J Soils Sediments (2017) 17:1101–1115
DOI 10.1007/s11368-016-1573-4

SOILS, SEC 5 • SOIL AND LANDSCAPE ECOLOGY • RESEARCH ARTICLE

Topography-soil relationships in a hilly evergreen broadleaf forest in subtropical China

Xiaopeng Li[1,2] · Scott X. Chang[1] · Jintao Liu[3] · Zemei Zheng[4,5] · Xihua Wang[4,5]

Received: 7 June 2015 / Accepted: 25 September 2016 / Published online: 14 October 2016

Abstract
Purpose Topography-soil relationships usually vary with climate, vegetation type, degree of human disturbance, type of parent material, and the scale being studied. In this paper, we studied the topography-soil relationship in a hilly forest in subtropical China.
Materials and methods The influence of topography on soil properties (soil moisture, organic carbon (C), total nitrogen (N) and total phosphorus contents, C:N ratio, and pH) was evaluated using a recursive partitioning conditional inference tree (CIT) as well as a multiple linear regression (MLR) method.
Results and discussion The CIT models generally performed better than MLR in describing the topography-soil relationships. Topographic parameters chosen by the CIT models, which indicate the mechanisms at play for the spatial variation of the soil properties, varied with the soil property of concern. The soil moisture, organic C, and total N models contained only primary terrain attributes, the soil C:N ratio and pH models contained both primary and secondary terrain attributes, while the total phosphorus model contained mostly secondary terrain attributes.
Conclusions The CIT method worked well for exploring the topography-soil relationships in the studied undisturbed hilly forest. We conclude that (1) soil moisture, organic C, and total N were strongly affected by location-specific topographic features such as gravitational potential, the amount of precipitation, temperature, and vegetation type; (2) total phosphorus was affected by catchment-related hydrological activities and soil C:N ratio; and (3) pH was affected by location-specific topographic features and catchment-related hydrological activities.

Keywords Conditional inference tree · Hill forest · Primary and secondary terrain attributes · Spatial variation

Responsible editor: Arnaud Temme

✉ Xihua Wang
xhwang@des.ecnu.edu.cn

[1] Department of Renewable Resources, University of Alberta, T6G 2E3, Edmonton, Alberta, Canada

[2] Institute of Soil Science, Chinese Academy of Sciences, Nanjing, Jiangsu 210008, People's Republic of China

[3] State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, Jiangsu 210098, People's Republic of China

[4] Tiantong National Station of Forest Ecosystems, Chinese National Ecosystem Observation and Research Network, Ningbo, Zhejiang 315114, People's Republic of China

[5] College of Resources and Environmental Science, East China Normal University, Shanghai 200241, People's Republic of China

1 Introduction

The formation and evolution of soils over time are influenced by environmental factors such as topography, parent material, climate, and biota (Chen et al. 1997; McKenzie and Ryan 1999; Buol et al. 2011). Among these factors, topography influences the distribution of soils and their properties not only by controlling regional water flow and material transport, but also by changing the local climate, vegetation composition, and other conditions (Moore et al. 1993; (Florinsky et al. 2002; Gessler et al. 2000). The relationship between topographic characteristics and soil properties, in the form of a catena, has been widely studied in ecosystems ranging from boreal to tropical (Seibert et al. 2007; Barthold et al. 2008; Sumfleth and Duttmann 2008; Dlugoß et al. 2010; Liu et al. 2013). Some studies showed that topography has a strong influence on soil properties. For instance, slope and topographic wetness index (*TWI*) significantly influenced A horizon thickness, soil organic matter content, pH, extractable phosphorus, and soil particle size composition in an agricultural landscape in Colorado (Moore et al. 1993). Topographic parameters derived from digital elevation models were found to be significantly correlated with soil organic layer and E horizon thicknesses, pH, and carbon (C) to nitrogen (N) ratio (CN ratio) in boreal forests throughout Sweden (Seibert et al. 2007). And elevation, slope, stream power index (*SPI*), and *TWI* affected the spatial distribution of soil organic matter in the suburb of Beijing (Zhang et al. 2012). On the other hand, in a forested watershed in the Catskill Mountains of New York, a single terrain attribute is not be sufficient to effectively predict soil properties (Johnson et al. 2000), while in a tropical moist forest in Panama, topographic attributes are not key factors determining the spatial distribution of soil properties (such as exchangeable K and Mg) (Barthold et al. 2008). Inconsistencies among the studies indicate that topography-soil relationships are usually regionally unique and vary widely among biomes and land uses (Hancock et al. 2010). Site-specific studies are therefore required to further understand how soil properties are influenced by topography (Garcia-Pausas et al. 2007).

A number of approaches, such as regression analysis, computational intelligence, geostatistics, and decision trees, have been applied to derive topography-soil relationships. Regression analysis has been the most widely used method for assessing topography-soil relationships (Moore et al. 1993; Gessler et al. 1995, 2000), attributable to its simplicity in data processing and model structure, and ease of interpretation. However, they may not work well in some nonlinear cases, and moreover, if the soils have been evolved independently in different subregions of a study area, globally built regression equations may not work in some locations (Ziadat 2005). Computational intelligence such as artificial neural network and geostatistical techniques such as ordinary kriging, cokriging, and regression kriging were also popular in exploring topography-soil relationships (Tso and Yau 2007; Motaghian and Mohammadi 2011; Zhang et al. 2012). The artificial neural network method is good at dealing with nonlinear problems or when the relationship is unknown, but it does not provide p values for the significance test and is hard to interpret due to hidden layers in the model (Tso and Yau 2007). Geostatistical approaches are mainly used for spatial prediction rather than relationship extrapolation, and may lead to overall inaccurate outputs due to the smoothing effect when interpolation is used (Goovaerts 1999). Over the last decade, decision tree methods such as classification and regression tree, “ID3,” “C4.5,” and “C5.0” have been widely used for exploring relationships between complex ecological data (De’ath and Fabricius 2000), including relationships between topography and soil properties (Mertens et al. 2002; Gmur et al. 2012; Mage and Porder 2012).

A decision tree is a recursive partitioning method that explores the relationship between a response variable and several input variables by splitting the dataset into subsets based on an attribute value test. Different algorithms use different metrics for choosing a variable that best splits the dataset at each partition and deciding when the recursion should be completed. These algorithms have advantages of having flexible data input, simple model assumptions for complicated relationships, intuitive results, robustness with respect to outliers, and the ability to deal with missing values, and these algorithms have already been applied to predict soil types and soil properties (Tittonell et al. 2008; Vega et al. 2009; Gmur et al. 2012; Mage and Porder 2012). In these methods, Gini index or information gain were used as splitting measures; however, they have limitations such as overfitting and biased predictor selection (Shih 2004; Bramer 2007). The conditional inference tree (CIT) method is a recently developed decision tree method that overcomes overfitting and biased predictor selection problems by using multiple testing as the splitting measure rather than the Gini index or information gain used in other decision tree methods (Hothorn et al. 2006). The CIT method has gradually gained popularity (Hu and Cheng 2013).

The Southeast Coastal Hill Region in China is one of the most prosperous regions in the country and is also one of the areas severely affected by human activities due to rapid development and the increasing population. The mountainous terrain, high annual precipitation, and long-term intensive land use resulted in specific topography-soil relationships and diverse soil properties, sometimes associated with serious soil erosion and land degradation problems that impede rural development (Chen and Wang 2003; Yang et al. 2010). The objective of this study was to apply the CIT as well as a multiple linear regression (MLR) method to evaluate how soil properties were influenced by topography in a natural evergreen hilly forest in the region. The effects of seven topographic parameters (elevation (m), slope (β, °), plan (C_h, m^{-1}), profile curvature (C_v, m^{-1}), *TWI*, *SPI*, and length-slope factor (*LS*)) on soil properties (volumetric water content (VWC), soil organic C (SOC), total N (TN), CN ratio, pH, and total phosphorus (TP)) have been evaluated, and the performances of the CIT and MLR methods have also been compared.

2 Materials and methods

2.1 Study area and soil sampling

The study was carried out in an evergreen broad-leaf forest in the core area of Tiantong National Forest Park

(29° 48.696′~29° 48.938′ N, 121° 46.953′~121° 47.278′ E) in Zhejiang Province, China, and about 240 km south of Shanghai. Tiantong has a sub-tropical monsoon climate with hot and humid summers and cool winters. Annual average air temperature, precipitation, and evaporation were 16.2 °C, 1374.7 mm, and 1320.1 mm, respectively (Song and Wang 1995).

The study area is located on a hillslope that spans three small catchments with a total area of 37.6 ha. Most of this area was occupied by a long-term ecological monitoring plot, which is 400 m from north to south and 500 m from east to west and has 500 quadrats (grids) of 20 × 20 m in size (Fig. 1). The area has hardly been disturbed by human activity (Wang 2011), and is representative of natural hill forests in southeast China. Within the catchments, the parent material is uniform. Mesozoic sediment forms the bedrock of the hill and is also the parent material for the soil developed at the surface. Soil depth in the study area varies but averages about 1 m (Song and Wang 1995). Evergreen broad-leaved plants are the dominant species and deciduous plants are found in canopy gaps, and the land surface is generally covered with a litter layer.

Soils were all sampled following the existing 20 × 20 m grid of the ecological monitoring plot to locationally match up with the topographic data, which was treated in a 20 × 20 m grid-based format. In the study, samples were collected only within the ecological monitoring plot and a total of 471 grids were sampled considering both the practical operability at the locations and labor limitation (Fig. 1). In each grid, soil properties were represented by samples collected within the grid. One sample was collected at the center of each grid, and one or two more samples were randomly collected in most grids at the distance of 2, 5, or 8 m from the grid center to improve the representative of the sampling. All samples within a grid were averaged to represent the property of the grid. Sampling was finished in a short period of time in March 2011. During the sampling, there was no significant precipitation, and air temperatures mostly ranged between 5 and 15 °C.

Soil moisture content, bulk density, SOC, TN, pH, and TP were analyzed for the samples. Soil bulk density and gravimetric water content (GWC) were determined by oven-drying and weighing the core samples, and VWC was derived from the measured GWC and bulk density. Soil pH was determined using glass and calomel electrodes after extraction with deionized water with a soil to water ratio of 1:2 (*w*:*w*). Soil organic C and TN contents were determined using an elemental analyzer (vario MICRO cube, Elementar, Germany). Since inorganic C in the studied acidic soils was negligible, the total soil C was the same as SOC, and soil CN ratio was calculated as SOC divided by TN. Soil samples were digested with concentrated sulfuric acid and a copper sulfate pentahydrate-sodium sulfate catalyst for 3 h, and the solution was diluted and used for TP measurement. Soil TP content was determined using the digested soil samples on a continuous flow injection analyzer (SAN++, Skalar, Netherlands).

2.2 Topographic parameters

In this study, elevations were measured using a total station (SET2110, SOKKIA Corporation) following the 20-m grid of the ecological monitoring plot. The elevation measurement covered all three catchments from the divides to the outlets, and the digital elevation model grid nodes within the plot were overlapped with the soil sampling points. All the other topographic parameters, i.e., slope, C_h, and C_v, and three secondary terrain attributes: *TWI*, *SPI*, and *LS*, were computed from the sampled digital elevation model at the catchment scale (Fig. 2)

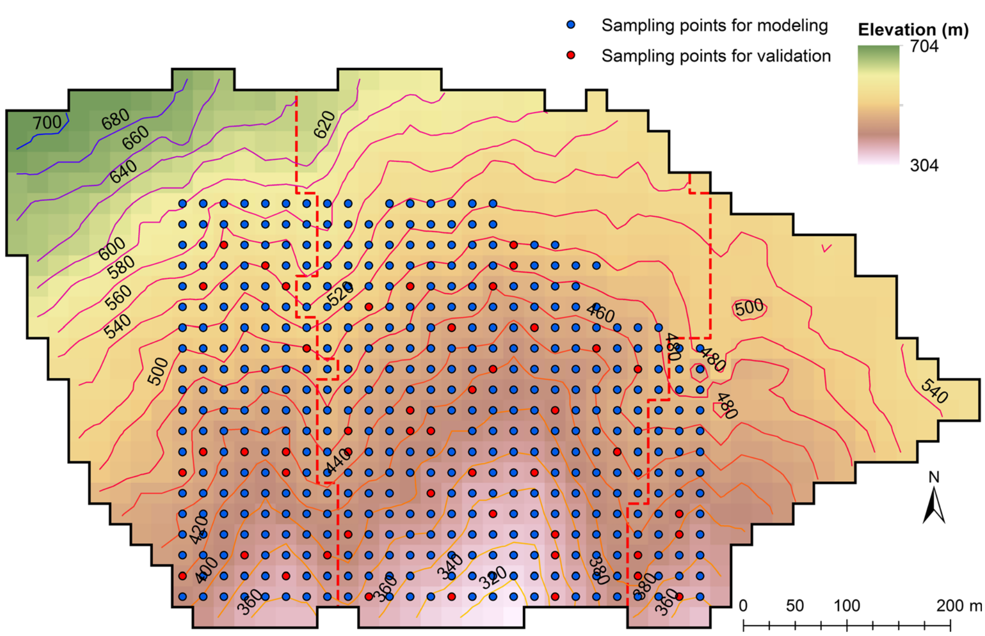


Fig. 1 Elevation map and sampling points in the study area (*dashed line* internal catchment divide)

using a terrain analysis tool called DigitalHydro, which is a GIS-based tool for elevation interpolation, topographic parameter calculation, and flow vector matrix and digital channel determination (Liu 2012).

Elevation determines the gravitational potential. Moreover, elevation not only contains geographic information but also reflects meteorological and vegetational conditions. Meteorological factors such as temperature and precipitation have clear trends with elevation in a hilly terrain (Montgomery 2006), and these changes further influence vegetation cover (Qiu and Zhong 2013). This means that if a soil property is influenced by elevation, it can be the result of gravitational substance flow, change in soil moisture content, temperature, vegetation, or all of them. Slope measures the topographic gradient along the steepest path on the land surface, and C_v and C_h represent the curvatures in the direction of the maximum slope and transverse to the slope, respectively. They represent location-specific influences on substance flows. Slope is the first-order derivative of elevation (O'Callaghan and Mark 1984). It determines local velocity of surface substance flows. A soil property correlated to slope means it is sensitive to changes in drainage. Profile and plan curvatures are second-order derivatives of elevation in the direction of maximum slope and transverse to the slope, respectively (Moore et al. 1993). They reflect the concave/convex nature of the land surface (Evans 1972; Schmidt et al. 2003). Flow tends to accelerate when $C_v > 0$ and decelerate when $C_v < 0$, and diverge when $C_h > 0$ and converge when $C_h < 0$ (Florinsky et al. 2002). A combination of slope and curvature features can influence the capacity for water, solute, and sediment conservation in a location: slight slope and negative C_h and/or C_v can hold substance, resulting in its accumulation, while steep slope and positive C_h and/or C_v are poor for substance conservation (Kirkby and Chorley 1967; Burt and Butcher 1985).

Secondary terrain attributes integrate the concepts of specific catchment area (A_s, m^2 m^{-1}) and slope, and are capable of characterizing the spatial variability of specific processes occurring in the landscape (Beven and Kirkby 1979; Moore et al. 1993; Sumfleth and Duttmann 2008). Their definitions are:

$$TWI = \ln\left(\frac{A_s}{\tan\beta}\right) \quad (1)$$

$$SPI = A_s \times \tan\beta \quad (2)$$

$$LS = \left(\frac{A_s}{22.13}\right)^{0.6}\left(\frac{\sin\beta}{0.0896}\right)^{1.3} \quad (3)$$

Secondary terrain attributes are not location-specific but catchment-related, because they integrate the concepts of specific catchment area and slope. Among secondary terrain attributes, *TWI* is an indicator of soil moisture distribution and the extent of flow accumulation at a location, *SPI* is directly proportional to the erosive power of overland flow, and *LS* is indicative of the erosion and deposition processes, especially the effect of topography on soil loss (Beven et al. 1984; Moore et al. 1991; Moore et al. 1993). Soil properties correlated to secondary terrain attributes should be sensitive to soil erosion and substance transport processes.

Fig. 2 Topographic parameters calculated from a digital elevation model (with a 20-m grid) using DigitalHydro: panels **a** to **f** represent the spatial distributions of slope, plan curvature (C_h), profile curvature (C_v), topographic wetness index (*TWI*), stream power index (*SPI*), and length-slope factor (*LS*), respectively

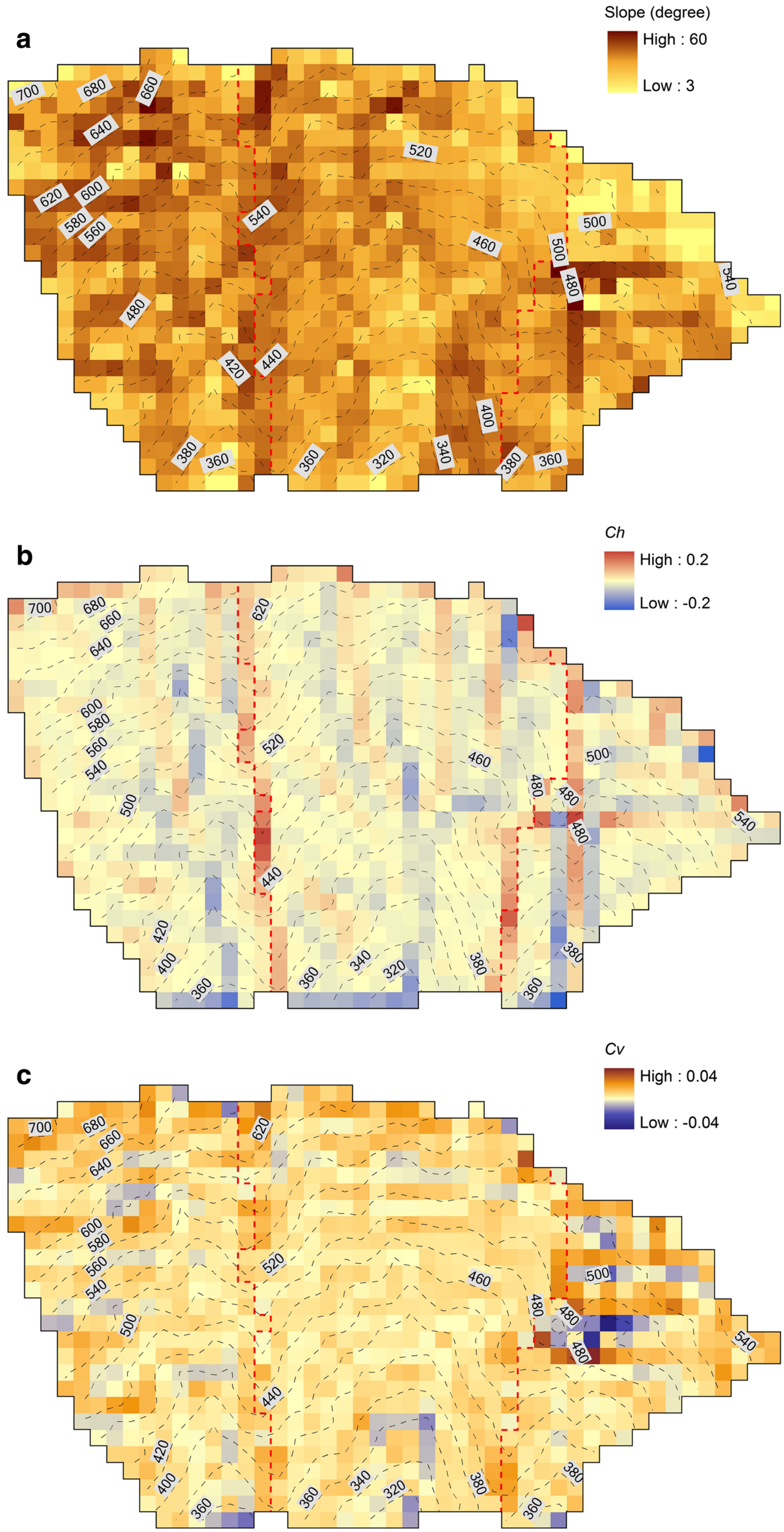
2.3 Conditional inference tree and multiple linear regression methods

A CIT is one of the forms of the decision tree method, which generates tree-structured regression models in a conditional inference framework using binary recursive partitioning. It uses multiple testing to determine when the recursion will stop and overcomes the problem of overfitting. This method is also able to select variables in an unbiased way based on the conditional distribution of statistics measuring the association between response and covariates (Hothorn et al. 2006).

The CIT model could intuitively tell us which topographic parameter has the highest influence on the tested soil property in a dataset (as the root node) and find out a threshold in the selected topographic parameter to optimally split the tested soil property into two disjoint subsets with the largest discrepancy (as two branches). Each internal (non-terminal) node represents a test on the selected soil property, the branches represent the binary partitioned outcomes of a test, and each terminal node represents a class of the response variable showing common influences from some input variables.

The procedure for establishing a topography-soil property CIT model is as follows: (1) the global null hypothesis of independence between each input topographic parameter and a soil property is tested, and the topographic parameter most strongly associated with the soil property is selected if the hypothesis is rejected, or stop splitting if the hypothesis cannot be rejected; (2) the optimal split for the selected topographic parameter is determined using permutation tests, and the optimal split is determined to separate the dataset into two disjoint subsets with the most significant discrepancy; and (3) steps 1 and 2 are repeated until no input topographic parameter can be selected to reject the null hypothesis. A more detailed description of the CIT algorithm can be found in Hothorn et al. (2006). In this study, the CIT-related analyses were conducted in R (R Core Team 2013), using the "party" package (Hothorn et al. 2006). The type of test statistic applied was "quadratic," and the minimum criterion for splitting was set at 0.95.

In this study, the CIT method was adopted to evaluate the degree of influence of topographic parameters (elevation,


a
Slope (degree)
High : 60
Low : 3
b
Ch
High : 0.2
Low : -0.2
c
Cv
High : 0.04
Low : -0.04

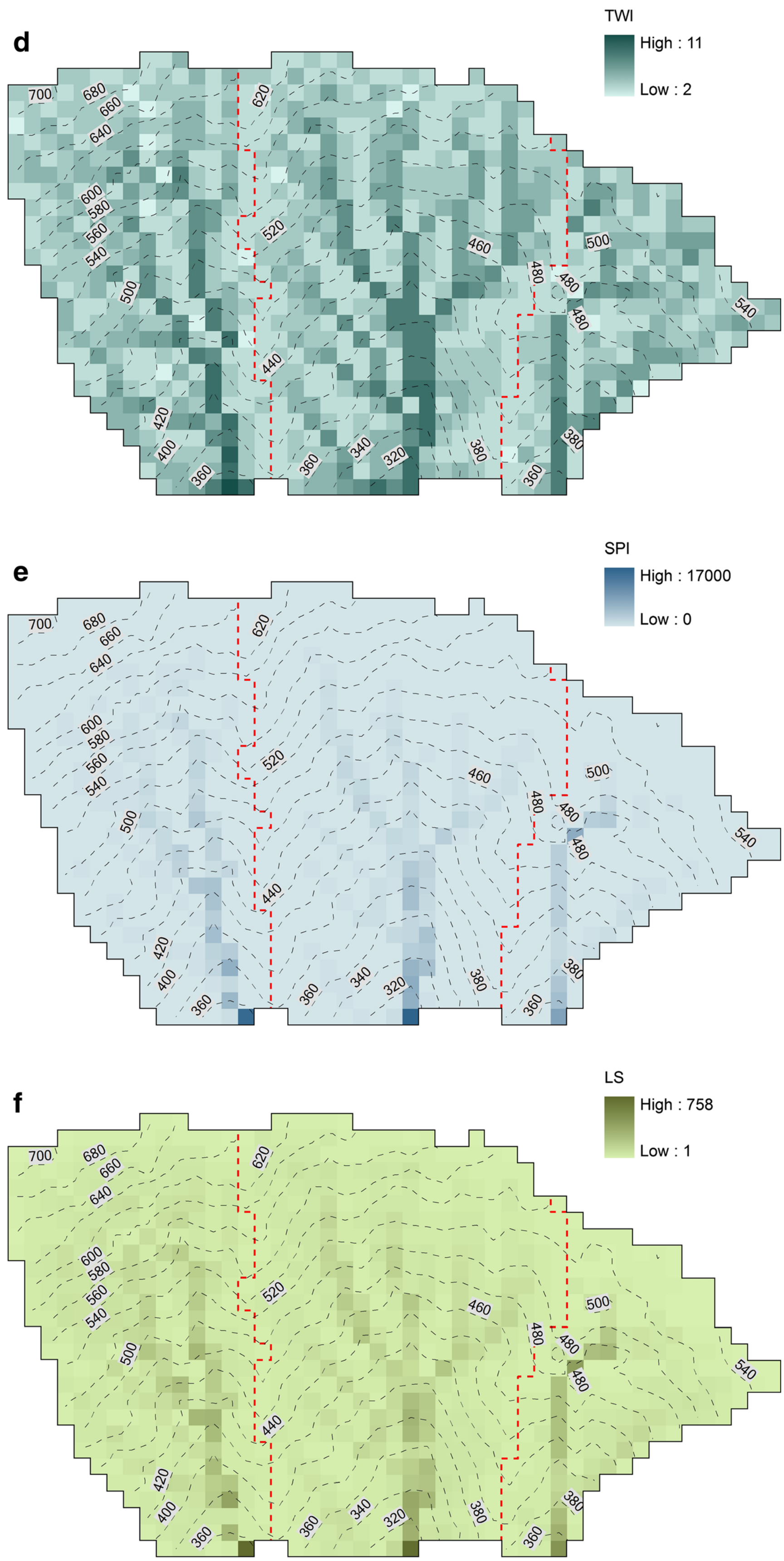


Fig. 2 continued.

slope, C_h, C_v, *TWI*, *SPI*, and *LS*) on soil properties (VWC, SOC, TN, CN ratio, pH, and TP), and a MLR method was also used to compare or validate the performance of the CIT method. In the MLR analysis, the stepwise method was used for independent variable selection, with the probability of the *F* value for entry and removal as 0.05 and 0.10, respectively.

2.4 Evaluation of model performance

The dataset containing 471 data points (samples) was randomly divided into sub-datasets for model development (424 samples) and validation (47 samples). The CIT and MLR models were derived from the model developing sub-dataset, and their performance was evaluated using the validation sub-dataset. The accuracy of estimates was assessed by mean error (ME), mean absolute error (MAE), and the root mean squared errors (RMSE). They are defined as:

$$\mathrm{ME} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left(z_i - \hat{z}_i\right) \quad (4)$$

$$\mathrm{MAE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}\left|z_i, -, \hat{z}_i\right| \quad (5)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}\left(z_i - \hat{z}_i\right)^2} \quad (6)$$

where n is the number of validation points, z_i is the observed soil property, and $\hat{z}_i$ is the predicted soil property by the CIT or MLR models.

3 Results

3.1 Descriptive statistics of topography-soil relationships

The sampled soils, which were loam- to clay-textured Acrisols according to the World Reference Base for Soil Resources (2014), were generally moist, acidic, and rich in organic matter and nutrients. The variation of soil pH values was low, with the coefficient of variation (CV) ≤ 0.10, and the variation of other soil properties at intermediate levels (0.10 < CV ≤ 1.00). Of the six soil properties evaluated, VWC, pH, and CN ratio were normally distributed, while SOC, TN, and TP were lognormally distributed (Table 1).

The range and spatial distribution of topographic parameters are presented in Fig. 2. Of the seven topographic parameters, only elevation followed normal distribution, and *TWI*, *LS*, and *SPI* were strongly non-normal. Since the majority of the soil properties and topographic parameters were in non-normal distribution, Spearman's rank correlation analyses were used to assess their interrelations. All of the evaluated soil properties were significantly correlated to at least one topographic parameter (Table 2). Elevation was the topographic parameter with the highest correlations with VWC (Spearman's $r = -0.33$; same below), SOC (0.33), and TN (0.25), and was also significantly correlated with pH (−0.30) and CN ratio (0.35); C_h had the highest correlations with pH (−0.41) and CN ratio (0.49); and *SPI* was the most significant topographic parameter for TP (0.38). The four primary terrain attributes (elevation, slope, C_h, and C_v) had similar effects on soil properties: negative on VWC and pH, positive on SOC and CN ratio, while effects on TN and TP were variable. Secondary terrain attributes (*TWI*, *SPI*, and *LS*) also had similar effects on soil properties, but their influences were generally the opposite to those of the primary attributes.

3.2 Conditional inference tree and multiple linear regression models for soil properties

The MLR models were built up using the model developing sub-dataset, after the non-normally distributed soil properties were logarithmically transformed (Table 3). All MLR models were significant, with the determination coefficients (R^2) ranging from 0.07 to 0.40. The same dataset was used for CIT model development as for MLR, and the model structures and the spatial distributions of the classifications (terminal nodes) are shown in Fig. 3.

Table 1 Descriptive statistics for sampled soil properties ($n = 471$)

Soil property	Mean	Median	Minimum	Maximum	STDEV	CV	Distribution
VWC ($cm^3\ cm^{-3}$)	0.28	0.28	0.10	0.49	0.06	0.23	Normal
SOC ($g\ kg^{-1}$)	46.50	42.26	13.25	185.44	20.73	0.45	Lognormal
TN ($g\ kg^{-1}$)	3.29	3.05	1.06	11.35	1.21	0.38	Lognormal
CN ratio	14.17	13.99	9.68	20.77	2.10	0.15	Normal
pH	4.14	4.14	3.54	4.78	0.19	0.05	Normal
TP ($g\ kg^{-1}$)	0.26	0.24	0.06	0.77	0.12	0.46	Lognormal

VWC volumetric water content, *SOC* soil organic carbon, *TN* total nitrogen content, *TP* total phosphorus content, *CN ratio* C:N ratio, *STDEV* standard deviation, *CV* coefficient of variation

Table 2 Spearman's rank correlations between soil properties and topographic attributes (n = 471)

Soil/terrain attributes	VWC	pH	TN	TP	SOC	CN ratio	Elevation	Slope	C_h	C_v	*TWI*	*LS*	*SPI*
VWC	1												
pH	0.28**	1											
TN	−0.40**	−0.22**	1										
TP	−0.09	0.024	0.39**	1									
SOC	−0.46**	−0.41**	0.83**	0.18**	1								
CN ratio	−0.31**	−0.47**	0.14**	−0.31**	0.52**	1							
Elevation	−0.33**	−0.30**	0.25**	0.05	0.33**	0.35**	1						
Slope	−0.14**	−0.38**	−0.02	−0.20**	0.20**	0.48**	0.31**	1					
C_h	−0.15**	−0.41**	−0.05	−0.37**	0.18**	0.49**	0.15**	0.49**	1				
C_v	−0.05	−0.17**	−0.06	−0.15**	0.06	0.24**	0.11*	0.41**	0.30**	1			
TWI	0.10*	0.41**	0.04	0.29**	−0.16**	−0.42**	−0.19**	−0.64**	−0.75**	−0.22**	1		
LS	0.08	0.37**	0.14**	0.37**	−0.06	−0.41**	−0.11*	−0.35**	−0.78**	−0.11*	0.80**	1	
SPI	0.09	0.40**	0.14**	0.38**	−0.08	−0.44**	−0.14**	−0.42**	−0.80**	−0.16**	0.83**	0.99**	1

VWC volumetric water content ($cm^3\ cm^{-3}$), *SOC* soil organic carbon ($g\ kg^{-1}$), *TN* total nitrogen ($g\ kg^{-1}$), *TP* total phosphorus ($g\ kg^{-1}$) contents, *CN ratio* C:N ratio, C_h plan curvature, C_v profile curvature, *TWI* topographic wetness index, *LS* length-slope factor, *SPI* stream power index

*statistically significant at $p < 0.05$ level; **statistically significant at $p < 0.01$ level

3.2.1 Volumetric soil moisture content

The CIT and MLR methods selected the same topographic properties, elevation and C_h, as input variables. The CIT model explained 14 % of the variation in soil VWC, while 12 % of the variation was explained by the MLR model. Low soil moisture contents were found in two types of terrains: high elevation (node 5 in Fig. 3a, b, with a mean soil VWC of 0.23 $cm^3\ cm^{-3}$) or low-elevation topographic locations with apparently divergent flow (node 4 in Fig. 3a, b, with a mean soil VWC of 0.23 $cm^3\ cm^{-3}$).

3.2.2 Soil organic C and total N

The SOC and TN contents were highly correlated (Table 2), and both showed similar distribution patterns as VWC (Fig. 3d, f), while SOC was more strongly influenced by topography than TN. In the study, 24 and 16 % of the variation of SOC explained by the CIT and MLR models, respectively, compared to only 8 and 7 %, respectively, of the variation of TN explained. The root node in the CIT model for SOC indicated that the largest difference in SOC content existed between the high- and low-elevation locations (Fig. 3c, d). The CIT model for TN only has two branches with elevation as independent variable and exactly the same threshold as that in the SOC model (508.01 m, Fig. 3e).

3.2.3 Soil CN ratio

The high percentage of the variation of soil CN ratio explained by CIT (45 %) and MLR (40 %) indicates that the CN ratio had the strongest relationship with topography in the area (Tables 2 and 3). The CN ratio also had the most complex

Table 3 Multiple linear regression models describing topography-soil property relationships (n = 424)

Soil property	Model	R^2	Significance
VWC	VWC = −0.001 × elevation − 0.263 × C_h + 0.402	0.12	<0.001
SOC	log(SOC) = 0.001 × elevation +0.731 × C_h + 1.237	0.16	<0.001
TN	log(TN) = −0.002 × slope + 0.001 × elevation + 0.263	0.07	<0.001
CN ratio	CN ratio =0.363 × *TWI* − 0.026 × *LS* + 0.002 × *SPI* + 0.117 × slope + 0.008 × elevation + 19.748 × C_h + 5.610	0.40	<0.001
pH	pH = 0.0005 × elevation − 0.006 × slope − 1.354 × C_h + 4.575	0.26	<0.001
TP	log(TP) = −0.003 × slope − 0.0001 × *SPI* + 0.002 × *LS* − 0.659 × C_h − 0.588	0.14	<0.001

VWC volumetric water content ($cm^3\ cm^{-3}$), *SOC* soil organic carbon ($g\ kg^{-1}$), *TN* total nitrogen ($g\ kg^{-1}$), *TP* total phosphorus ($g\ kg^{-1}$) contents, *CN ratio* C:N ratio, C_h plan curvature, C_v profile curvature, *TWI* topographic wetness index, *LS* length-slope factor, *SPI* stream power index

CIT model structure with 13 nodes including 7 terminal nodes (Fig. 3g). Soils with high CN ratios were found in two disparate terrains: one had parallel or convergent flow ($C_h \leq 0.006$), high elevation (>510 m), and small *TWI* (≤4.07) (node 7 in Fig. 3g, h, with a mean CN ratio of 17.1), while the other had apparently divergent flow (node 13 in Fig. 3g, h, $C_h > 0.049$, with a mean CN ratio of 17.3).

3.2.4 Soil pH

Soil pH also showed a close relationship to topography with 31 and 26 % of its variation explained by the CIT and MLR models, respectively. Both models used the same topographic parameters: C_h, slope, and elevation. The same as soil CN ratio, the root node in the CIT model for soil pH was C_h (threshold 0.007). Low soil pH values were found in two types of topographic locations: locations with apparently divergent flow (node 11 in Fig. 3i, j, with a mean soil pH of 3.90), or locations with both divergent flow and steep slope (node 10 in Fig. 3i, j, with a mean soil pH of 3.86).

3.2.5 Soil TP

The TP model contained more secondary than primary terrain attributes (Fig. 3k, l). The CIT model (22 % of the variation explained) also had a better fit than the MLR model (14 % of the variation explained). The CIT model selected *TWI* as the root splitting factor (threshold 6.08). Soils with high TP generally distributed in locations with large *TWI* (node 7 in Fig. 3k, l, with a mean TP content of 0.33 g kg^{-1}) or with large *LS* (node 6 in Fig. 3k, l, with a mean TP content of 0.31 g kg^{-1}).

3.3 Model performance

As described above, the CIT models established on the subdataset (424 samples) generally explained the topography-soil relationships better than the MLR method. The R^2 of CIT models for various soil properties ranged from 0.08 to 0.45, which are generally higher than those of the MLR models (Table 3). The predictability of the CIT and MLR

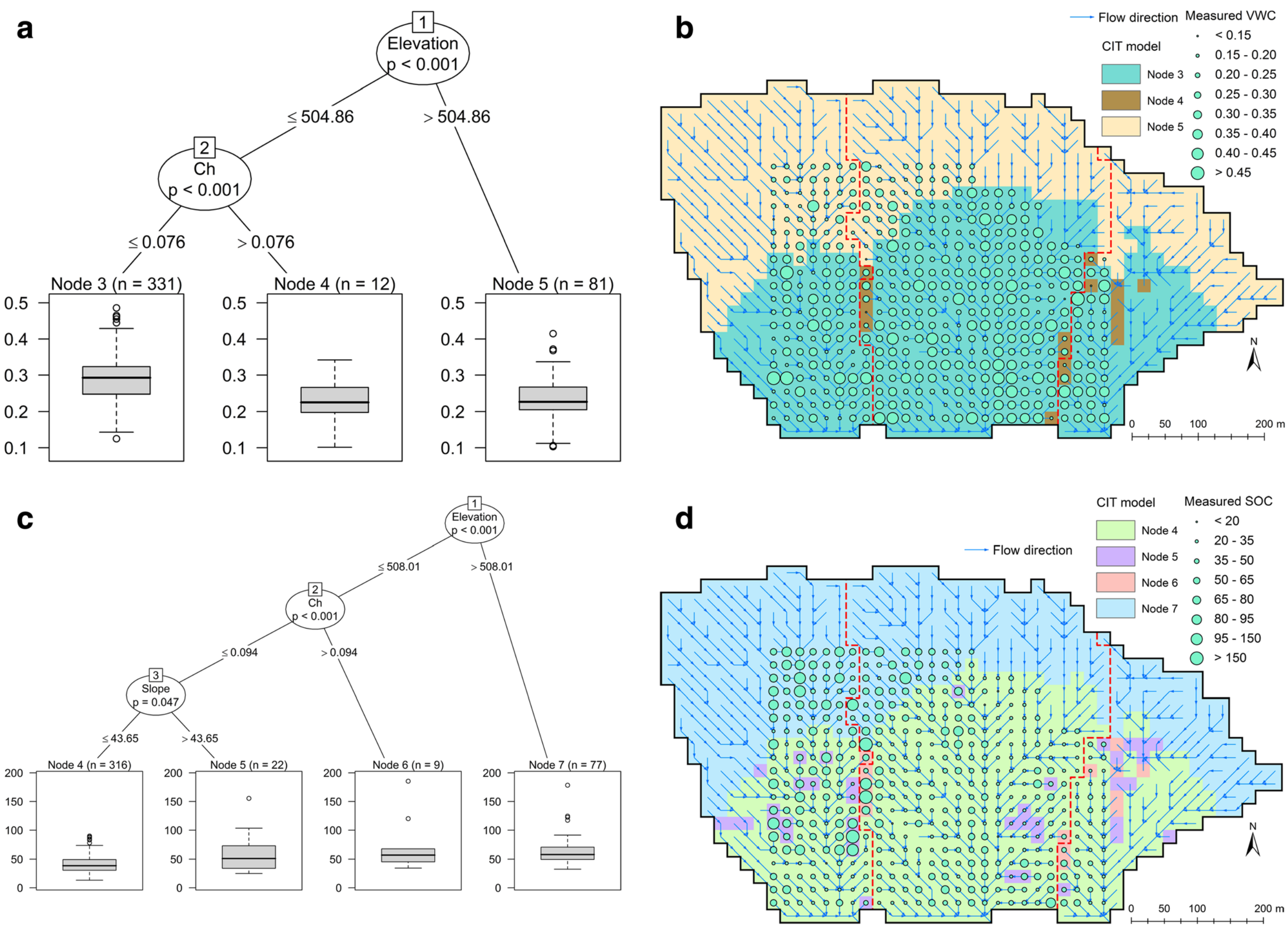


Fig. 3 Conditional inference tree (CIT) models and the spatial distribution of the subsets for **a** and **b** soil volumetric water content (cm^3 cm^{-3}), **c** and **d** organic matter content (g kg^{-1}), **e** and **f** total nitrogen (g kg^{-1}), **g** and **h** C:N ratio, **i** and **j** pH, and **k** and **l** total phosphorus (g kg^{-1}). In each model, *n* is the number of samples in a subset

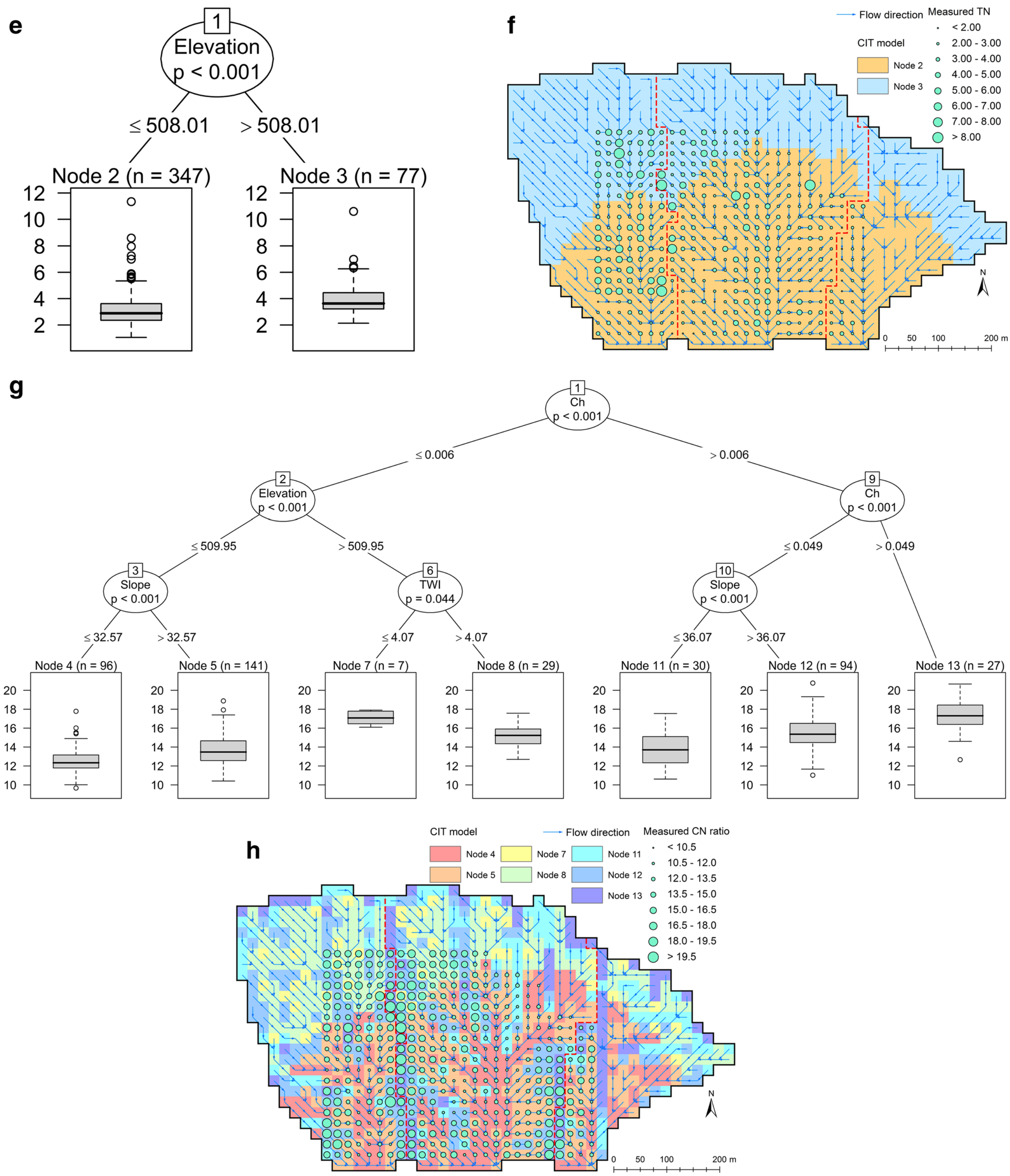


Fig. 3 continued.

models for various soil properties in unknown locations was validated using the validation sub-dataset (47 samples). It showed that the CIT method generally had smaller ME, MAE, and RMSE than the MLR method (Table 4).

4 Discussion

By virtue of the intuitive form of the CIT models, they provide more information on soil-topography relationships than the

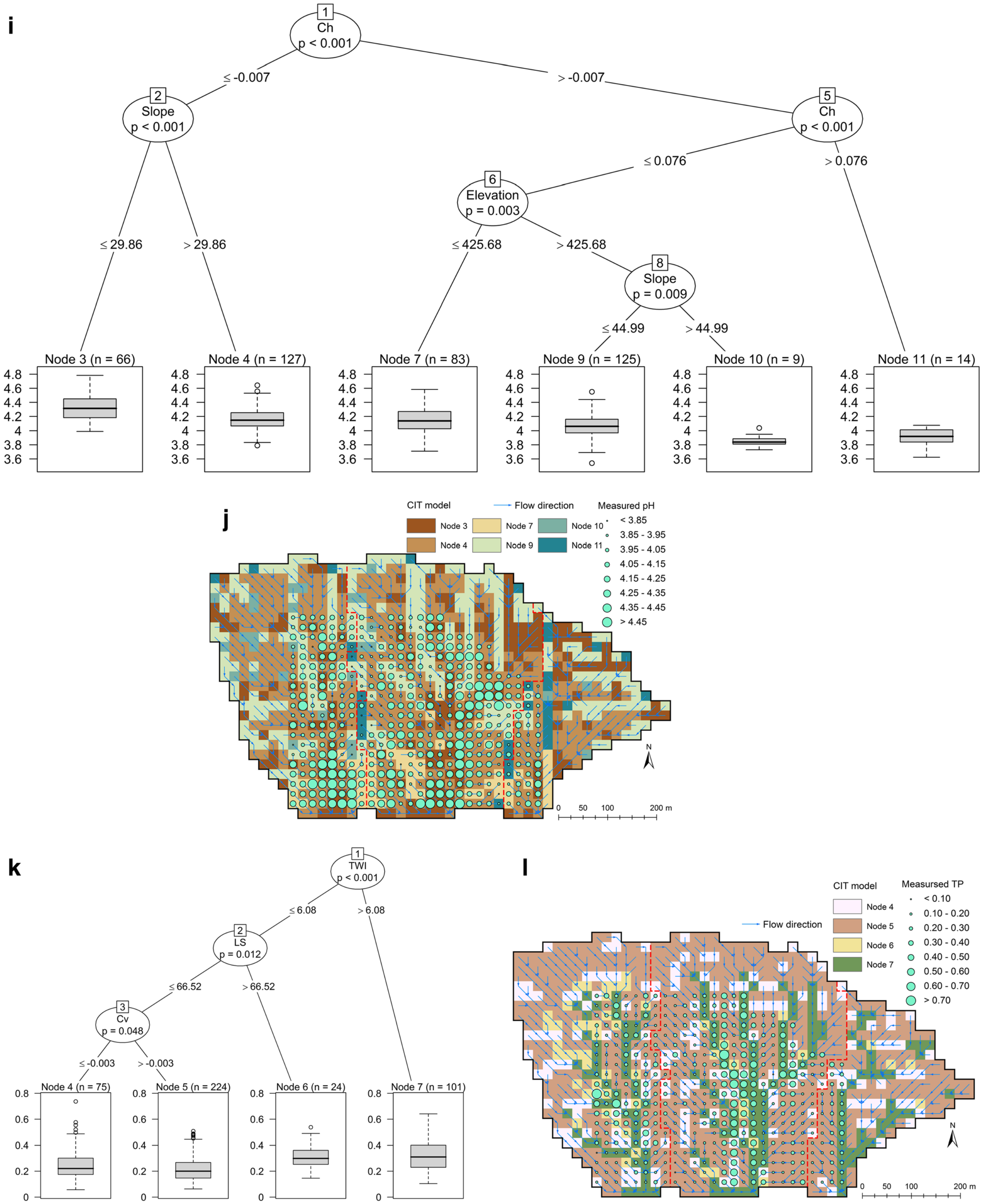


Fig. 3 continued.

MLR method; meanwhile, the CIT models also provide topographic thresholds that have practical meanings and make the results more easily interpreted (Fig. 3). For elevation, the optimal splitting values were consistently between 500 and

Table 4 Validation results for the conditional inference tree and multiple linear regression models

Method	Evaluation index	VWC	SOC	TN	CN ratio	pH	TP
CIT	ME	−0.02	−2.16	−0.10	0.04	−0.02	−0.02
	MAE	0.05	11.98	0.84	1.14	0.13	0.10
	RMSE	0.07	16.75	1.22	1.40	0.17	0.13
MLR	ME	0.02	4.42	0.28	0.03	0.01	0.03
	MAE	0.05	11.32	0.78	1.19	0.13	0.09
	RMSE	0.07	17.37	1.23	1.49	0.17	0.13

VWC volumetric water content (cm^3 cm^{-3}), *SOC* soil organic carbon (g kg^{-1}), *TN* total nitrogen (g kg^{-1}), *TP* total phosphorus (g kg^{-1}) contents, *CN ratio* C:N ratio, *ME* mean error, *MAE* mean absolute error, *RMSE* root mean squared error

510 m in the VWC, SOC, TN, and CN ratio models, indicating that the largest differences in soil properties existed above and below this elevation. Referring to the local average elevation of 493.34 m, this height should be the division for high and low elevation. For slope, the consistent split value of about 45° in the SOC and pH models represents the divide for moderate and steep terrains, and correspondingly, the splits of about 30° to 35° in the pH and CN ratio models represent the divides for mild, mild-moderate, and moderate terrains, respectively. And for C_h, some splitting values in the SOC, CN ratio, and pH models were found close to 0 ($-0.01 \leq C_h \leq 0.01$). Plan curvatures in this range are expected to result in parallel flow. Larger and smaller than it will cause convergent and divergent flow, respectively.

From the CIT and MLR results, it can be easily discovered that soil VWC, SOC, and TN were significantly influenced by primary terrain attributes, soil CN ratio and pH were affected by both primary and secondary terrain attributes, and soil TP was markedly affected by secondary terrain attributes. The different model structures indicate that the mechanisms for topography to affect soil properties vary with each soil property.

4.1 Mechanisms affecting soil VWC, SOC, and TN

There are two possible mechanisms for elevation to affect soil moisture content: (1) in hilly areas, usually precipitation increases and temperature decreases (reducing evaporation) with elevation, resulting in high soil moisture content in high-elevation sites and (2) downslope flow and gravitational redistribution of water tend to result in a decreasing pattern of soil moisture content with elevation, especially in humid regions (Zhu and Lin 2011). The negative correlation between VWC and elevation and the strong relationship between VWC and primary terrain attributes in the area suggest that the second mechanism was more important to influence the spatial pattern of soil water content (Table 2). The low soil moisture content along the ridges (with positive C_h) is attributed to the poor water conservation capacity of the ridges (Kirkby and Chorley 1967; Burt and Butcher 1985; Qiu et al. 2001).

Soil organic C content is the balance between C input and output (Schimel et al. 1994; Stallard 1998; Polyakov and Lal 2004). In this study, soils in high-elevation sites generally had higher SOC, associated with lower soil moisture content and temperature, and thus lower organic matter decomposition rates (Fig. 3c, d). The stronger correlation of SOC content with primary than with secondary terrain attributes (Table 2) indicates that SOC in the area was not strongly affected by catchment-related erosion processes, but by local conditions such as moisture availability, temperature, and plant species composition; the amount of soil C lost through runoff should be much less than the loss through decomposition.

There is no consensus as to how SOC content is impacted by topographic features (Moore et al. 1993; Terra et al. 2004; Zehetner and Miller 2006; Chai et al. 2008; Hancock et al. 2010; Feng et al. 2011; Zhang et al. 2012); however, human disturbance or changes in land uses may play a deterministic role. Frequent disturbance increases SOC decomposition rates and soil erosion in cultivated lands (Pennock et al. 1994; Shukla and Lal 2005). In undisturbed systems (forest and grassland, either in boreal or tropical areas), as the result in this study, SOC is typically positively correlated with elevation (Schimel et al. 1985; Johnson et al. 2000; Luizão et al. 2004; Tsui et al. 2004; Umali et al. 2010), and more strongly affected by primary terrain attributes (especially elevation) (Hancock et al. 2010; Zhang et al. 2012). However, the opposite is true in cultivated systems: SOC was generally less related to elevation (and other primary terrain attributes) but more to secondary terrain attributes such as *TWI* (Moore et al. 1993; Florinsky et al. 2002; Terra et al. 2004; Sumfleth and Duttmann 2008; Dlugoß et al. 2010).

The weaker correlation between TN (as compared to SOC) and topographic parameters (Fig. 3e, f and Table 2) suggests that biophysical factors such as vegetation, microbial activities, and soil texture but not topography were the major control on soil TN (Luizão et al. 2004).

4.2 Mechanisms affecting soil CN ratio and pH

Soil CN ratio and pH were strongly correlated with both primary and secondary terrain attributes, suggesting that both location-specific properties and catchment-related hydrological processes strongly affected their distribution on the landscape.

Soil CN ratio is an indicator of the degree of soil organic matter decomposition (Garten et al. 1994), and is also influenced by soil erosion (Seibert et al. 2007). In this study, soil CN ratio showed a stronger correlation with topography than with SOC and TN, similar to Garten et al. (1994), Tokuchi et al. (1999), and Seibert et al. (2007). Soils with high CN ratios in locations with non-divergent (node 7 in Fig. 3g, h) and divergent (node 13 in Fig. 3g, h) flows could be explained by two mechanisms: (1) a low degree of organic matter decomposition leads to high CN ratio. The high CN ratios in locations belonging to node 7, which usually had low soil temperature, moisture content, and correspondingly low organic matter decomposition rate and high soil CN ratios, should be caused by this (Yimer et al. 2006), and (2) soil CN ratio can also be increased by heavy flux of water that washes away NH_4^+ and NO_3^- (Seibert et al. 2007). Locations belonging to node 13 had the greatest water loss and soil erosion potentials in the area. Soil mineral N loss would be the main factor contributing to the high CN ratios in these locations. The actual distribution of CN ratio should be jointly affected by both mechanisms, and result in a strong topography-CN ratio correlation.

Soil pH is influenced by several factors such as soil organic matter content and decomposition rate, cation exchange and leaching (Tokuchi et al. 1999; Seibert et al. 2007), and influence of the vegetation (Jung et al. 2011; Jung and Chang 2013). In this study, soils with high pH were often found in locations with convergent flow, mild slope, or low elevation (Fig. 3i, j), topographic features that tend to be associated with high soil temperature and moisture content, where organic matter decomposition tends to be fast and more complete, resulting in low SOC content. As a result, the amount of H^+ release to soils is reduced, decreasing soil acidity. Another potential reason was that base cations taken up by the trees from deeper soil layers were released to the soil through the decomposition of plant residues, and increases soil pH (Seibert et al. 2007).

4.3 Mechanisms affecting soil TP

Phosphorus is easier to migrate with water in suspended and dissolved forms, and then it is more influenced by hydrological processes at the catchment scale (Roberts et al. 1985; Honeycutt et al. 1990; Gburek et al. 2002). The distribution of soil TP in relation to topographic features in this study is consistent with that reported by McDowell and Srinivasan (2009), in which TP was more strongly correlated to secondary terrain attributes such as *TWI* and *LS* in some small catchments. Therefore, the spatial distribution of soil TP is considered to be a reliable indicator of water movement in the landscape (Smeck and Runge 1971; Roberts et al. 1985; Gburek and Sharpley 1998). However, the relationship between topography and soil phosphorus status may be weakened by other factors such as soil disturbance (Page et al. 2005).

5 Conclusions

We conclude that the CIT method not only provided better fitted models for topography-soil relationships but also had less prediction errors than the MLR method. Based on the CIT analysis, soil properties were classified into three categories according to the way they were affected by topography: those that were highly influenced by primary terrain attributes (such as VWC, SOC, and TN), by secondary terrain attributes (TP), or by both (CN ratio and pH). The distribution of VWC, SOC, and TN on the landscape was more affected by location-specific features such as differences in gravitational potential, precipitation, temperature, and vegetation in association with changes in topography. The spatial distribution of soil TP was mostly affected by catchment-related hydrological processes, while those of CN ratio and pH were affected by topography in both location-specific and catchment-related ways, resulting in the strongest topography-soil relationships in the study. The topography-soil property relationships generalized in this study should help us evaluate topography-soil property relationships in similar regions and help with ecological health evaluation and ecological restoration of degraded lands in the southeast hilly region in China, a region that is severely disturbed by anthropogenic activities.

Acknowledgments We thank the National Science Foundation of China (31210103920), Chinese Academy of Sciences and Technology of Strategic Pilot Projects (XDA0505020401), and the Natural Science and Engineering Research Council of Canada for funding.

References

Barthold FK, Stallard RF, Elsenbeer H (2008) Soil nutrient-landscape relationships in a lowland tropical rainforest in Panama. Forest Ecol Manag 255:1135–1148

Beven K, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. Hydrol Sci J 24:43–69

Beven K, Kirkby M, Schofield N, Tagg A (1984) Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. J Hydrol 69:119–143

Bramer M (2007) Principles of data mining. Springer, London

Buol SW, Southard RJ, Graham RC, McDaniel PA (2011) Soil genesis and classification. John Wiley & Sons, Inc

Burt T, Butcher D (1985) Topographic controls of soil moisture distributions. J Soil Sci 36:469–486

Chai X, Shen C, Yuan X, Huang Y (2008) Spatial prediction of soil organic matter in the presence of different external trends with REML-EBLUP. Geoderma 148:159–166

Chen Y, Wang YL (2003) Soil erosion and countermeasures in hilly region of southeast China (in Chinese with English abstract). Chinese J Trop Agric 23:38–42

Chen ZS, Hsieh CF, Jiang FY, Hsieh TH, Sun IF (1997) Relations of soil properties to topography and vegetation in a subtropical rain forest in southern Taiwan. Plant Ecol 132:229–241

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192

Dlugoß V, Fiener P, Schneider K (2010) Layer-specific analysis and spatial prediction of soil organic carbon using terrain attributes and erosion modeling. Soil Sci Soc Am J 74:922–935

Evans IS (1972) General geomorphometry, derivatives of altitude, and descriptive statistics. In: Chorley RJ (ed) Spatial analysis in geomorphology. Harper & Row, London, pp. 17–90

Feng D, Liu J, Chen X (2011) Spatial variation of hillslope soil chemical attributes (in Chinese with English abstract. J Mt Sci 29:427–432

Florinsky IV, Eilers RG, Manning GR, Fuller LG (2002) Prediction of soil properties by digital terrain modelling. Environ Model Softw 17:295–311

Garcia-Pausas J, Casals P, Camarero L, Huguet C, Sebastià M, Thompson R, Romanyà J (2007) Soil organic carbon storage in mountain grasslands of the Pyrenees: effects of climate and topography. Biogeochemistry 82:279–289

Garten CT, Huston MA, Thoms CA (1994) Topographic variation of soil nitrogen dynamics at Walker Branch Watershed, Tennessee. For Sci 40:497–512

Gburek WJ, Sharpley AN (1998) Hydrologic controls on phosphorus loss from upland agricultural watersheds. J Environ Qual 27:267–277

Gburek WJ, Drungil CC, Srinivasan MS, Needelman BA, Woodward DE (2002) Variable-source-area controls on phosphorus transport: bridging the gap between research and design. J Soil Water Conserv 57:534–543

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

Gessler PE, Chadwick OA, Chamran F, Althouse L, Holmes K (2000) Modeling soil-landscape and ecosystem properties using terrain attributes. Soil Sci Soc Am J 64:2046–2056

Gmur S, Vogt D, Zabowski D, Moskal LM (2012) Hyperspectral analysis of soil nitrogen, carbon, carbonate, and organic matter using regression trees. Sensors 12:10639–10658

Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89:1–45

Hancock GR, Murphy D, Evans KG (2010) Hillslope and catchment scale soil organic carbon concentration: an assessment of the role of geomorphology and soil erosion in an undisturbed environment. Geoderma 155:36–45

Honeycutt C, Heil R, Cole C (1990) Climatic and topographic relations of three Great Plains soils: II. Carbon, nitrogen, and phosphorus. Soil Sci Soc Am J 54:476–483

Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 15:651–674

Hu Y, Cheng H (2013) Application of stochastic models in identification and apportionment of heavy metal pollution sources in the surface soils of a large-scale region. Environ Sci Technol 47:3752–3760

Johnson CE, Ruiz-Méndez JJ, Lawrence GB (2000) Forest soil chemistry and terrain attributes in a Catskills watershed. Soil Sci Soc Am J 64:1804–1814

Jung K, Chang SX (2013) Soil and tree chemistry reflected the cumulative impact of acid deposition in *Pinus banksiana* and *Populus tremuloides* stands in the Athabasca oil sands region in western Canada. Ecol Indic 25:35–44

Jung K, Chang SX, Arshad MAC (2011) Effects of canopy-deposition interaction on H+ supply to soils in *Pinus banksiana* and *Populus tremuloides* ecosystems in the Athabasca oil sands region in Alberta, Canada. Environ Pollut 159:666–673

Kirkby M, Chorley R (1967) Throughflow, overland flow and erosion. Hydrol Sci J 12:5–21

Liu J (2012) Digital Drainage Network Automatic Extraction Software, DigitalHydro V1.1, Copyright no. 2012SR020314, China

Liu J, Chen X, Lin H, Liu H, Song H (2013) A simple geomorphic-based analytical model for predicting the spatial distribution of soil thickness in headwater hillslopes and catchments. Water Resour Res 49:7733–7746

Luizão RCC, Luizão FJ, Paiva RQ, Monteiro TF, Sousa LS, Kruijt B (2004) Variation of carbon and nitrogen cycling processes along a topographic gradient in a central Amazonian forest. Glob Chang Biol 10:592–600

Mage SM, Porder S (2012) Parent material and topography determine soil phosphorus status in the Luquillo Mountains of Puerto Rico. Ecosystems 16:284–294

McDowell R, Srinivasan M (2009) Identifying critical source areas for water quality: 2. Validating the approach for phosphorus and sediment losses in grazed headwater catchments. J Hydrol 379:68–80

McKenzie NJ, Ryan PJ (1999) Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94

Mertens M, Nestler I, Huwe B (2002) GIS-based regionalization of soil profiles with classification and regression trees (CART. J Plant Nutr Soil Sci 165:39

Montgomery K (2006) Variation in temperature with altitude and latitude. J Geogr 105:133–135

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5:3–30

Moore ID, Gessler PE, Nielsen GA, Peterson GA (1993) Soil attribute prediction using terrain analysis. Soil Sci Soc Am J 57:443–452

Motaghian H, Mohammadi J (2011) Spatial estimation of saturated hydraulic conductivity from terrain attributes using regression, kriging, and artificial neural networks. Pedosphere 21:170–177

O'Callaghan JF, Mark DM (1984) The extraction of drainage networks from digital elevation data. Comput Vision Graph 28:323–344

Page T, Haygarth PM, Beven KJ, Joynes A, Butler T, Keeler C, Freer J, Owens PN, Wood GA (2005) Spatial variability of soil phosphorus in relation to the topographic index and critical source areas. J Environ Qual 34:2263–2277

Pennock D, Anderson D, De Jong E (1994) Landscape-scale changes in indicators of soil quality due to cultivation in Saskatchewan, Canada. Geoderma 64:1–19

Polyakov V, Lal R (2004) Modeling soil organic matter dynamics as affected by soil water erosion. Environ Int 30:547–556

Qiu B, Zhong M (2013) Spatiotemporal variability of vegetation phenology with reference to altitude and climate in the subtropical mountain and hill region, China. Chin Sci Bull 58:2883–2892

Qiu Y, Fu B, Wang J, Chen L (2001) Spatial variability of soil moisture content and its relation to environmental indices in a semi-arid gully catchment of the Loess Plateau, China. J Arid Environ 49:723–750

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Roberts T, Stewart J, Bettany J (1985) The influence of topography on the distribution of organic and inorganic soil phosphorus across a narrow environmental gradient. Can J Soil Sci 65:651–665

Schimel D, Stillwell M, Woodmansee R (1985) Biogeochemistry of C, N, and P in a soil catena of the shortgrass steppe. Ecology 66:276–282

Schimel DS, Braswell BH, Holland EA, McKeown R, Ojima DS, Painter TH, Parton WJ, Townsend AR (1994) Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. Global Biogeochem Cy 8:279–293

Schmidt J, Evans IS, Brinkmann J (2003) Comparison of polynomial models for land surface curvature calculation. Int J Geogr Inf Sci 17:797–814

Seibert J, Stendahl J, Sørensen R (2007) Topographical influences on soil properties in boreal forests. Geoderma 141:139–148

Shih YS (2004) A note on split selection bias in classification trees. Comput Stat Data An 45:457–466

Shukla M, Lal R (2005) Erosional effects on soil organic carbon stock in an on-farm study on Alfisols in west central Ohio. Soil Till Res 81: 173–181

Smeck NE, Runge E (1971) Phosphorus availability and redistribution in relation to profile development in an Illinois landscape segment. Soil Sci Soc Am J 35:952–959

Song YC, Wang XR (1995) The vegetation and flora in Zhejiang Tiantong National Forest Park (In Chinese). Shanghai Scientific and Technical Literature Publishing House, Shanghai

Stallard RF (1998) Terrestrial sedimentation and the carbon cycle: coupling weathering and erosion to carbon burial. Global Biogeochem Cy 12:231–257

Sumfleth K, Duttmann R (2008) Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. Ecol Indic 8:485–501

Terra JA, Shaw JN, Reeves DW, Raper RL, van Santen E, Mask PL (2004) Soil carbon relationships with terrain attributes, electrical conductivity, and a soil survey in a coastal plain landscape. Soil Sci 169:819–831

Tittonell P, Shepherd K, Vanlauwe B, Giller K (2008) Unravelling the effects of soil and crop management on maize productivity in smallholder agricultural systems of western Kenya—an application of classification and regression tree analysis. Agric Ecosyst Environ 123:137–150

Tokuchi N, Takeda H, Yoshida K, Iwatsubo G (1999) Topographical variations in a plant-soil system along a slope on Mt Ryuoh, Japan. Ecol Res 14:361–369

Tso GK, Yau KK (2007) Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy 32:1761–1768

Tsui CC, Chen ZS, Hsieh CF (2004) Relationships between soil properties and slope position in a lowland rain forest of southern Taiwan. Geoderma 123:131–142

Umali B, Chittleborough D, Kookana R, Ostendorf B (2010) DEM and terrain analysis to predict spatial pattern of SOC, 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Austrilia

Vega FA, Matias JM, Andrade ML, Reigosa MJ, Covelo EF (2009) Classification and regression trees (CARTs) for modelling the sorption and retention of heavy metals by soil. J Hazard Mater 167:615–624

Wang DL (2011) Relationships between the community characteristics and topography parameters of evergreen broad-leaved forest in Tiantong's 20 ha dynamic plot. Eastern China, East China Normal University, Shanghai, Zhejiang Province

Yang ZJ, Cui JC, Xie JS, Chen GS, Yang YS (2010) Effects of land use/cover chang on soil properties in mid-subtropical mountainous area of southern China (in Chinese with English abstract. Sci Geogr Sin 30:475–480

Yimer F, Ledin S, Abdelkadir A (2006) Soil organic carbon and total nitrogen stocks as affected by topographic aspect and vegetation in the Bale Mountains, Ethiopia. Geoderma 135:335–344

Zehetner F, Miller W (2006) Soil variations along a climatic gradient in an Andean agro-ecosystem. Geoderma 137:126–134

Zhang S, Huang Y, Shen C, Ye H, Du Y (2012) Spatial prediction of soil organic matter using terrain indices and categorical variables as auxiliary information. Geoderma 171–172:35–43

Zhu Q, Lin H (2011) Influences of soil, terrain, and crop growth on soil moisture variation from transect to farm scales. Geoderma 163:45–54

Ziadat FM (2005) Analyzing digital terrain attributes to predict soil attributes for a relatively large area. Soil Sci Soc Am J 69:1590–1599