presents 90% confidence interval maps drawn using data segmentation and cross-validation techniques. These prediction limit intervals can be regarded as indicators of the model's uncertainty. In the downstream areas of the study region, the confidence interval widths were generally below 6 kg m^{-2}, whereas in the mountainous regions they were substantially higher, with some areas reaching up to 40 kg m^{-2}.

3.6 SOC stock distribution with various landscape types and land uses

The SOC stock spatial distribution was categorized on the basis of topography to demonstrate the distribution of SOC stocks under various landscape types. As shown in Fig. 5, in the lowlands of the ZRW, the lowest average SOC stock was identified in upland farming

areas (1.93 kg m^{-2}), whereas the highest average SOC stock was identified in paddy fields (3.08

kg m^{-2}). In the lowlands of the LRW, the lowest average SOC stock was identified in "other"

land cover types (1.89 kg m^{-2}), whereas the highest average SOC stock was identified in forest

areas (3.17 kg m^{-2}). In the uplands of both catchments, the lowest cover was identified in paddy

fields (3.01 and 2.16 kg m^{-2}), whereas the highest cover was identified in forests (4.22 and 3.6

kg m^{-2}). In terms of land cover, the highest SOC stock was identified in forest areas, with 6.54

kg m^{-2} in the ZRW and 8.02 kg m^{-2} in the LRW, whereas the lowest SOC stock was identified

in orchard areas, with 4.82 kg m^{-2} in the ZRW and 5.60 kg m^{-2} in the LRW.

3.7 Extreme climate index parameter estimates in two emission scenarios

Extreme climate indices in two SSPs were compared: SSP1-2.6 (sustainable development) and SSP2-4.5 (middle of the road). Projections were evaluated for mid-century (2050) and end-century (2100) time points at units of sub-catchment for each watershed. These units were classified as total area, lowlands, uplands, and mountainous areas, denoted as T, L, U, and M, respectively, in Tables S1 and S2.

In all emission scenarios, major spatial heterogeneity and temporal increases were found in SOC stocks (Table 3, Figs. 6 and 7). These findings underscore the importance of modifying the management practices of land use in the future, especially if climate change is severe. In mountainous areas in both watersheds, significant SOC accumulation was predicted. By contrast, lowland agricultural zones are expected to maintain relatively low SOC stocks (<9 Mg C ha^{-1}), with minor gains across scenarios. Scenario SSP2-4.5 was found to result in the higher projected increase in SOC stocks as a result of elevated CO_2 and potential biomass input, although spatial disparities are expected to increase, particularly in erosion-prone or intensively cultivated lands (Fig. S3).

The CDD, CWD, R95p/R99p, and TNx/TXx were analyzed as extreme climate indices. In both SSPs, the increases in temperature- and precipitation-related extremes in the two watersheds were significant. In scenario SSP2-4.5, the magnitude and spatial heterogeneity of these changes were predicted to intensify toward 2100 compared to the 2020 baseline (Fig. 6a). In the ZRW, scenario SSP2-4.5 was predicted to result in a 36% increase in CDD, especially by 2050 in mountainous areas (Table S1). For scenario SSP1-2.6, the CDD was predicted to decrease slightly by 26.3% and 35.2%, particularly in lowlands and uplands. For both watersheds, CWD was predicted to increase in the emission scenarios for 2050 and 2100, but different trends were discovered for the ZRW and LRW. Regarding the ZRW, CWD was

predicted to significantly increase in lowlands and decrease in mountainous areas. However, in the LRW, it was predicted to increase only in uplands (Fig. 6a). These results indicated the polarization of wet–dry periods. A major increase in rainfall extremes was predicted, with R95p increasing by 1785 mm and 1889 mm in the total and mountainous areas, respectively, under the scenario SSP2-4.5 by 2100 in the ZRW. For the same scenario and time frame, R99p was predicted to reach 3535 mm in total areas, with uplands and mountainous areas receiving rainfall of 3625 and 3588 mm, respectively. Temperature extremes were also predicted to increase, where the TXx in ZRW was predicted to increase by up to 22.7% in mountainous areas (scenario SSP1-2.6 and SSP2-4.5 for the year 2100), whereas TNx was predicted to increase by 53.7% in mountainous areas under the scenario SSP2-4.5 by 2100, indicating pronounced warming.

Regarding the LRW, which is characterized by a tropical monsoon climate, high warming and precipitation extremes were predicted, particularly in uplands and mountainous areas, indicating climate spatial heterogeneity. For scenario SSP1-2.6, an approximately 22.2% CDD increase for the entire watershed was predicted by 2100 (Table S2), with the largest increase predicted for lowland areas (52.9%). In the mountainous areas, R95p and R99p under scenario SSP2-4.5 were predicted to reach 1142 mm and 3525 mm, respectively, by 2100. Moreover, CWD was predicted to increase in uplands by up to 85.7%, suggesting prolonged wet conditions. Notably, under the same scenario and time, TXx and TNx were predicted to substantially increase in uplands and mountainous areas, reaching up to 26.8% and 62.6%, respectively, emphasizing the intensification of heat extremes. Overall, this scenario may present another type of threat: long-term droughts with torrential downpours and extreme heat.

3.8 Relationships between extreme climate indices and SOC stocks

Principal component analysis (PCA) was conducted to examine the relationships between extreme climate indicators and SOC stock and to determine their topographic distribution characteristics for two scenarios. Regarding scenario SSP1-2.6, SOC stock variation exhibited negative correlations with R95p and R99p, indicating that extreme precipitation may be detrimental to the maintenance of SOC stocks (Fig. 8a). Pearson's correlation analysis revealed significant negative correlations of SOC stocks with R95p ($r = -0.32$, $p < 0.05$) and R99p ($r = -0.29$, $p < 0.01$; Fig. 8b). Regarding scenario SSP2-4.5, SOC stock variation exhibited a positive correlation with R95p and R99p with $r = 0.17$ and $r = 0.18$, ($p < 0.05$), respectively (Fig. 8c and 8d), indicating that stable wet conditions may promote

SOC accumulation under moderate emission conditions. For all scenarios, strong positive R95p–R99p and TXx–TNx correlations were found (e.g., $r = 0.93$ between R95p and R99p in scenario SSP2-4.5). Taken together, these findings suggest uniform increases in the frequencies of extreme rainfall and extreme heat events in terms of both spatial distribution and climatic mechanisms.

Our results revealed clear topographic effects, with distinct spatial variations in SOC stock dynamics predicted for the different emission scenarios. For scenario SSP1-2.6, SOC stocks were projected to decrease by 3% to 21%, whereas for scenarios SSP2-4.5, SOC stocks were projected to increase by 7.91% to 58.3%, particularly in mountainous areas (because of their enhanced net primary productivity). In addition, the variance and coefficient of variation (CV) of SOC stock percentages show a significant increase in interquartile range and CV over time under SSP1-2.6 and SSP2-4.5 emission scenarios relative to the 2020 baseline. For both emission scenarios, the variance and coefficient of variation of the SOC stock percentage distribution were predicted to increase (Table 3). Collectively, these results indicate that future climatic extremes are projected to significantly increase the spatial heterogeneity of the percentage distribution of SOC stocks. In the majority of scenarios, uplands and mountainous areas are expected to exhibit a drastic response to extreme climate indicators, including major increases in CDD, R99p, TXx, and TNx. In uplands, SOC stocks are expected to respond strongly to the extreme climate, suggesting the susceptibility of slope soils to extreme rainfall and thermal destabilization. Furthermore, more extreme values are predicted for the LRW than for the ZRW, which is likely attributable to the topographic elevation distribution and baseline tropical monsoon climate in the LRW.

In our PLS-PM analysis, we discovered a goodness of fit (GOF) value ranging from 43.0 to 45.7, indicating the high explanatory power of our findings (Fig. 9). The PLS-PM results also revealed distinct differences in the controls for SOC stocks between the ZRW and LRW. For the ZRW (GOF = 45.7%), topographic variables (elevation and slope gradient) were found to have the strongest positive total effect on SOC stocks (standardized total effect = 0.579), followed by consecutive dry and wet periods (CDD and CWD, total effect = 0.238). Despite these findings, extreme rainfall events and temperature did not appear to have a direct effect on SOC stocks. For the LRW (GOF = 43.0%), stronger topographic control of SOC stocks was discovered (total effect = 0.753), with the direct path being positive (0.84, $p < 0.01$), indicating that the SOC accumulation patterns observed predicted this watershed will be closely linked to its land type. Regarding scenarios SSP2-4.5, most of the predicted increases in SOC stocks are

434 concentrated in mountainous areas. Despite these results, the effects of temperature extremes

435 on SOC stocks will presumably be weakened by topographic and hydrological stability,

436 particularly in the LRW. For both watersheds, prolonged dry spells were predicted to indirectly

437 increase SOC stocks, whereas rainfall extremes were predicted to reduce SOC stocks,

438 particularly in uplands (slope lands). Taken together, these findings underscore the importance

439 of incorporating topographic and extreme climate variables into SOC modeling and climate-

440 resilient soil management strategies.

441

442 **4. Discussion**

443 ***4.1 Ability of machine learning models to predict SOC stocks***

444   Lamichhane et al. (2019) argued that several predictive models can be used to predict the

445 spatial distribution of SOC. In the present study, common machine learning models—the

446 Cubist, RF, and regression kriging models—were used to predict SOC stock in the sampling

447 area. Shaik and Srinivasan (2019) highlighted the likelihood of overfitting in RF models,

448 indicating that while the model may predict training data accurately, it may perform poorly

449 when applied to unseen data outside the training set. Therefore, in this study, some of the

450 collected data were randomly selected for validation (Fig. 2). The results indicated that

451 integrating regression kriging into the Cubist model yielded the highest predictive performance

452 ($R^2 = 0.48$, RMSE = 0.50), which was significantly higher than that achieved by the Cubist

453 model alone ($R^2 = 0.43$, RMSE = 0.45). However, incorporation of regression kriging into the

454 RF model resulted in limited improvements, consistent with the results of Vaysse and

455 Lagacherie (2015). This limited improvement may be attributable to the inherently low

456 residuals in the RF model, indicating that even standardizing these residuals and adding them

457 to the model would have minimal effects on predictions. As shown in Fig. 4, SOC stocks were

458 underestimated for several samples from mountainous regions with high organic carbon stock

459 due to plant residues, suggesting that this type of variance was not captured by the models or

460 residuals. Other studies have demonstrated various predictive disparities. These studies include

461 those conducted by Lacoste et al. (2014), who applied a Cubist model to predict organic carbon

462 stock in a France-based study ($R^2 = 0.12$, RMSE = 12.64); Adhikari et al. (2014), who

463 employed regression kriging in a Denmark-based study ($R^2 = 0.41$, RMSE = 0.24); and Ma et

464 al. (2017), who combined regression kriging with a Cubist model in a China-based study ($R^2 =$

465 0.25, RMSE = 0.12). Although the models used in the present study exhibited several predictive

466 disparities, they nonetheless exhibited high reliability in terms of their overall predictive ability.

467   In SOC stock forecasting models, empirical estimations of uncertainty involve geographic
468   spatial segmentation. In Cubist models, input data are divided into groups on the basis of a
469   series of rule-based classifications. Therefore, examining the empirical distribution of the
470   regression kriging residuals in each category is appropriate (Malone et al., 2014). According to
471   classification rules, if low-elevation areas such as plains, foothills, and valleys involve young,
472   weakly developed soils or miscellaneous lands, these areas are considered to exhibit a wide
473   distribution of residuals. By contrast, other soil order categories are considered to exhibit a
474   more concentrated distribution of residuals. In mountainous regions, classification is based on
475   the NDVI, where a higher NDVI indicates more vegetation, whereas a low NDVI suggests the
476   presence of water bodies or bare soils. In this study, analysis of 90% confidence interval maps
477   (Fig. 4) revealed that for low-altitude areas with high sampling density, abundant data were
478   available for model construction, leading to generally low prediction residuals. Even with
479   various landscape classifications, the prediction limit intervals were low. By contrast, in
480   mountainous areas, the existence of few samples and substantial variability in environmental
481   covariates increased the difficulty of prediction. Generally, when establishing empirical
482   divisions with insufficient data, outliers can easily influence the residual distribution in a given
483   category. Therefore, future sample planning in these areas can be guided by such map data.

484

485   *4.2 Effects of environmental covariates on SOC stocks*

486   In this study, SOC stocks were influenced by the following key topographical attributes:
487   elevation, the multiresolution index of valley bottom flatness, slope, and the topographic
488   wetness index (Fig. 3a). As reported by Mishra and Riley (2015), who conducted an Alaska-
489   based study, elevation is a crucial predictor of SOC stock, regardless of resolutions. According
490   to Adhikari et al. (2014), the multiresolution index of valley bottom flatness (MrVBF) and the
491   topographic wetness index (TWI) are important covariates, ranking just below precipitation.
492   The MrVBF is used to identify flat valley bottoms and thereby indicate potential areas of
493   erosion or deposition, whereas the TWI is used to indicate terrain's control over soil moisture,
494   reflecting wet or dry conditions (Lamichhane et al., 2019). Slope affects SOC stocks by
495   influencing solar radiation and moisture retention. Regarding meteorological covariates,
496   annual cumulative precipitation and mean annual temperature are crucial in determining SOC
497   stock. Gray et al. (2015) identified positive correlations of the SOC content of topsoil in New
498   South Wales, Australia, with precipitation and relative humidity. According to Lamichhane et
499   al. (2019), high precipitation may enhance vegetation growth or create anoxic conditions that

slow soil carbon oxidation. Rial et al. (2017) reported a negative correlation between temperature and SOC content in Europe, with higher altitude and latitude found to correspond to slower SOC decomposition. In the present study, the effect of elevation was attributable to temperature, particularly because the original resolution of the temperature data was 1 km, and high-resolution elevation data were necessary to obtain a detailed spatial distribution. Of the two biotic factors used in this study, the NDVI was the primary environmental covariate. Wang et al. (2018) highlighted the role of the NDVI as an indicator of vegetation cover, which is strongly correlated with SOC. They emphasized that convincing results could be obtained if long-term remote sensing data could be obtained to calculate NDVI values for multiple time periods.

4.3 Influence of land cover on the spatial distribution of SOC stocks

In this study, a trend of increasing SOC stock with increasing elevation was discovered, and this effect is likely driven by elevation and temperature (Fig. 3b, Fig. S3). The SOC stock in differing land coverage types was found to exhibit variation for different terrains. For instance, for lowlands, SOC stocks are predicted to rapidly decrease as a result of intensive cultivation, leading to low SOC stocks in agricultural production lands (Fig. 3b). Regarding the LRW, which has a tropical climate, SOC stock in farmlands was considerably lower than that in forests. This finding may be attributable to the farmlands in the LRW being frequently tilled for triple cropping or the mean annual temperature in the area being higher than that on the Zhuoshui River plains. In uplands, the rice fields in the two basins had SOC stock levels similar to those observed for lowlands, although an increase was discovered for orchard and forests. In mountainous areas, the SOC stock predictions were higher than those for lowlands and uplands across all types of land cover. According to the literature, the eastern region of the LRW experiences high precipitation and low temperatures, which result in higher organic carbon storage than that observed in the ZRW, particularly in mountainous areas (Fig. 3b; Guo et al., 2019). Furthermore, forests cover the majority of the landscape in mountainous areas and represent the primary reservoirs of SOC stocks.

4.4. Adaptation strategies for the management of SOC stocks in various emission scenarios

In this study, future climatic variables calculated from global climate models (GCMs) were used as inputs to estimate the spatiotemporal variation in global topsoil organic carbon stocks in 2020, 2050, and 2100. All GCM were obtained from the Coupled Model Intercomparison

Project. CMIP6 was specifically selected for estimating future topsoil organic carbon stocks. Tables S1 and S2 list the extreme climate indicators involved in the considered scenarios (scenarios SSP1-2.6 and SSP2-4.5) on the basis of CMIP6 data.

Our results indicated increasingly pronounced spatial heterogeneity in SOC stocks in the scenario involving severe greenhouse gas emissions (Table 3, Fig. 6). Severe warming in the future will cause increasingly regional climate variability, resulting in greater spatial heterogeneity in SOC stocks. This phenomenon will be particularly evident in countries with complex topography, such as those located on the frontline of the Pacific Rim, which is directly exposed to the threats of rapid climate change. Generally, the combination of steep terrain and intricate terrain complicates management of the SOC distribution because spatial variability driven by extreme climatic events is difficult to predict and control.

According to our results, extreme climate and land type are the most crucial determinants of SOC stocks in regions near the Pacific Rim (Dialynas et al., 2016; Wei et al., 2024; Chen et al., 2024). As shown in Figs. 6 and 7c, in scenario SSP1-2.6, SOC stocks are projected to be depleted by 2050 (-21%) and 2100 (-3.75%), with this depletion most severe in lowlands and uplands (slope lands). A significant increase in R95p, R99p, or CWD may increase soil erosion, leading to major losses in SOC stocks. Intense rainfall events may also cause topsoil erosion and the leaching of dissolved organic carbon, and episodic carbon export may exceed respiratory losses (Olaya-Abril et al., 2017; Rillig et al., 2021). In certain mountainous areas, localized SOC gains are predicted, even for the low emission projections, which may be related to spatially uneven warming and rainfall, leading to enhanced vegetation productivity and underground carbon input (Fig. 6; Guo et al., 2019). These findings are consistent with the region-specific SOC responses to temperature and precipitation anomalies in previous modeling studies (Wang et al., 2023).

In scenario SSP2-4.5, which involves CO_2 emissions that approach the current levels until the mid-century time point before declining but do not reach net zero by 2100, the SOC stocks are predicted to be controlled by R95p, R99p, and TXx (not statistically significant). For all study areas, slight warming and increased extreme rainfall events (smallest increase among all emission scenarios) will facilitate vegetation growth, which will in turn increase SOC stocks. Despite these findings, losses in SOC stocks are still predicted to occur in certain upland regions (Fig. 6) as a result of erosion events caused by increases in CWD, particularly in the LRW. Overall, these results underscore the importance of drainage devices and specific agricultural management practices in uplands (Vereecken et al., 2022; Wang et al., 2023).

In contrast to previous findings, our results indicate that SOC stocks will likely increase by an average of 14.2% to 35.5% in the total study area under scenario SSP2-4.5 (Fig. 7c). Under this scenario, TNx and TXx are crucial factors influencing SOC stocks, particularly due to increased net primary productivity driven by warming temperatures, CO_2 fertilization, and extended growing seasons that promote vegetation growth. (Elbasiouny et al., 2022). When the amount of biomass, including wood debris and roots, increases, more biomass can be transformed into stable SOC for storage. This finding is consistent with the microbial efficiency matrix stabilization framework, which treats root-derived microbial residues as the main precursors to mineral-associated organic matter, a stable SOC pool (Cotrufo et al., 2013; Sokol and Bradford, 2019).

Based on the SSP1-2.6 and SSP2-4.5 emission scenarios, severe warming and prolonged droughts in mountainous areas may result in wildfires or drought-induced dieback, which may reverse previous carbon gains in cases of ecological equilibrium within forests. These climate conditions may overwhelm current agricultural systems and infrastructure and undermine the current ecological carrying capacity (Zhang et al., 2020). Although a slight increase in SOC stocks was projected on the emission scenarios, some uplands will experience clear losses in SOC stocks. Previous study reported a reduced SOC stocks in slope-lands (Sarstedt et al., 2014). This effect indicates the vulnerability of uplands to erosion caused by runoff under extreme precipitation, thereby threatening long-term carbon retention. As observed in this study, the increases in CDD might decrease vegetation growth and soil moisture, which in turn will lead to less organic input and greater carbon losses. This condition is common in uplands, not mountainous areas, because the effects of temperature extremes on SOC stocks may be weakened by topography. These findings were confirmed by our PLS-PM analysis (Fig. 9). In terms of strategies for adaptation to temperature extremes, firebreak corridors or buffer zones in mountainous areas and drainage constructions in uplands should be prioritized.

Although the Cubist, Random Forest, and regression kriging models in this study demonstrated reasonable predictive performance, several limitations should be acknowledged. First, these models are data-driven and have limited extrapolation ability in regions beyond the range of the training data (Meyer and Pebesma, 2021). In addition, the uneven distribution of soil samples represents a major source of uncertainty; the scarcity of sampling points in mountainous areas results in higher prediction uncertainty (Jien et al., 2025). These limitations should be considered when interpreting the model results and the spatial distribution of SOC. In addition to climatic factors, land type plays a major role in SOC stock responses. In this

study, mountainous areas were found to have higher levels of SOC and to be more sensitive to climate change compared with other land types. The varied responses across different land types emphasize the need to include topography, climate, and land management practices in SOC stock models and the importance of developing carbon mitigation strategies (IPCC, 2019b). Ultimately, these findings highlight the importance of topography, climate, and land type as essential factors for improving SOC predictions and guiding carbon management planning.

5. Conclusions

This study established effective DSM models for predicting SOC stock, achieving an R^2 range of 0.43–0.50, and identified key environmental covariates, such as topography, climate, NDVI, and the prediction interval maps for identifying areas not covered in the sampling distribution. The findings suggest that even with observed predictive disparities, the Regression Kinging with Cubist and RF models can still be relied upon for overall predictive ability. The results demonstrated spatio-temporal variability in projected topsoil SOC under different emission scenarios, with clear sensitivity to landscape type and climate extremes. Under a moderate emissions scenario, SOC dynamics were highly sensitive to extreme climate events, with land type also playing a key role. These effects pose both location- and time-specific challenges for SOC management in studies on mid- to late-century time points. In the low-emission scenario (scenario SSP1-2.6), extreme rainfall events are predicted to induce a significant reduction in SOC stocks through erosion in upland areas. However, in the moderate emission scenarios (scenario SSP2-4.5), warming (TNx and TXx) and extreme rainfall events (R95p and R99p) may simultaneously increase biomass input and increase soil erosion risks.

These findings indicate that SOC management strategies should be highly specific to the site and time. In mountainous areas of the ZRW and LRW, even though the SOC stock dynamics in mountainous areas are likely to be affected by extreme rainfall events, heat waves, and prolonged droughts, future mitigation strategies should focus on reducing warming, preventing wildfires, and promoting heat-tolerant tree species. In upland areas, SOC stock changes in ZRW and LRW are predicted to be mainly driven by R95p, R99p, and CWD, where significant SOC losses will occur in certain upland areas for all emission scenarios. Therefore, management strategies should emphasize soil and water conservation to ensure that excess rainfall can be infiltrated into the soil without triggering erosion. These strategies should include the implementation of eco-engineering techniques on slope lands, maintaining vegetation cover and soil permeability, and establishing effective drainage systems. Overall,

clarifying the interactions between climatic extremes, land types, and SOC stocks to develop site-specific management practices is key to enhancing soil's resilience to sustain ecosystem functions in a changing climate.

Funding

This work was supported by the National Science and Technology Council of Taiwan (NSTC 113-2321-B-005-004 and NSTC 112-2313-B-005-059-MY2).

CRediT authorship contribution statement

SHJ: Writing—Review & Editing, Writing—Original Draft, and Data Curation. **CHS and CCY:** Methodology, Investigation, and Conceptualization. **SM and BJY:** Visualization and Methodology. **YMT:** Writing—Review & Editing.

Declaration of competing interests

The authors confirm that they have no financial or personal relationships that could have appeared to influence the research presented in this paper.

Data availability

The data used or analyzed in this study are available from the corresponding author upon reasonable request.

References

Adhikari, K., Hartemink, A. E., Minasny, B., Bou Kheir, R., Greven, M. B., and Greve, M. H.: Digital mapping of soil organic carbon contents and stocks in Denmark, PLoS ONE 9, e105519, https://doi.org/10.1371/journal.pone.0105519, 2014.

Blake, G. R., and Hartge, K. H.: Bulk density, in: Klute, A. (Ed.), Methods of soil analysis. Part 1. Physical and Mineralogical Methods, 2nd ed. ASA, SSSA, Madison, WI, 363–375, https://doi.org/10.2136/sssabookser5.1.2ed.c13, 1986.

Breiman, L.: Random forests, Machine Learning. 41, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Chalchissa, F. B., and Kuris, B. K.: Modeling the impacts of extreme climate scenarios on soil acidity (pH and exchangeable aluminum) in Abbay River Basin, Ethiopia, Heliyon, 10, 12, e33448, https://doi.org/10.1016/j.heliyon.2024.e33448, 2024.

Chalchissa, F. B., Diga, G. M., Feyisa, G. L., and Tolossa, A.R.: Impacts of extreme agroclimatic indicators on the performance of coffee *(Coffea arabica L.)* aboveground biomass in Jimma Zone, Ethiopia, Heliyon, 8, 8, e10136, https://doi.org/10.1016/j.heliyon.2022.e10136, 2022.

Chen, F., Wang, S. J., Dong, Q. J., Esper, J., Büntgen, U., Meko, D., Linderholm, H. W., Wang, T., Yue, W. P., Zhao, X. E., Hadad, M., González-Reyes, Á., and Chen, F.H.: Role of Pacific Ocean climate in regulating runoff in the source areas of water transfer projects on the Pacific Rim. npj Clim, Atmos. Sci., 7, 153, https://doi.org/10.1038/s41612-024-00706-1, 2024.

Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K., and Paul, E.: The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter?, Glob. Chan. Bio., 19, 988–995, https://doi.org/10.1111/gcb.12113, 2013.

Dialynas, Y. G., Bastola, S., Bras, R. L., Billings, S. A., Markewitz, D., and Richter, D. d B.: Topographic variability and the influence of soil erosion on the carbon cycle, Glob. Biogeochem. Cycles, 30, 644–660, https://doi.org/10.1002/2015GB005302, 2016.

Edmondson, J. L., Davies, Z. G., McCormack, S. A., Gaston, K. J., and Leake, J.R.: Land–cover effects on soil organic carbon stocks in a European city, Sci. Total Environ, 472, 444–453, https://doi.org/10.1016/j.scitotenv.2013.11.025, 2014.

Elbasiouny, H., El-Ramady, H., Elbehiry, F., Rajput, V. D., Minkina, T., and Mandzhieva, S.: Plant Nutrition under Climate Change and Soil Carbon Sequestration, Sustainability, 14, 914, https://doi.org/10.3390/su14020914, 2022.

Grace, J.: Understanding and managing the global carbon cycle, J. Ecol., 92, 189–202, https://doi.org/10.1111/j.0022-0477.2004.00874.x, 2004.

Gray, J. M., Bishop, T. F. A., and Wilson, B. R.: Factors controlling soil organic carbon stocks with depth in eastern Australia, Soil Sci. Soc. Am. J., 79, 1741–1751, https://doi.org/10.2136/sssaj2015.06.0224, 2015.

Grunwald, S.: Multi–criteria characterization of recent digital soil mapping and modeling approaches, Geoderma, 152, 195–207, https://doi.org/10.1016/j.geoderma.2009.06.003, 2009.

Guo, L., Shi, T. Z., Linderman, M., Chen, Y. Y., Zhang, H. T., and Fu, P.: Exploring the Influence of Spatial Resolution on the Digital Mapping of Soil Organic Carbon by

Airborne Hyperspectral VNIR Imaging, Remote Sens, 11, 1032, https://doi.org/10.3390/rs11091032, 2019.

Hijmans, R. J. raster: Geographic data analysis and modeling. https://doi.org/10.32614/CRAN.package.raster, 2022.

IPCC: 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Agriculture, Forestry and Other Land Use. Volume 4, https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html (last access: October 22, 2025), 2006.

IPCC: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Agriculture, Forestry and Other Land Use. Volume 4, https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol4.html (last access: October 22, 2025), 2019a.

IPCC: Climate Change and Land: Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems, https://doi.org/10.1017/9781009157988, 2019b.

Jenny, H.: Factors of soil formation: a system of quantitative pedology, McGraw Hill, New York, 1941.

Jien, S. H., Minasny, B., Yang, B. J., Liu, Y. T., Yen, C. C., Ocba, M. A., Zhang, Y. T., and Syu, C. H.: Enhancing Soil Carbon Storage: Developing high-resolution maps of topsoil organic carbon sequestration potential in Taiwan, Geoderma, 459, 117369, https://doi.org/10.1016/j.geoderma.2025.117369, 2025.

Johnston, C. A., Groffman, P., Breshears, D. D., Cardon, Z. G., Currie, W., Emanuel, W., Gaudinski, J., Jackson, R. B., Lajtha, K., Nadelhoffer, K., Jr., D. N., and Post, W. M.: Carbon cycling in soil. Front, Ecol. Environ, 2, 522–528, https://doi.org/10.1890/1540-9295(2004)002[0522:CCIS]2.0.CO;2, 2004.

Keskin, H., and Grunwald, S.: Regression kriging as a workhorse in the digital soil mapper's toolbox, Geoderma, 326, 22–41, https://doi.org/10.1016/j.geoderma.2018.04.004, 2018.

Kuhn, M.: Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5), 1–26. https://doi.org/10.18637/jss.v028.i05, 2008.

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., and Walter, C.: High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape, Geoderma, 213, 296–311, https://doi.org/10.1016/j.geoderma.2013.07.002, 2014.

Lamichhane, S., Kumar, L., and Wilson, B.: Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review, Geoderma, 352, 395–413, https://doi.org/10.1016/j.geoderma.2019.05.031, 2019.

730 Li, H. W., Wu, Y. P., Chen, J., Zhao, F., Wang, F., Sun, Y. H., Zhang, G. C., and Qiu, L. J.:
731   Responses of soil organic carbon to climate change in the Qilian Mountains and its future
732   projection, J. Hydrol, 596, 126110, https://doi.org/10.1016/j.jhydrol.2021.126110, 2021.

733 Liaw, A., and Wiener, M.: Classification and regression by RandomForest, R news, 2, 18-22,
734   https://CRAN.R-project.org/doc/Rnews/, 2002.

735 Ma, Y. X., Minasny, B., and Wu, C.: Mapping key soil properties to support agricultural
736   production in Eastern China, Geoderma Regional, 10, 144–153,
737   https://doi.org/10.1016/j.geodrs.2017.06.002, 2017.

738 Malone, B. P., Minasny, B., POdgers, N., and McBratney, A. B.: Using model averaging to
739   combine soil property rasters from legacy soil maps and from point data, Geoderma, 232,
740   34–44, https://doi.org/10.1016/j.geoderma.2014.04.033, 2014.

741 McBratney, A. B., Santos, M. M., and Minasny, B.: On digital soil mapping, Geoderma, 117,
742   3–52, https://doi.org/10.1016/S0016-7061(03)00223-4, 2003.

743 Meyer, H., & Pebesma, E.: Predicting into unknown space? Estimating the area of applicability
744   of spatial prediction models. Methods in Ecology and Evolution, 12(9), 1620-1633,
745   https://doi.org/10.1111/2041-210X.13650, 2021.

746 Minasny, B., and McBratney, A.B.: Digital soil mapping: A brief history and some lessons,
747   Geoderma, 264, 301–311, https://doi.org/10.1016/j.geoderma.2015.07.017, 2016.

748 Mishra, U., and Riley, W. J.: Scaling impacts on environmental controls and spatial
749   heterogeneity of soil organic carbon stocks, Biogeosciences, 12, 3993–4004,
750   https://doi.org/10.5194/bg-12-3993-2015, 2015.

751 Nelson, D. A., and Sommers, L.E.: Total carbon, organic carbon, and organic matter, in: Sparks,
752   D.L. (Ed.), Methods of soil analysis. Part 3, Chemical Methods. ASA, SSSA, Madison,
753   WI, 961–1010, https://doi.org/10.2136/sssabookser5.3.c34, 1996.

754 Olaya-Abril, A., Parras-Alcántara, L., Lozano-García, B., and Obregón-Romero, R.: Soil
755   organic carbon distribution in Mediterranean areas under a climate change scenario via
756   multiple linear regression analysis, Sci. Total Enviro., 592, 134–143,
757   https://doi.org/10.1016/j.scitotenv.2017.03.021, 2017.

758 Peng, Y., Zhou, W., Xiao, J., Liu, H., Wang, T., & Wang, K.: Comparison of Soil Organic
759   Carbon Prediction Accuracy Under Different Habitat Patches Division Methods on the
760   Tibetan Plateau. Land Degradation & Development. https://doi.org/10.1002/ldr.70184,
761   2025.

Quinlan, J.R.: Learning with Continuous Classes, in: Proceedings of Australian Joint Conference on Artificial Intelligence, Hobart, 1992.

Rial, M., Cortizas, A. M., and Rodríguez-Lado, L.: Understanding the spatial distribution of factors controlling topsoil organic carbon content in European soils, Sci. Total Enviro., 609, 1411–1422, https://doi.org/10.1016/j.scitotenv.2017.08.012, 2017.

Rillig, M. C., Leifheit, E., and Lehmann, J.: Microplastic effects on carbon cycling processes in soils, PLoS Biol., 19, e3001130, https://doi.org/10.1371/journal.pbio.3001130, 2021.

Rudiyanto, Minasny, B., Setiawan, B. I., Saptomo, S. K., and McBratney, A. B.: Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands, Geoderma, 313, 25–40, https://doi.org/10.1016/j.geoderma.2017.10.018, 2018.

Sanchez, G., Trinchera, L., Russolillo, G.: plspm: tools for partial least squares path modeling (PLS-PM), R Package Version (4), p. 9 https://CRAN.R-project.org/package=plspm, 2017

Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., and Hair, J. F.: Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers, J. Fam. Bus. Strategy, 5, 105–115, https://doi.org/10.1016/j.jfbs.2014.01.002, 2014.

Shaik, A. B., and Srinivasan, S.: A Brief Survey on Random Forest Ensembles in Classification Model, in: Bhattacharyya, S., Hassanien, A., Gupta, D., Khanna, A., Pan, I. (Eds.), International Conference on Innovative Computing and Communications. Lecture Notes in Networks and Systems, 56. Springer, Singapore, 253–260, 2019.

Siewert, M. B.: High-resolution digital mapping of soil organic carbon in permafrost terrain using machine learning: a case study in a sub-Arctic peatland environment, Biogeosci., 15, 1663–1682, https://doi.org/10.5194/bg-15-1663-2018, 2018.

Singh, B. P., Setia, R., Wiesmeier, M., and Kunhikrishnan, A.: Chapter 7 - Agricultural management practices and soil organic carbon storage, Soil carbon storage, 207–244, https://doi.org/10.1016/B978-0-12-812766-7.00007-X, 2018.

Sokol, N. W., and Bradford, M. A.: Microbial formation of stable soil carbon is more efficient from belowground than aboveground input, Nat. Geosci., 12, 46–53, https://doi.org/10.1038/s41561-018-0258-6, 2019.

Therneau, T., Atkinson, B.: Rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-15, https://CRAN.R-project.org/package=rpart (last access: November 3, 2025).

794 Tsui, C. C., Liu, X. N., Guo, H. Y., and Chen, Z.S.: Effect of Sampling Density on Estimation
795   of Regional Soil Organic Carbon Stock for Rural Soils in Taiwan, Geospatial Technology–
796   Environmental and Social Applications, 35–53, https://doi.org/10.5772/64210, 2016.

797 Vaysse, K., and Lagacherie, P.: Evaluating digital soil mapping approaches for mapping
798   GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France),
799   Geoderma Reg., 4, 20–30, https://doi.org/10.1016/j.geodrs.2014.11.003, 2015.

800 Vereecken, H., et al.: Soil hydrology in the Earth system, Nat. Rev. Earth Environ., 3, 573–
801   587, https://doi.org/10.1038/s43017-022-00324-6, 2022.

802 Viaud, V., Angers, D. A., and Walter, C.: Toward landscape-scale modeling of soil organic
803   matter dynamics in agroecosystems, Soil Sci. Soc. Am. J., 74, 1847–1860,
804   https://doi.org/10.2136/sssaj2009.0412, 2010.

805 Villanueva, R. A. M., and Chen, Z. J.: ggplot2: Elegant Graphics for Data Analysis (2nd ed.),
806   Measurement-Interdisciplinary Research and Perspectives, 17, 160-167,
807   https://doi.org/10.1080/15366367.2019.1565254, 2019.

808 Wang, B., Waters, C., Orgill, S., Gary, J., Cowie, A., Clark, A., and Liu, D. L.: High resolution
809   mapping of soil organic carbon stocks using remote sensing variables in the semi-arid
810   rangelands of eastern Australia, Sci. Total Enviro., 630, 367–378,
811   https://doi.org/10.1016/j.scitotenv.2018.02.204, 2018.

812 Wang, M. M., Zhang, S., Guo, X. W., Xiao, L. J., Yang, Y. H., Luo, Y. Q., Mishra, U., and
813   Luo, Z. K.: Responses of soil organic carbon to climate extremes under warming across
814   global biomes, Nat. Clim. Change, 14, 98–105, https://doi.org/10.1038/s41558-023-
815   01874-3, 2023.

816 Wei, Y. C., Wang, M. M., Viscarra Rossel, R. A., Chen, H., and Luo, Z. K.: Extreme Climate
817   as the Primary Control of Global Soil Organic Carbon Across Spatial Scales, Global
818   Biogeochem. Cycles, 38, e2024GB008277, https://doi.org/10.1029/2024GB008200, 2024.

819 Wickham, H.: ggplot2: Elegant graphics for data analysis, 2nd ed. 2016. ed, Use R! Springer
820   International Publishing: Imprint: Springer, Cham, https://doi.org/10.1007/978-3-319-
821   24277-4, 2016.

822 Wiesmeier, M., Urbanski, L., Hobley, E., Lützow, M. V., Marin-Spiotta, E., Wesemael, B. V.,
823   Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H. J., and Kögel-Knabner,
824   I.: Soil organic carbon storage as a key function of soils–A review of drivers and indicators
825   at various scales, Geoderma, 333, 149–162,
826   https://doi.org/10.1016/j.geoderma.2018.07.026, 2019.

827 Zhang, G. Y., Luo, S., Jing, Z. W., Wei, S., and Ma, Y. H.: Evaluation and Forewarning
828   Management of Regional Resources and Environment Carrying Capacity: A Case Study
829   of Hefei City, Anhui Province, China, Sustainability, 12, 1637,
830   https://doi.org/10.3390/su12041637, 2020.

831 Zhu, Q., and Lin, H. S.: Comparing ordinary kriging and regression kriging for Soil properties
832   in contrasting landscapes, Pedosphere, 20, 594–606, https://doi.org/10.1016/S1002-
833   0160(10)60049-5, 2010.

834

835

836 **Table 1.** Environmental covariates.

Type of data	Environmental covariates	Soil forming factor	Type
Remote sensing	Normalized difference vegetation index (NDVI)	o; t	Q
Digital elevation model	Elevation	r	Q
	Slope	r	Q
	Aspect	r	Q
	Terrain ruggedness index (TRI)	r	Q
	Topographic wetness index (TWI)	r	Q
	Terrain position index (TPI)	r	Q
	Multiresolution Index of Valley Bottom Flatness (MrVBF)	r	Q
	Multiresolution Ridge Top Flatness (MrRTF)	r	Q
	Stream power index (SPI)	r	Q
	Curvature	r	Q
	Flow accumulation	r	Q
	R-value		Q
	K-value		Q
Climate	Mean annual temperature (MAT)	c; t	Q
	Total annual precipitation (TAP)	c; t	Q
Land cover	Land cover	o; t	C
Soil	Soil Order	s	C

837 s: soil, r: relief, c: climate, t: time, o: organism, Q: quantitative, C: categorical.

838

839 **Table 2.** Data partitioning rules and 5th/95th percentiles of leave-one-out validation residuals.

Class	Conditions	Residual percentile	
		P_5	P_{95}
1	Soil Order in Other	-1.08	0.80
2	Topographic Position Index ≤ -10.8889	-1.99	0.95
3	MAT > 17.09029, Topographic Position Index > -10.8889, Soil Order in Inceptisol, Entisol, Alfisol, Spodosol, Ultisol	-0.85	0.70
4	MAT ≤ 17.09029	-0.67	0.80

840

841

842   **Table 3.** Variance and coefficient of variance (CV) of the spatiotemporal distribution of SOC

843   stocks for various land uses in the two emission scenarios.

844

		2020				2050				2100			
		T	L	U	M	T	L	U	M	T	L	U	M
	Variance	6.45	0.66	1.35	5.11								
	CV(%)	44.3	32.3	30.7	32.5								
SSP1	Variance					5.52	0.35	0.85	4.37	8.24	0.76	1.56	6.45
2.6	CV(%)					50.4	29.8	33.9	36.0	51.3	33.5	39.2	36.4
SSP2	Variance					10.1	1.14	2.25	7.95	21.4	1.74	2.53	18.7
4.5	CV(%)					48.5	35.0	37.5	34.9	59.7	35.2	40.4	43.7

845   CV: coefficient of variance; T: total area; L: lowland regions; U: Upland regions; M: Mountainous regions.

846

847

848

849

850

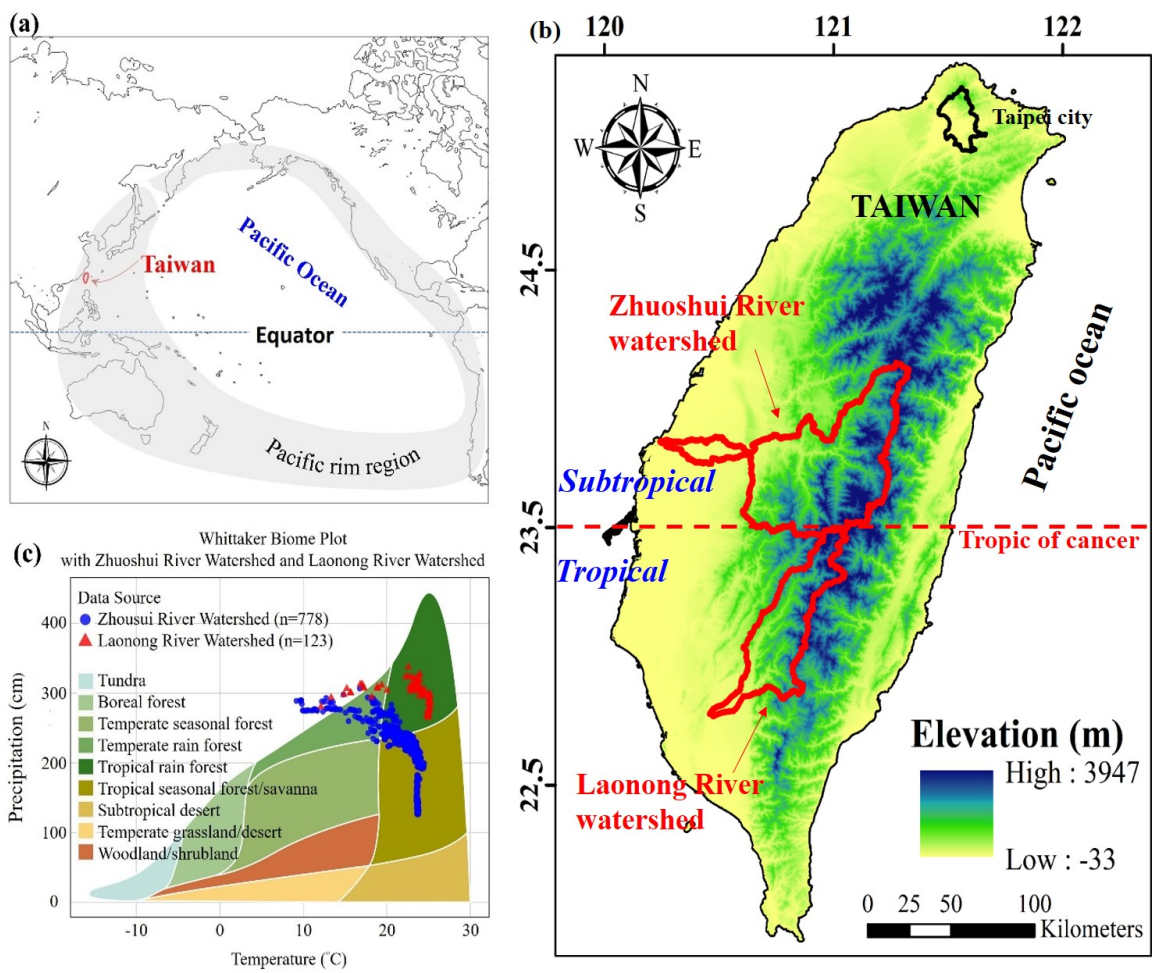
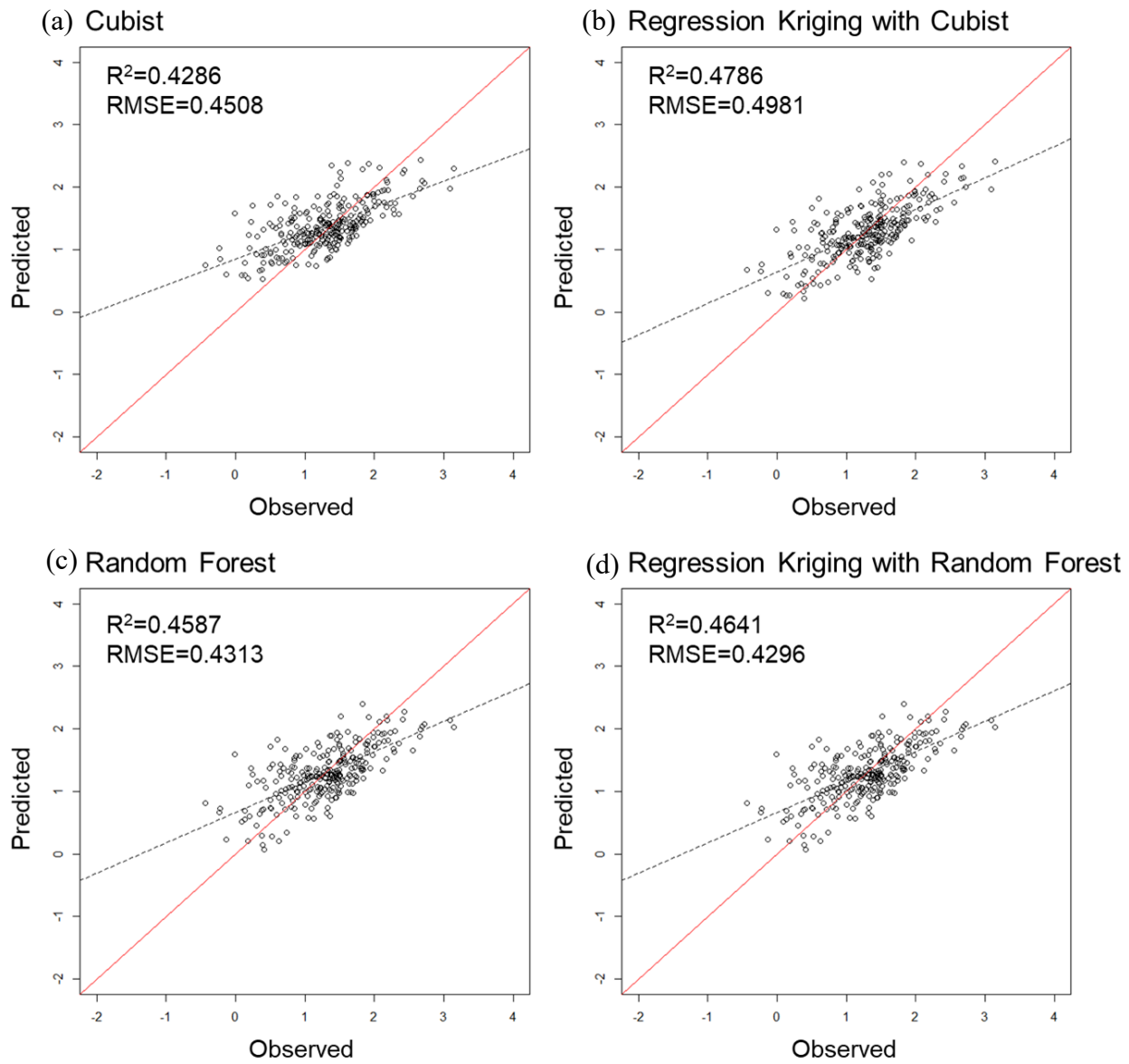



Fig. 1. Location of Taiwan in Pacific Ocean regions (a); location of Zhuoshui River watershed

(ZRW) and Laonong River watershed (LRW) in Taiwan; (c) Whittaker Biome Plot of ZRW and

LRW.

854   **Fig. 2.** Scatter plots of predicted versus observed soil organic carbon (SOC) stock in the topsoil

855   (0–30 cm) where predictions were obtained on the basis of validation data and by using the (a)

856   Cubist, (b) regression kriging (RK) with Cubist, (c) Random forest (RF), and (d) regression

857   kriging with RF models. The *x*-axis represents the observed values, and the *y*-axis represents

858   the predicted values. The solid line is the fitted line.

859

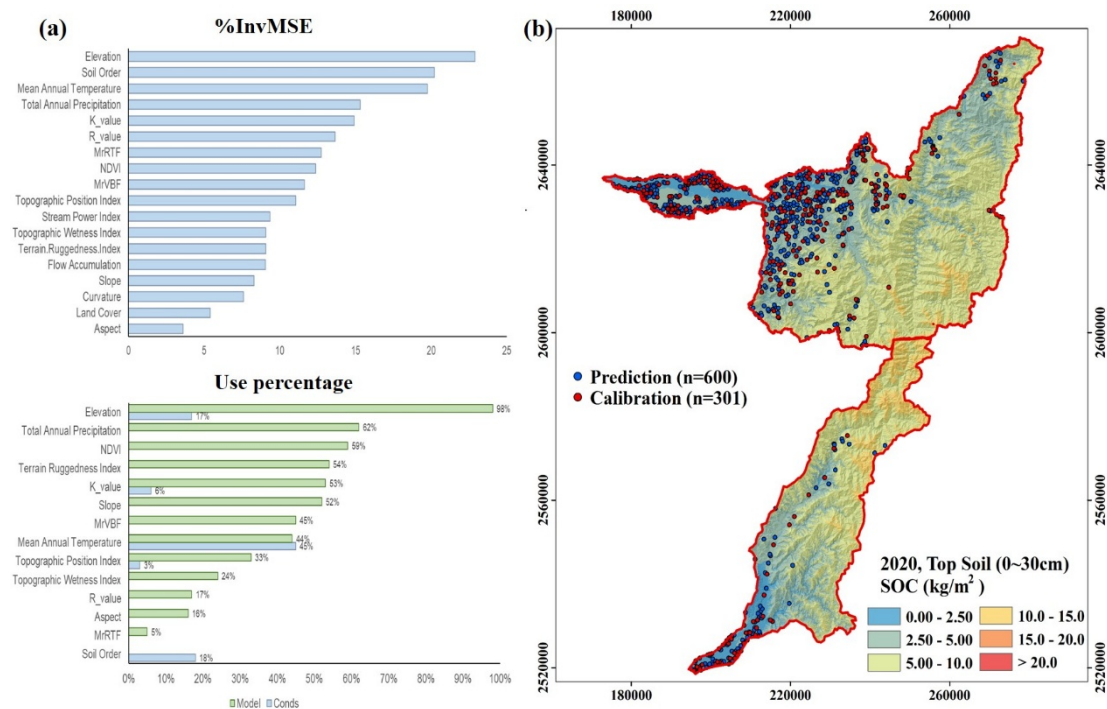
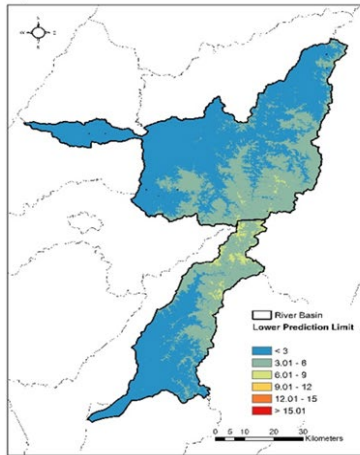
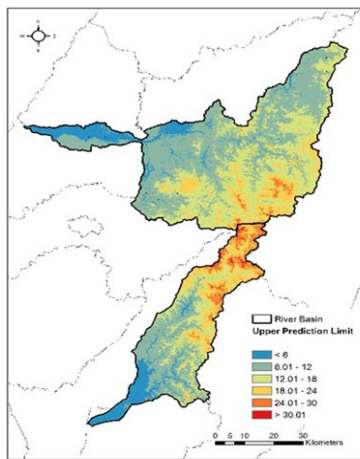
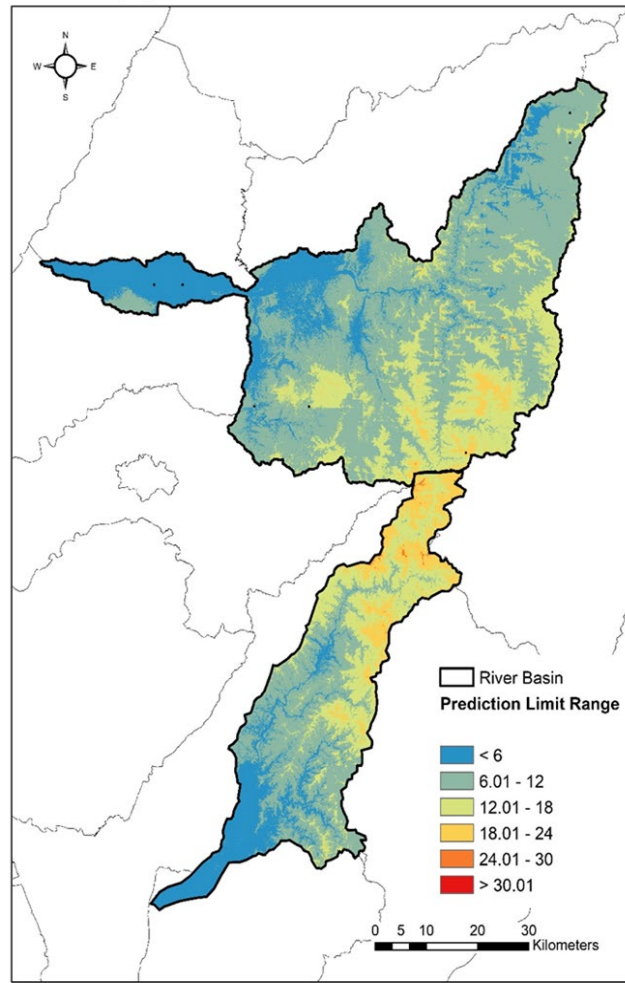



Fig. 3. (a) Variable importance of Random Forest and Cubist models for SOC stock in surface

soils (0–30 cm) and (b) predictive map of SOC stock in Zhuoshui River watershed and Laonong

River watershed.

863

864 **Fig. 4.** Topsoil (0–30 cm) soil organic carbon (SOC) stock maps of the (a) 90% lower prediction

865 limit, (b) 90% upper prediction limit, and (c) prediction limit range derived using bootstrapping.

866

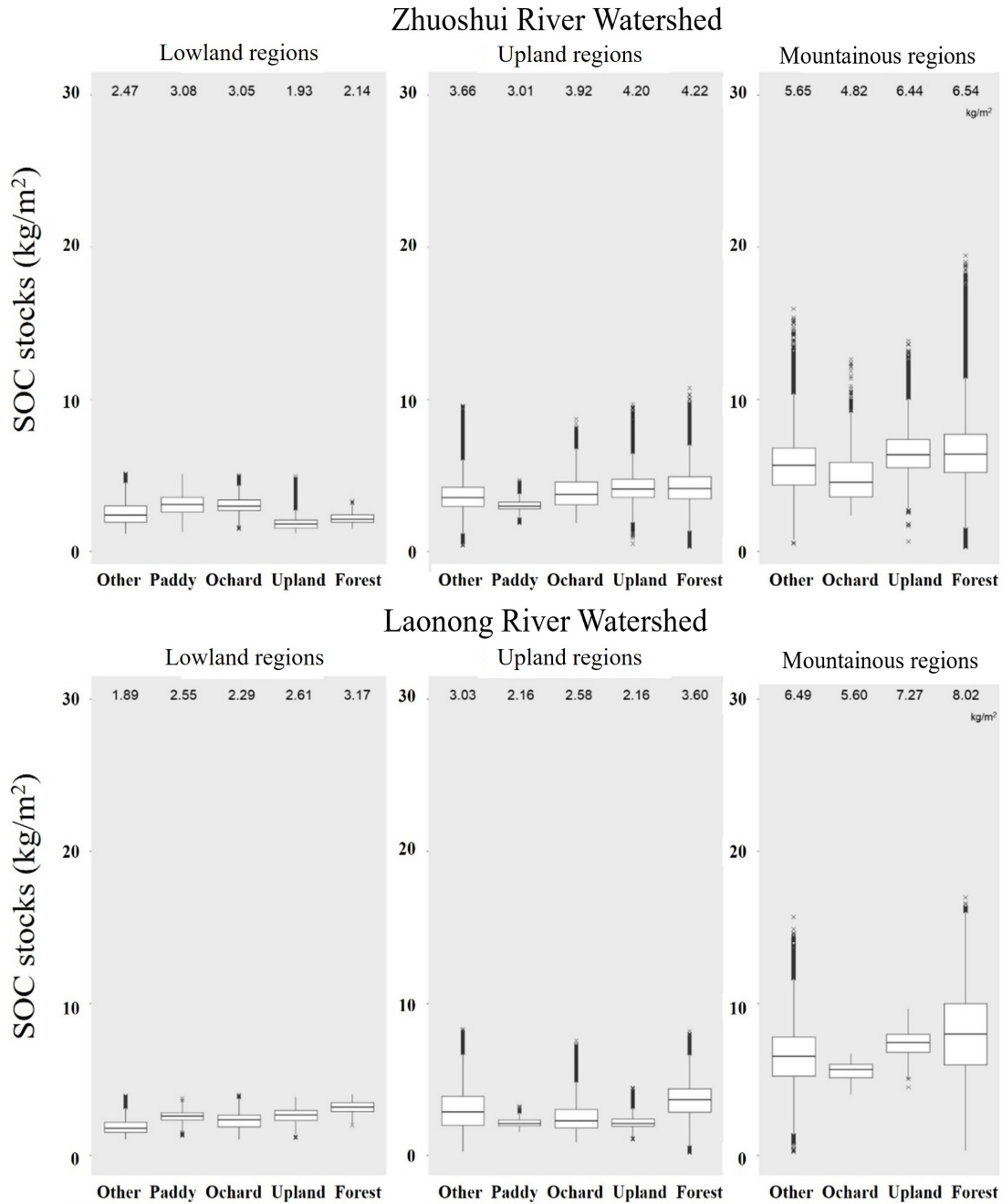
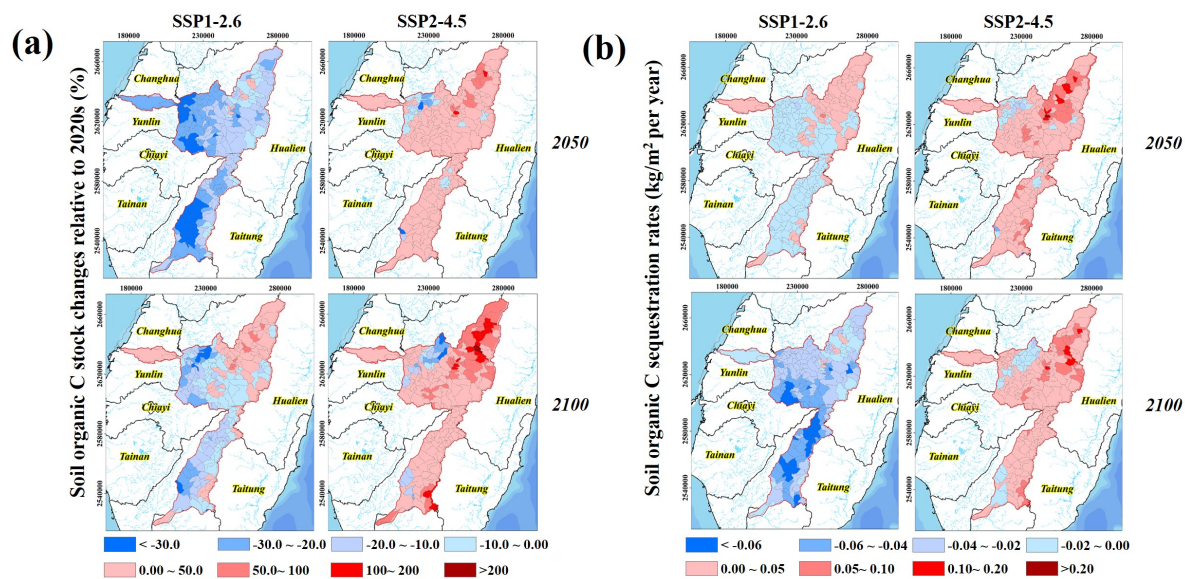



Fig. 5. Boxplots of topsoil (0–30 cm) soil organic carbon (SOC) stocks for various land cover: (left) lowland regions (<100 m in elevation), (middle) upland regions (100–1000 m in elevation), and (right) mountainous regions (>1000 m in elevation) at Zhuoshui River watershed and Laonong River watershed. The upland land cover represents the upland farming.

873

874 **Fig. 6.** Spatiotemporal predictions of (a) SOC stocks ($kg\,m^{-2}$) and (b) SOC sequestration rates

875 ($kg\,m^{-2}$ per year) relative to the 2020s under two emission scenarios. The mapping unit is sub-

876 catchments in Taiwan.

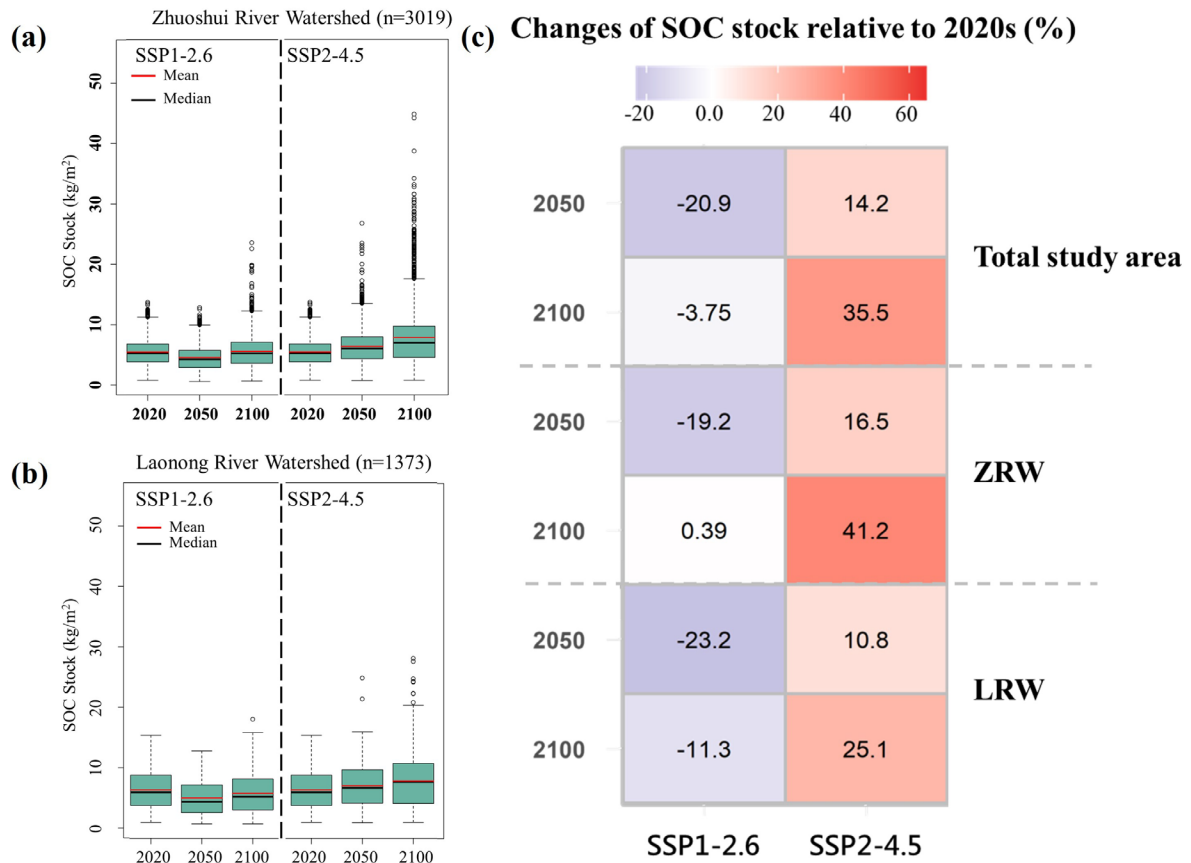
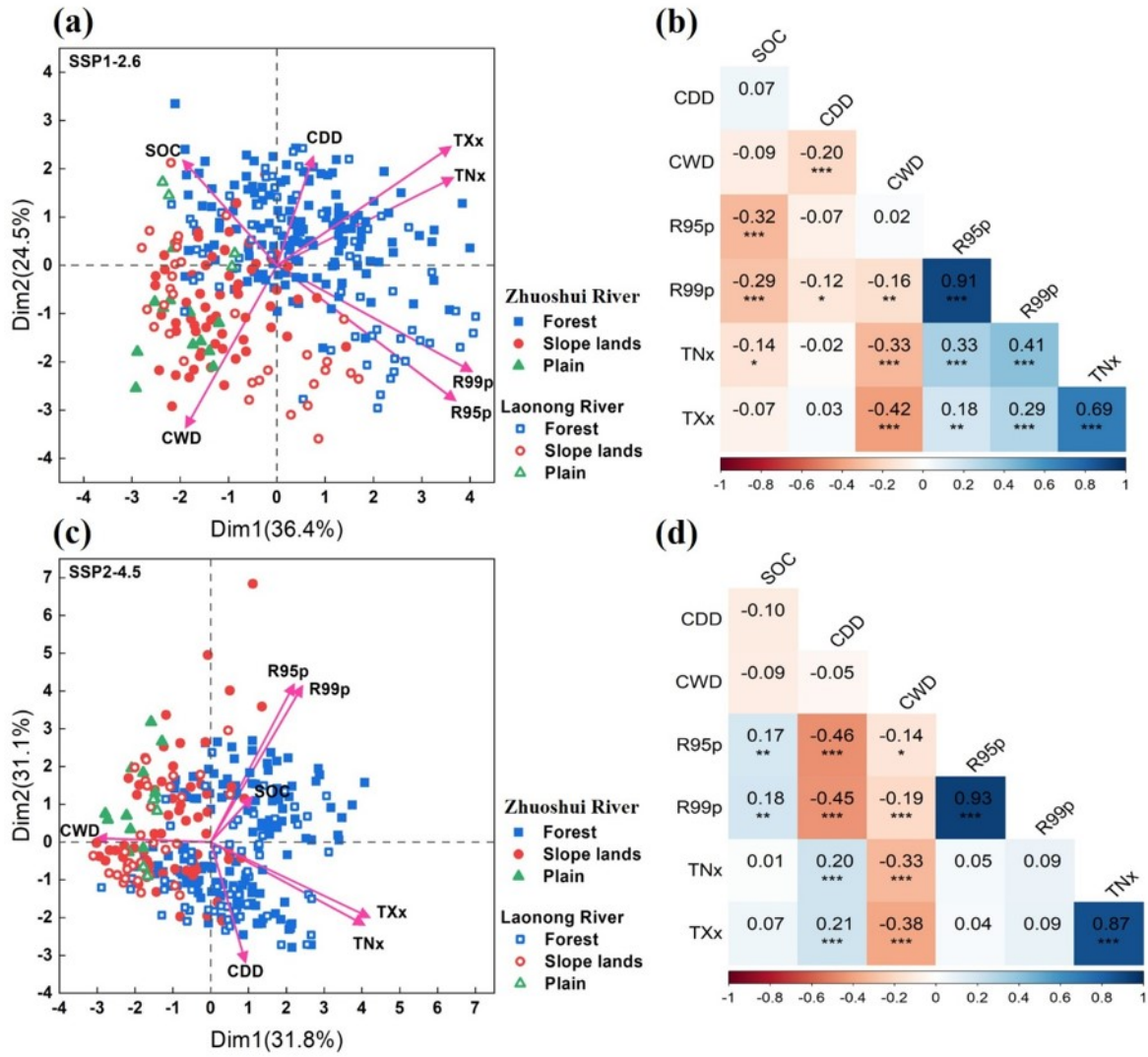



Fig. 7. Boxplots showing the temporal trends in predicted SOC stocks across two emission scenarios (SSP1-2.6 and SSP2-4.5) for 2020, 2050, and 2100 in (a) Zhuosui River watershed; and (b) Laonong River watershed, and (c) Increase in the ratio of SOC stocks relative to the 2020s in the ZRW and LRW for the two emission scenarios for 2050 and 2100.

883

884 **Fig. 8.** Principal component analysis and Pearson's correlation coefficient of extreme climate

885 indices and SOC stocks: (a, b) and SSP1-2.6, (c, d) SSP2-4.5

886

887

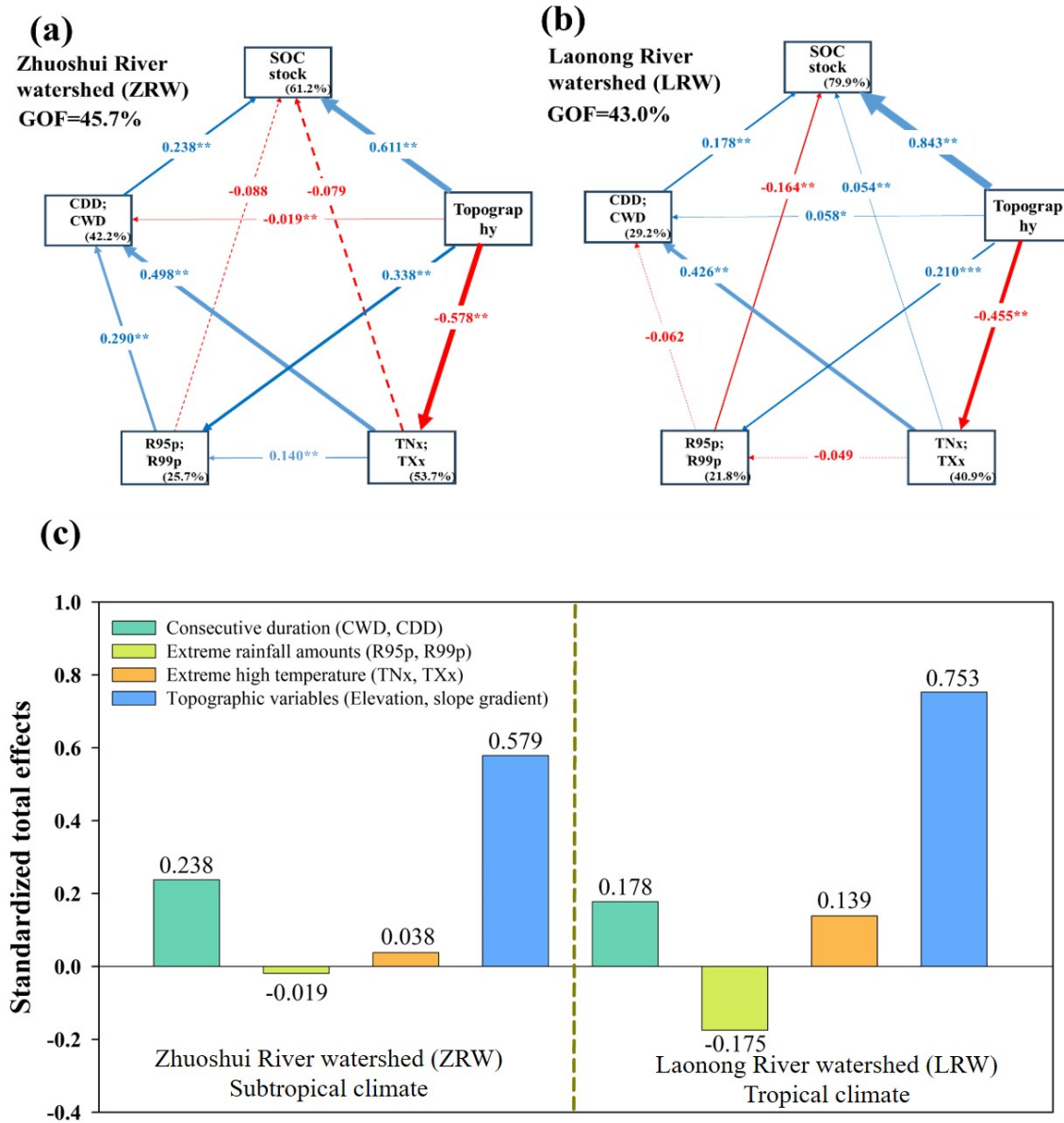



Fig. 9. Partial least squares path modeling (PLS-PM) analysis of the relationships among SOC

stocks, consecutive durations of extreme climatic events (CDD and CWD), extreme rainfall

amounts (R95p and R99p), extreme temperatures (TNx and TXx), and topographic variables

(elevation and slope gradient) (a) Zhuoshui River watershed; (b) Laonong River watershed; (c)

standardized total effects. Positive and negative effects are represented by blue and red arrows,

respectively. Path coefficients that do not significantly differ from zero are depicted as gray

dashed lines: $^*p < 0.05$ and $^{**}p < 0.01$. The percentages in the boxes represent the explanatory

power of the variables. The goodness-of-fit was used to assess the model.