Kasetsart J. (Nat. Sci.) 47 : 358 - 373 (2013)

Spatial Modeling for Soil Properties Prediction in Mountainous Areas Using Partial Least Squares Regression

Suwit Ongsomwang[1,*] and Rawee Rattanakom

ABSTRACT

Soil properties are one of the most important categories of information for land management and environmental modeling. Unfortunately, soil properties in mountainous areas with slopes of more than 35% are rarely investigated in Thailand due to the complexity of their landscapes and the cost and time requirements. The main objective was to predict soil properties in mountainous areas relating to soil forming factors using partial least squares regression (PLSR). The combination of topographic position index values from two different scales and criteria sets was firstly used to classify landform for *in situ* soil survey. Then, analyzed soil properties of the topsoil and subsoil (sand, silt, clay, pH, organic matter, total N, available P, exchangeable K, cation exchange capacity (CEC) and base saturation) and soil forming factors (rainfall, normalized difference vegetation index, elevation, slope, aspect, plan curvature, profile curvature, curvature, topographic wetness index and Al/Si ratio) were used to construct soil-landscape models using PLSR. It was found that the best predictive model for topsoil prediction was sand (R^2 = 0.92) and the worst was silt (R^2 = 0.52) while the best predictive model for subsoil property prediction was CEC (R^2 = 0.85) and the worst was total N and available P (R^2 = 0.59). Accuracy assessment for the topsoil and subsoil properties prediction models using normalized root mean square error varied between 0.18 to 0.25 and 0.18 to 0.36, respectively. In addition, the selected predictive soil properties were used for soil texture classification and soil fertility assessment. In conclusion, it is suggested that soil-landscape modeling using PLSR can be efficiently used as a tool for spatial soil property prediction in mountainous areas where soil characteristics and properties are not available.

Keywords: landform classification, soil forming factor, soil property prediction, spatial modeling

INTRODUCTION

Soil properties are one of the important information categories for land management and environmental modeling (Florinsky *et al*., 2002; Herbst *et al*., 2006; Ziadat, 2007; Boettinger, 2010). Unfortunately, soil properties in mountainous areas with slopes of more than 35% are rarely investigated in Thailand due to the complexity of their landscape. These complexities in mapping soils and their properties mean it is difficult to classify and map large uniform units. Therefore, the soil maps in the mountainous areas of Thailand are mostly described as slope complex (SC) or Soil Unit 62 where soil characteristics and properties are not available (Land Development Department, 1992). In addition, soil properties extraction in these areas requires a lot of time and money by using conventional soil survey techniques. However, there is a close relationship

[1] School of Remote Sensing, Institute of Science, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand.
* Corresponding author, e-mail: suwit@sut.ac.th

Received date : 03/01/13 Accepted date : 03/04/13

between the soil properties and soil forming factors in the soils in mountainous areas (Gessler *et al*., 1995; Gobin *et al*., 2001). In these areas, the topography is often used in soil studies including the modeling and prediction of soil properties and it is overwhelming and influences most of the other soil forming factors (Jenny, 1980; Pennock *et al*. 1987; Moore *et al*., 1993; McKenzie *et al*., 2000; Ballabio, 2009).

Understanding the soil distribution patterns in relation to landscape attributes is seen as a step to improve the accuracy of prediction of soil properties at unsampled locations. However, this variation is not random because the properties of soil vary from place to place. Natural soil bodies are the result of climate and living organisms acting on parent material, with topography or local relief exerting a modifying influence and with time required for soil-forming processes to act (Soil Survey Division Staff, 1993).These relationships are ideal for the application of regression techniques to predict soil properties. Thus, understanding the soil distribution pattern in relationship to the soil properties and their soil forming factors is very important for the prediction of soil properties in mountainous areas.

During the last decade, several studies have attempted to characterize and predict the spatial distribution of soil properties using more readily available soil forming factors or environmental variables, namely soil-landscape modeling (Huggett, 1975; Moore *et al*., 1993; Odeh *et al*., 1995; Gessler *et al*., 1995; Thompson *et al*., 1997; Gessler *et al*., 2000; Wilson and Gallant, 2000; Gobin *et al*., 2001; Grunwald, 2006). This modeling has been developed as a quantitative method to predict patterns of soil properties from observed patterns in soil forming factors. The main advantage is the improvement in soil information and the reduced cost and time involved in field sampling (Thompson *et al*., 2006).

The objectives of the current study were to classify landforms and to generate soil forming factors using geoinformatics, to quantify the relationship between soil properties and soil forming factors using partial least squares regression (PLSR) and by application of the results to predict soil properties in a mountainous area.

STUDY AREA

The Mae Sa watershed was chosen as the study area as it represents a site of small-scale rural development and integrated watershed management in a mountainous area. It covers 138.85 km^2 in Chiang Mai province, northern Thailand. The watershed is an upland area with mountainous terrain and an altitude range from 300 to 1,600 m above mean sea level. Most soils in this area (about 70%) have been classified as SC (Land Development Department, 1992) as shown in Figure 1.

According to the geological map (Department of Mineral Resources, 2006), the petrography of the study area consists of 67.16% Triassic granites and 32.84% Precambrian gneiss. The dominant land use and land cover types are forest land covering an area of 73.10% including hill evergreen (17.03%), mixed deciduous (39.68%), and dry dipterocarp forests (16.40%), and agricultural land covering an area of 24.09% (Royal Forest Department, 2007). The mean annual rainfall for a 10-year period (2000–2009) was 1,267 mm (Thai Meteorological Department, 2010).

MATERIALS AND METHODS

The methodology of spatial modeling for soil properties prediction using PLSR consisted of four components: 1) landform classification, 2) soil sampling unit identification and soil sample data collection and analysis, 3) soil forming factor generation and 4) soil-landscape model development and its application (Figure 2).

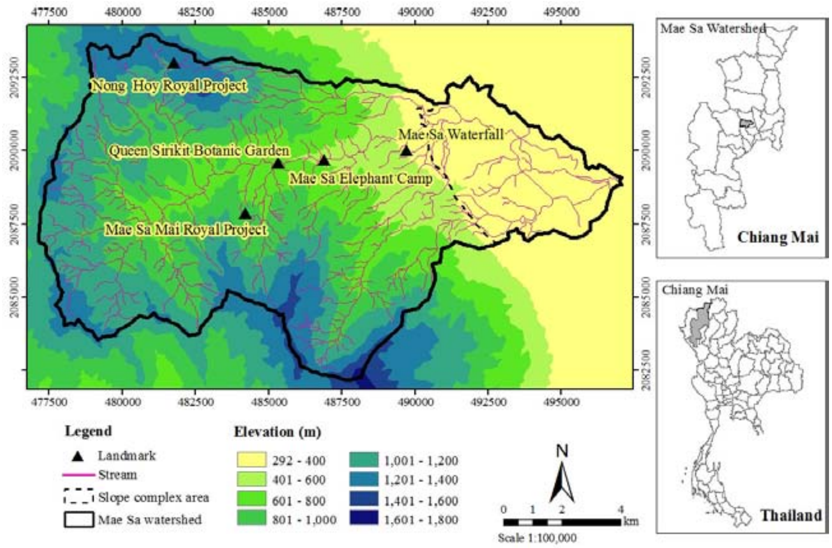

Figure 1 Distribution of slope complex in the study area, Mae Sa Watershed, Chiang Mai province.

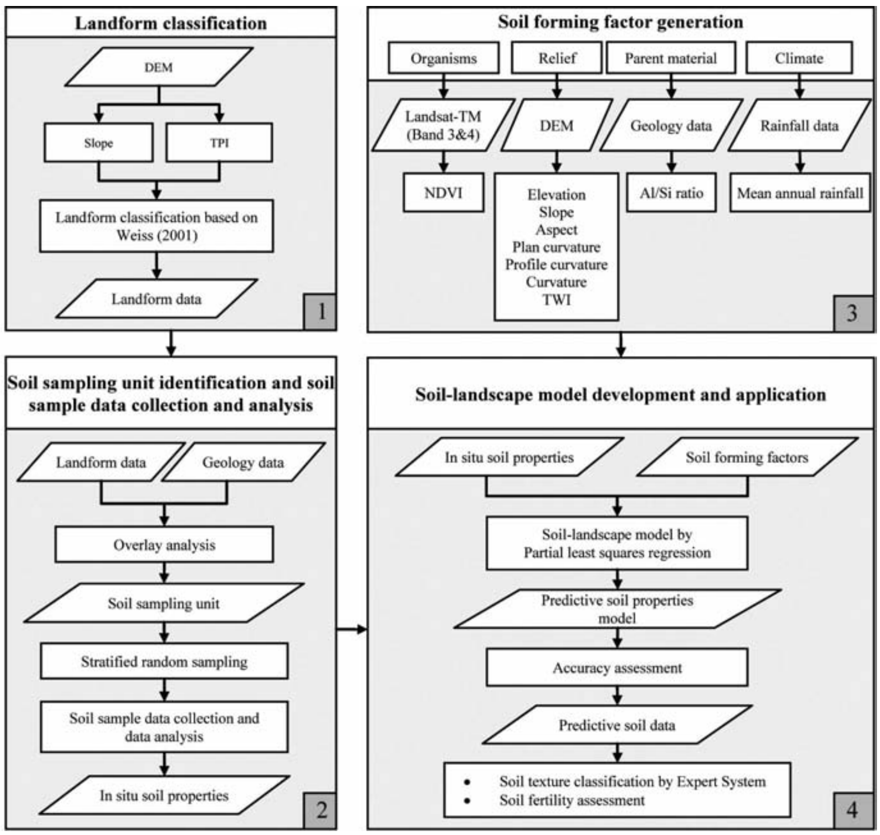

Figure 2 Research methodology and workflow for: 1) landform classification, 2) soil sampling unit identification and soil sample data collection and analysis, 3) soil forming factor generation and 4) soil-landscape model development and its application (DEM = Digital elevation model, TPI = Topographic position index, NDVI = Normalized difference vegetation index).

Landform classification

The input data for landform classification were the slope and topographic position index (TPI) at two scales (small kernel of 15 × 15 cells and large kernel of 45 × 45 cells), which were extracted from a digital elevation model (DEM) with a spatial resolution of 25 × 25 m. In principle, TPI compares the elevation of each cell in a DEM to the mean elevation of a specified neighborhood around that cell. Positive TPI values represent locations that are higher than the average of their surroundings (tending to be hilltops) while negative TPI values represent locations that are lower than their surroundings (tending to be valleys) and TPI values near zero are normally flat areas (Weiss, 2001). The combination of TPI values from different scales (kernel size) and criteria sets suggests various landform types. In practice, two data sets of TPI, which were extracted at two different scales, were reclassified using the standard deviation (SD) into three categories: 1) TPI with standard deviation value less than or equal to -1; 2) TPI with standard deviation value between -1 and 1; and 3) TPI with standard deviation value greater than or equal to 1.

Then, the derived TPI data were overlaid with slope data, which were classified into two classes (less than or equal to 5 degrees and greater than 5 degrees), for landform classification. In this study, the criteria sets for landform classification were modified from Weiss (2001) as shown in Table 1.

Table 1 Landform category and criteria (modified from Weiss, 2001).

No	Landform category	Criteria	Description
1	Canyons, Deeply incised streams	Small scale TPI: TPI ≤ -1 SD Large scale TPI: TPI ≤ -1 SD	Areas are lowest in the landscape, having negative plan and/or profile curvature
2	Midslope drainages, Shallow valleys	Small scale TPI: TPI ≤ -1 SD Large scale TPI: -1 SD < TPI < 1 SD	Areas are low in mid slope, channel in mid slope
3	Upland drainages, Headwaters	Small scale TPI: TPI ≤ -1 SD Large scale TPI: TPI ≥ 1 SD	Areas are low in upper slope channel in upper slope
4	U-shaped valleys	Small scale TPI: -1 SD < TPI < 1 SD Large scale TPI: TPI ≤ -1 SD	Areas in lower slope, footslope adjacent below an open slope and adjacent above a flat or streams
5	Plains	Small scale TPI: -1 SD < TPI < 1 SD Large scale TPI: -1 SD < TPI < 1 SD Slope ≤ 5°	Areas are flat having a slope ≤ 5°
6	Open slopes	Small scale TPI: -1 SD < TPI < 1 SD Large scale TPI: -1 SD < TPI < 1 SD Slope > 5°	Areas are rectilinear transition in mid slope, having a slope > 5°
7	Upper slopes, Mesas	Small scale TPI: -1 SD < TPI < 1 SD Large scale TPI: TPI ≥ 1 SD	Areas are having high slope, shoulder adjacent below a top
8	Local ridges/Hills in valleys	Small scale TPI: TPI ≥ 1 SD Large scale TPI: TPI ≤ -1 SD	Areas are high in lower slope, ridge in lower slope
9	Midslope ridges, Small hills in plains	Small scale TPI: TPI ≥ 1 SD Large scale TPI: -1 SD < TPI < 1 SD	Areas are high in mid slope, ridge in mid slope
10	Mountain tops, High ridges	Small scale TPI: TPI ≥ 1 SD Large scale TPI: TPI ≥ 1 SD	Areas are highest in the landscape, having positive plan and/or profile curvature

TPI = Topographic position index, SD = Standard deviation.

Soil sampling unit identification and soil sample data collection and data analysis

The derived landform category was firstly overlaid with geological formation (Triassic granite and Precambrian gneiss) for soil sampling unit stratification. Then, the number of soil samples was calculated at the detailed reconnaissance soil survey level (1:40,000–1:100,000) with a soil sample intensity of one sample per 2 km^2 (Kheoruenromne, 2005). Finally, a stratified random sampling scheme was applied to allocate soil sample sites using the ERDAS Imagine software package Version 8.7 (Leica Geosystems, 2004) as shown in Table 2. At each soil sampling site, soil samples were taken from the topsoil (0–25 cm) and subsoil (between 25–50 cm) with a soil auger. These samples were then analyzed for physical properties (soil texture) and chemical soil properties of pH, organic matter (OM), total nitrogen (N), available phosphorus (P), exchangeable potassium (K), cation exchange capacity (CEC) and base saturation (BS) in the soil laboratory. In addition, the accuracy of landform classification was also assessed based on ground stratified random points by field observation using overall accuracy and the Kappa hat coefficient of agreement.

Table 2 Soil sampling unit stratification between geological formation and landform for sample site allocation and number of soil sample sites.

Geological formation	Landform	Area (km^2)	Percentage	Number of soil sample sites
Gr: Granite	1 Canyons, Deeply incised streams	7.78	6.68	4
Gr: Granite	2 Midslope drainages, Shallow valleys	7.53	6.46	3
Gr: Granite	3 Upland drainages, Headwaters	0.11	0.09	1
Gr: Granite	4 U-Shaped valleys	5.65	4.85	3
Gr: Granite	5 Plains	1.18	1.01	1
Gr: Granite	6 Open slopes	32.82	28.14	9
Gr: Granite	7 Upper slopes	6.40	5.49	3
Gr: Granite	8 Local ridges, Hills in valleys	0.01	0.01	1
Gr: Granite	9 Midslope ridges, Small hills in plains	5.13	4.40	2
Gr: Granite	10 Mountain tops	11.71	10.04	5
PE: Gneiss	1 Canyons, Deeply incised streams	3.79	3.25	2
PE: Gneiss	2 Midslope drainages, Shallow valleys	3.07	2.64	2
PE: Gneiss	3 Upland drainages, Headwaters	0.02	0.01	-
PE: Gneiss	4 U-Shaped valleys	2.53	2.17	1
PE: Gneiss	5 Plains	3.47	2.98	1
PE: Gneiss	6 Open slopes	18.35	15.74	5
PE: Gneiss	7 Upper slopes	1.09	0.93	1
PE: Gneiss	8 Local ridges, Hills in valleys	0.00	0.00	-
PE: Gneiss	9 Midslope ridges, Small hills in plains	2.87	2.46	2
PE: Gneiss	10 Mountain tops	3.11	2.67	2
	Total	116.61	100.00	48

Soil forming factor generation

Under this component, attributes of soil forming factors excluding time were firstly reviewed from research work (Moore *et al*., 1993; Gessler *et al*., 1995; Ryan *et al*., 2000; Gobin *et al*., 2001; Hengl *et al*., 2002; Putthapibun, 2002; McBratney *et al*., 2003; Ballabio, 2009; Castrignanò *et al*., 2011; United States Geological Service, 2012) and then they were selected for soil forming factor generation in the study (Table 3). The extracted value of the soil forming factor attribute in the slope complex area was directly applied in PLSR as summarized in Table 4.

Soil-landscape model development and its application

PLSR was firstly used to identify the relationship between *in situ* soil properties (dependent variables) and soil forming factors (independent variables) in the form of a multiple linear regression equation. Basically, PLSR is a technique that combines features from generalizes principal component analysis (PCA) and multiple linear regressions. This prediction is achieved by extracting from the predictors a set of orthogonal factors called latent variables (from PCA) which have the best predictive power and it is able

Table 3 Attribute of soil forming factors and method for data generation.

Soil forming factor	Description	Attribute	Method
Organism	Organisms relate to the effect of vegetation and human activity.	NDVI	$NDVI = \left(\frac{TM\ Band4 - TM\ Band3}{TM\ Band4 + TM\ Band3}\right)$
Relief	Relief affects run-off and erosion. Herein primary and secondary terrain attributes which were extracted from DEM with cell size of 25 × 25 m	Elevation (m)	Extract from DEM with cell size 25×25m
		Slope (°)	Extract from DEM with cell size 25×25m
		Aspect (°)	Extract from DEM with cell size 25×25m
		Plan curvature	Extract from DEM with cell size 25×25m
		Profile curvature	Extract from DEM with cell size 25×25m
		Curvature	Extract from DEM with cell size 25×25m
		TWI	$TWI = \ln\left(\frac{\text{Upslope contribute areas}}{\tan \text{slope}}\right)$
Parent material	Chemical composition of parent materials has an effect on weathering process and it can affect to soil properties	Al/Si ratio	Ratio between alumina (Al_2O_3) and silica (SiO_2) of geological formation as suggestion by Putthapiban (2002) and United States Geological Service (2012)
Climate	Rainfall affects both vegetative production and soil horizon development. Its interacting with parent material also affects to soil physical and chemical properties	Mean annual rainfall (mm)	Interpolate from mean annual rainfall of from 2000 to 2009 by IDW

NDVI = Normalized difference vegetation index, TWI = Topographic wetness index, IDW = Inversed distance weighting.

Table 4 Quantitative attributes of soil forming factors.

Soil forming factor	Attribute	Minimum value	Maximum value	Note
Organism	NDVI	-0.48	0.71	Landsat data: 17 Jan 2009
Relief	Elevation (m)	339.26	1,680.51	
	Slope (°)	0.01	47.29	
	Aspect (°)	0.0008	360.00	
	Plan curvature	-2.10	2.58	
	Profile curvature	-3.22	2.94	
	Curvature	-2.88	5.18	
	TWI	3.18	20.37	
Parent material	Al/Si ratio	0.19	0.22	
Climate	Mean annual rainfall (mm)	1,147.63	1,530.70	

NDVI = Normalized difference vegetation index, TWI = Topographic wetness index to avoid the multicollinearity problem among independent variables (Abdi, 2010). The derived multiple linear regression equations were then used to predict 10 soil properties using the Map Algebra module of the software package ArcGIS Version 9.0 (ESRI, 2004). The derived physical and chemical soil properties from PLSR were further used to compare actual soil properties from the dataset for accuracy assessment using the root mean square error (RMSE) and normalized root mean square error (NRMSE) as shown in Equations 1 and 2, respectively:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\text{Predicted value} - \text{Observed value}\right]^2} \quad (1)$$

$$\text{NRMSE} = \frac{\text{RMSE}}{\text{Maximum observed value} - \text{Minimum observed value}} \quad (2)$$

where n is number of observations.

In addition, some derived predictive soil properties (soil texture classification and soil fertility assessment) were selected to demonstrate the application. The predictive values of sand, silt and clay of topsoil and subsoil were used for soil texture classification based on the criteria of the United States Department of Agriculture (USDA) soil texture class (Soil Survey Division Staff, 1993) under the Expert System of ERDAS Imagine (Table 5). The selected predictive chemical soil properties (OM, P, K, CEC and BS) of the topsoil and subsoil were used to assess the soil fertility pattern based on the criteria of Land Development Department (1980) according to USDA Soil Survey Laboratory Method (United States Department of Agriculture, 2004). In practice, each chemical soil property was first assigned a standard score according to the soil fertility level: 1 for low, 2 for moderate and 3 for high (Table 6). Then, the total score from each soil property was simply added and the total score was reclassified into three soil fertility levels: low (total score: 5–7), moderate (total score: 8–12) and high (total score: 13–15).

Table 5 Rules and conditions for soil texture classification under the expert system based on Soil Survey Division Staff (1993).

Soil texture class	Rule and conditions
Sand	%Sand ≥ 85 and %Silt + (1.5 x %Clay) < 15
Loamy sand	%Sand ≥70 and %Sand < 91 and %Silt + (1.5 x %Clay) ≥ 15 and %Silt + (2 x %Clay) < 30
Sandy loam	%Sand > 52 and %Silt + (2 x %Clay) ≥ 30 and %Clay ≥ 7 and %Clay < 20 **OR** %Sand > 43 and %Silt < 50 and %Silt + (2 x %Clay) > 30 and %Clay < 7
Loam	%Sand ≤ 52 and %Silt ≥ 28 and %Silt < 50 and %Clay ≥ 7 and %Clay < 27
Silt loam	%Silt ≥ 50 and %Clay ≥ 12 and %Clay < 27 **OR** %Silt ≥ 50 and %Silt < 80 and %Clay < 12
Silt	%Silt ≥ 80 and %Clay < 12
Sandy clay loam	%Sand > 45 and %Silt < 28 and %Clay ≥ 20 and %Clay < 35
Clay loam	%Sand > 20 and %Sand ≤ 45 and %Clay ≥ 27 and %Clay < 40
Silty clay loam	%Sand ≤ 20 and %Clay ≥ 27 and %Clay < 40
Sandy clay	%Sand > 45 and %Clay ≥ 35
Silty clay	%Silt ≥40 and %Clay ≥ 40
Clay	%Sand ≤ 45 and %Silt < 40 and %Clay ≥ 40

Table 6 Chemical soil properties and standard score for soil fertility assessment based on Land Development Department (1980).

Fertility level	Standard Score	OM (%)	P (mg kg^{-1})	K (mg kg^{-1})	CEC (cmol kg^{-1})	BS (%)
Low	1	< 1.5	< 10	< 60	< 10	< 35
Moderate	2	1.5-3.5	10-25	60-90	10-20	35-75
High	3	> 3.5	> 25	> 90	> 20	> 75

OM = Organic matter, P = Available P, K = Exchangeable K, CEC = Cation exchange capacity, BS = Base saturation.

RESULTS AND DISCUSSION

Landform classification

The most dominant landform was open slopes covering an area of 51.17 km^2 or 43.88% of the trial watershed. The second dominant landform was mountain tops (14.82 km^2 or 12.71%), while upland drainage or headwaters and local ridges or hills in valleys covered an area of less than 1%. The overall accuracy of landform classification and the Kappa hat coefficient of agreement were 92.00% and 0.91, respectively. According to Landis and Koch (1977), a Kappa hat coefficient of agreement value of more than 0.80 represents strong agreement or accuracy between the classification map and the ground reference information. The distribution of the landform classification and an example of the accuracy assessment using field observation are displayed in Figure 3.

The landform classification based on TPI values was a useful method for soil landscape analysis, because the criteria parameters used were simple and thus, this method was able to identify major landform elements in the mountainous areas such as mountain tops, open slopes, plains and canyons which correlated to soil erosion, the deposition process and soil horizon development. However, the accuracy of this method depends on the DEM resolution and an optimal kernel size specification.

Soil sampling unit identification and sample sites allocation

The combination of the derived landform categories and geological formation was used to stratify the soil sampling units and to allocate soil sample sites using a stratified random sampling scheme. In this study, 48 soil sample sites were selected and data was collected from them analysis. This dataset was divided into two datasets: 28 sites for modeling and 20 sites for assessment of the model accuracy in soil property prediction.

Soil sample data analysis

The major soil properties from the 48 soil sample sites of topsoil and subsoil were qualitatively and quantitatively described based on soil laboratory reports.

Of the physical soil properties, the soil texture of the topsoil was dominated by sandy clay loam (30 samples) and the remainder was represented by clay loam (17 samples) and loam (1 sample), while the subsoil was dominated by clay (32 samples) with the remainder represented by clay loam (8 samples), sandy clay (4 samples) and sandy clay loam (4 samples). For the chemical soil properties, the pH of the topsoil and subsoil varied from extremely acid to neutral (pH 4.33–6.74), the organic matter content of the topsoil was moderate to high (1.79–10.05%) and of the subsoil was low to high (0.53–3.98%), while the total nitrogen of the topsoil was very low to moderate (0.09–0.51%) and of the subsoil was very low to low (0.03–0.18%). The available phosphorus concentration of the topsoil was low to moderate (0.81–11.15 mg.kg^{-1}) while in the subsoil, it was low (0.41–2.57 mg.kg^{-1}) and the exchangeable potassium concentration of the

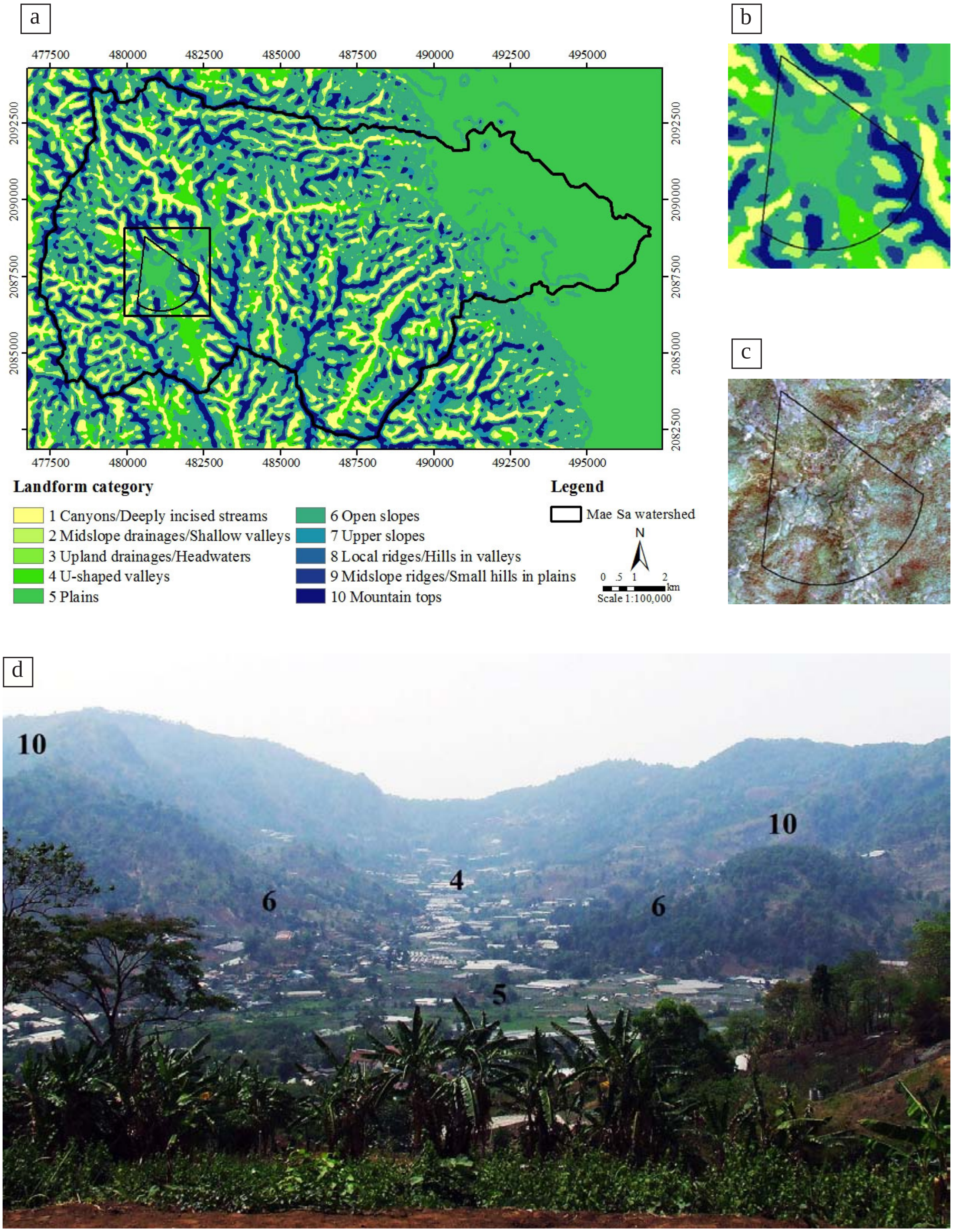

Figure 3 Distribution of landform classification and ground visiting sites for accuracy assessment: (a) landform classification, (b) landform data, (c) Landsat data and (d) ground visiting sites (numbers indicate the landform categories visited as listed in subfigure 3d).

topsoil was moderate to very high (81.95–262.27 mg.kg^{-1}) whereas in the subsoil, it was low to high (27.72–148.34 mg.kg^{-1}). The cation exchange capacity of the topsoil and subsoil were both moderate to high (11.15–35.44 cmol.kg^{-1}) and the base saturation of the topsoil was low to moderate (15.15–43.59%) while in the subsoil, it was low (11.56–31.27%).

These results showed that most of the topsoil was more fertile than the subsoil, as most of the topsoil property values were higher than in subsoil except for the clay content. This latter result might have been due to the rather high OM content in the topsoil.

Soil-landscape model for soil property prediction using PLSR

The relationship between *in situ* soil properties (dependent variables) from the 28 sample sites and soil forming factors (independent variables) for each soil property of the topsoil and subsoil were analyzed using PLSR. The intercepts and coefficients of the predictor variables for each soil property from PLSR are summarized by columns in Table 7. For the topsoil properties, the best predictive model was sand (R^2 = 0.92) while the worst predictive model was silt (R^2 = 0.52). The best predictive model for the subsoil properties was CEC (R^2 = 0.85) while the worst predictive model was with nitrogen and phosphorus (R^2 = 0.59). From these results, the variation of the predictive model can be inferred for explaining the soil properties using soil forming factors. In addition, Table 8 summarizes the values of variable importance in the projection (VIP) in each soil property prediction model using PLSR. These

Table 7 Summary of regression coefficients for the prediction of topsoil and subsoil properties with R^2 values.

Multiple linear regression of soil property prediction										
Topsoil	Sand	Silt	Clay	pH	OM	N	P	K	CEC	BS
Intercept	16.78	40.86	30.43	6.56	-5.13	-0.21	1.30	113.92	3.47	61.17
Rain	0.02	-0.01	-1.0×10^{-3}	-9.0×10^{-4}	7.8×10^{-2}	3.2×10^{-4}	7.5×10^{-5}	-0.01	0.01	-0.03
NDVI	-8.57	1.08	-3.79	-0.88	6.15	0.24	5.94	-46.59	8.77	-20.56
Elevation	2.5×10^{-3}	-2.0×10^{-3}	8.2×10^{-5}	-1.6×10^{-4}	1.4×10^{-3}	5.3×10^{-5}	1.8×10^{-4}	4.8×10^{-3}	2.4×10^{-3}	-3.7×10^{-3}
Slope	0.46	-0.09	-0.12	-0.03	0.02	4.67×10^{-4}	-3.5×10^{-3}	-1.73	0.01	-0.23
Aspect	2.5×10^{-3}	8.3×10^{-4}	4.0×10^{-3}	5.1×10^{-4}	-8.6×10^{-4}	3.2×10^{-5}	-5.8×10^{-4}	0.09	2.0×10^{-3}	0.01
Plan	1.77	-1.32	-2.47	-0.06	-0.67	-0.054	-1.39	-21.25	-2.25	-2.64
Profile	-0.64	1.27	1.52	0.10	1.12	0.067	1.46	18.71	2.39	1.12
Curvature	0.73	-0.77	-1.19	-0.05	-0.53	-0.036	-0.85	-11.87	-1.37	-1.13
TWI	-1.43	0.59	0.41	0.11	0.14	1.1×10^{-2}	0.32	7.68	0.32	0.81
Al/Si ratio	29.89	13.72	1.87	1.76	-17.09	-0.90	-8.39	113.50	-34.38	60.27
R^2	0.92	0.52	0.67	0.82	0.68	0.80	0.74	0.86	0.77	0.79
Subsoil	Sand	Silt	Clay	pH	OM	N	P	K	CEC	BS
Intercept	30.12	19.74	43.99	4.85	-2.74	-0.40	0.85	42.29	3.50	41.77
Rain	0.01	-7.8×10^{-4}	-3.2×10^{-3}	1.3×10^{-4}	3.1×10^{-3}	9.7×10^{-4}	3.7×10^{-5}	0.01	0.01	-0.02
NDVI	7.76	3.72	-13.56	0.26	2.59	0.40	0.53	12.20	5.03	-8.57
Elevation	-8.4×10^{-4}	8.0×10^{-5}	2.2×10^{-3}	5.7×10^{-5}	5.2×10^{-4}	1.7×10^{-4}	1.4×10^{-5}	0.01	2.6×10^{-3}	-2.5×10^{-3}
Slope	0.51	-0.01	-0.51	-0.01	8.8×10^{-3}	-1.4×10^{-5}	-2.0×10^{-3}	-0.40	-3.5×10^{-3}	-0.17
Aspect	-3.6×10^{-3}	1.8×10^{-4}	8.0×10^{-3}	3.3×10^{-4}	-1.7×10^{-4}	8.5×10^{-5}	3.1×10^{-4}	0.02	$3.7\ \times10^{-3}$	1.8×10^{-3}
Plan	3.50	-2.13	-1.58	-0.26	-0.31	-0.12	-0.26	-16.11	-2.26	-1.54
Profile	-4.71	2.51	1.84	0.28	0.54	0.16	0.28	17.52	2.67	1.09
Curvature	2.42	-1.37	-1.01	-0.16	-0.25	-0.08	-0.16	-9.96	-1.46	-0.78
TWI	-1.59	0.42	1.03	0.05	0.07	0.02	0.06	3.63	0.32	0.56
Al/Si ratio	-40.36	-10.28	42.83	-1.68	-6.58	-1.83	-2.12	-70.16	-36.19	21.89
R^2	0.82	0.78	0.77	0.66	0.66	0.59	0.59	0.78	0.85	0.60

OM = Organic matter, P = Available P, K = Exchangeable K, CEC = Cation exchange capacity, BS = Base saturation.

values represent the significance of the predictor variables for soil properties prediction. Chong and Jun (2005) stated that any independent variable with a VIP value greater than 1 was considered as a highly important predictor. For example, TWI, which had a negative relationship with the percentage of sand, had the highest VIP value (Tables 7 and 8) and it dictated the distribution of the low percentage of sand (Figure 4). In contrast, the Al/Si ratio, which had a positive relationship with the percentage of sand, demonstrated the lowest VIP value (Tables 7 and 8) and it did not control the distribution of the low percentage of sand (Figure 4).

As a result, the relief factors of TWI, curvature, plan curvature and profile curvature are important for soil property prediction in mountainous areas.

Accuracy assessment of soil property prediction model

The predictive physical and chemical soil properties from PLSR were used to compare the *in situ* soil properties from the 20 sites to assess the accuracy of the soil property prediction model using RMSE and NRMSE (Table 9). In principle, RMSE provides the absolute average error between an estimated value and an observed value with a measured unit. This value is not appropriate for accuracy assessment comparison when the measured units are different. As NRMSE is a normalized value of RMSE and thus, has no unit. This value is appropriate for accuracy assessment comparison. Based on NRMSE values, the best predictive topsoil property model was clay while the best predictive subsoil property model was potassium and CEC. The worst predictive topsoil and subsoil property model was phosphorus.

Table 8 Variable importance in the projection (VIP) values of each predictor variable for soil property prediction.

Soil properties		Variable importance in the projection (VIP)									
		Rainfall	NDVI	Elevation	Slope	Aspect	Plan curvature	Profile curvature	Curvature	TWI	Al/Si ratio
Sand	Topsoil	0.66	0.22	0.40	**1.57**	0.52	**1.21**	0.85	1.13	**1.77**	0.20
	Subsoil	0.17	0.18	0.14	**1.35**	0.59	1.27	1.23	**1.36**	**1.66**	0.03
Silt	Topsoil	**1.27**	0.15	0.87	**1.27**	0.16	0.78	0.73	0.82	**2.01**	0.36
	Subsoil	0.08	0.55	0.04	0.17	0.04	1.39	**1.59**	**1.62**	**1.58**	0.29
Clay	Topsoil	0.09	0.50	0.04	**1.67**	0.76	**1.45**	0.87	1.27	**1.39**	0.05
	Subsoil	0.20	0.55	0.18	**1.84**	0.82	1.08	0.91	**1.09**	**1.54**	0.20
pH	Topsoil	0.75	0.78	0.55	**1.54**	0.81	**1.06**	0.69	0.96	**1.64**	0.54
	Subsoil	0.10	0.30	0.22	0.64	0.57	1.39	**1.44**	**1.54**	**1.61**	0.39
OM	Topsoil	**1.27**	**1.52**	1.10	0.55	0.30	0.74	**1.19**	1.04	0.86	0.82
	Subsoil	**1.15**	**1.49**	0.96	0.52	0.14	0.80	**1.33**	1.16	0.98	0.74
N	Topsoil	0.97	1.12	0.80	0.22	0.22	1.12	**1.35**	**1.34**	**1.25**	0.81
	Subsoil	**1.25**	0.78	1.10	0.00	0.24	1.02	**1.38**	**1.30**	1.20	0.70
P	Topsoil	0.01	1.24	0.12	0.08	0.17	1.28	**1.31**	**1.41**	**1.72**	0.34
	Subsoil	0.03	0.61	0.05	0.23	0.51	1.30	**1.38**	**1.46**	**1.83**	0.47
K	Topsoil	0.05	0.21	0.20	1.09	0.98	**1.34**	1.17	**1.37**	**1.65**	0.06
	Subsoil	0.12	0.23	0.40	0.77	0.57	1.32	**1.39**	**1.47**	**1.72**	0.25
CEC	Topsoil	1.10	1.08	0.95	0.18	0.36	**1.22**	**1.26**	**1.35**	1.02	0.82
	Subsoil	0.87	0.62	1.03	0.04	0.66	**1.23**	**1.41**	**1.44**	1.01	0.87
BS	Topsoil	1.13	**1.32**	0.76	**1.52**	0.87	0.75	0.31	0.58	**1.33**	0.75
	Subsoil	**1.15**	0.91	0.86	**1.81**	0.27	0.73	0.50	0.67	**1.53**	0.45

NDVI = Normalized difference vegetation index, OM = Organic matter, P = Available P, K = Exchangeable K, CEC = Cation exchange capacity, BS = Base saturation, TWI = Topographic wetness index.
The three best significant factors for each soil property prediction are shown as bold numbers.

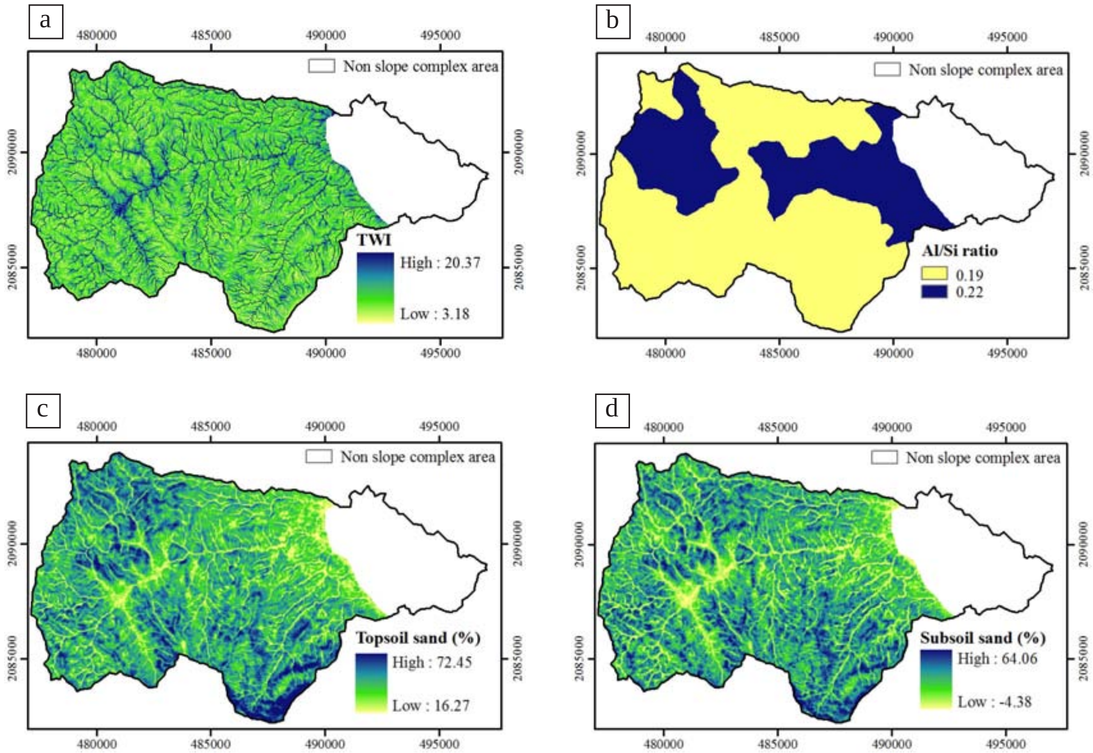

Figure 4 Distribution in study area of: (a) topographic wetness index, (b) Al/Si ratio, (c) predictive sand of topsoil, and (d) predictive sand of subsoil.

Table 9 Summary of root mean square error (RMSE) and normalized root mean square error (NRMSE) values for topsoil and subsoil properties prediction.

Topsoil	Sand	Silt	Clay	pH	OM	N	P	K	CEC	BS
RMSE	3.53	1.79	1.95	0.26	1.80	0.08	2.01	19.58	3.13	4.82
NRMSE	0.22	0.23	0.18	0.20	0.24	0.23	0.25	0.19	0.22	0.21
Subsoil	Sand	Silt	Clay	pH	OM	N	P	K	CEC	BS
RMSE	5.43	2.61	5.08	0.37	0.84	0.04	0.60	22.27	2.57	2.78
NRMSE	0.22	0.23	0.25	0.26	0.27	0.24	0.36	0.18	0.18	0.20

OM = Organic matter, P = Available P, K = Exchangeable K, CEC = Cation exchange Capacity, BS = Base saturation.

Classification of soil texture using the expert system

There were four soil texture classes of topsoil: sandy clay loam (58.30%), clay loam (41.65%), clay (0.04%), and silty clay (0.02%), while there were four soil texture classes of subsoil: clay (66.31%), clay loam (22.21%), sandy clay loam (7.72%) and sandy clay (3.76%) as shown in Figure 5. In addition, the accuracy for topsoil and subsoil texture classification was assessed based on the 48 *in situ* soil samples. It was found that the overall accuracy and Kappa hat coefficient values for the topsoil and subsoil texture classification were 81.25% and 0.61 and 73.92% and 0.47, respectively. As a result, the dominant soil texture of the topsoil was sandy clay loam and it was mostly distributed in the mountainous area, while the clay material was mostly deposited in the subsoil as a significant soil texture.

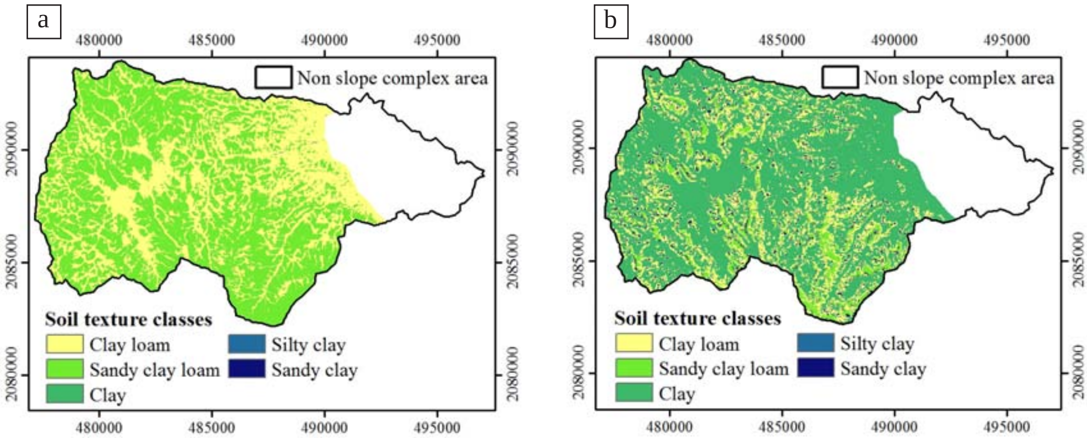

Figure 5 Distribution of United States Department of Agriculture soil texture classes in the study area: (a) topsoil and (b) subsoil.

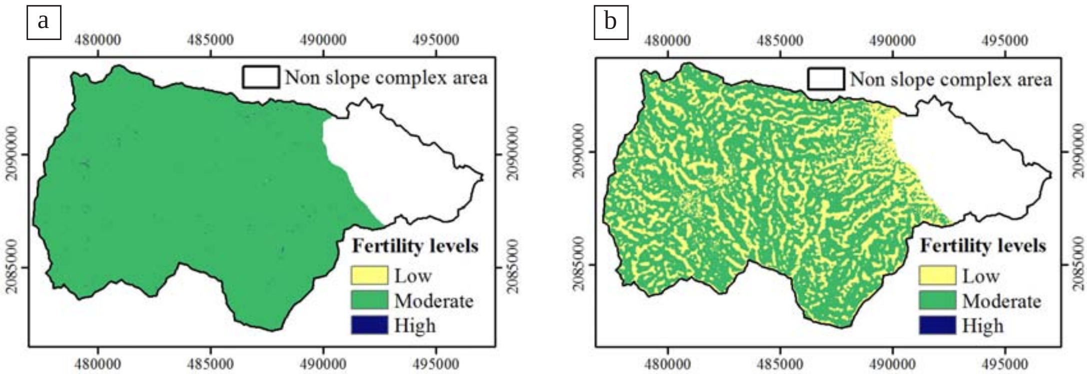

Figure 6 Distribution of soil fertility pattern in the study area: (a) topsoil and (b) subsoil.

Soil fertility assessment

The pattern of soil fertility distribution is presented in Figure 6. The fertility of the topsoil was mostly moderate (99.89%) because the base saturation and available phosphorus of the topsoil varied from low to moderate. At the same time, fertility of subsoil was low and moderate because the base saturation of the subsoil varied from low to moderate and available phosphorus was low covering an area of 41.73 (low) and 74.88 km^2 (moderate) or about 35.78 and 64.22%, respectively. Based on these findings, the topsoil and subsoil fertility can be improved by adding P fertilizers.

CONCLUSION

Under spatial modeling, PLSR was used to quantify the relationship between bio-physical soil-forming factors consisting of: rainfall, NDVI, elevation, slope, aspect, plan curvature, profile curvature, curvature, TWI and the Al/Si ratio as predictor variables and the soil properties of sand, silt, clay, pH, OM, N, P, K, CEC, and BS of the topsoil and subsoil as dependent variables. The results showed that PLSR can be used to predict the physical and chemical soil properties in mountainous areas. In fact, the R^2 of the predictive soil property model for topsoil and subsoil varied between 52 to 92% and 59 to 85%, respectively, whilst the accuracy assessment for the topsoil and subsoil property prediction models by NRMSE varied between 0.18 and 0.25 and 0.18 and 0.36, respectively. In addition, the selected predictive soil properties were used for soil texture classification and soil fertility assessment.

In conclusion, it is suggested that spatial modeling using PLSR can be efficiently used as a tool for soil property prediction in mountainous areas where soil characteristics and properties are not available.

ACKNOWLEDGEMENT

The authors would like to express their appreciation and gratitude to the anonymous reviewers for their valuable comments and contributions to the manuscript.

LITERATURE CITED

Abdi, H. 2010. Partial least squares regression and projection on latent structure regression (PLS Regression). **WIREs Computational Statistics** 2: 97–106.

Ballabio, C. 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. **Geoderma** 151: 338–350.

Boettinger, J.L. 2010. Environmental covariates for digital soil mapping in the western USA, pp. 17–27. *In* J.L. Boettinger, D.W. Howell, A.C. Moore, A.E. Hartemink and S. Kienast-Brown, (eds.). **Digital Soil Mapping: Bridging Research, Environmental Application, and Operation.** Springer Science+Business Media. Dordrecht, the Netherlands.

Castrignanò, A., G. Buttafuoco, R. Comolli and A. Castrignanò. 2011. Using digital elevation model to improve soil pH prediction in an alpine doline. **Pedosphere** 21(2): 259–270.

Chong, I.G. and C.H. Jun. 2005. Performance of some variable selection methods when multicollinearity is present. **Chemometr. Intell. Lab.** 78: 103–112.

Department of Mineral Resources. 2006. **Provincial Geological Map 1:50,000 of Chiang Mai.** [CD-ROM]. Department of Mineral Resources, Ministry of Natural Resources and Environment. Bangkok, Thailand.

Environmental Systems Research Institute. (ESRI). 2004. **ArcGIS software package Version 9.0.** Environmental Systems Research Institute, Inc. Redlands, CA, USA.

Florinsky, I.V., R.G. Eilers, G.R. Manning and L.G. Fuller. 2002. Prediction of soil properties by digital terrain modeling. **Environ. Modell. Softw.** 17: 295–311.

Gessler, P.E., I.D. Moore, N.J. McKenzie and P.J. Ryan. 1995. Soil landscape modelling and spatial prediction of soil attributes. **Int. J. Geogr. Inf. Sci.** 9: 421–432.

Gessler, P.E., O.A. Chadwick, F. Chamran, L. Althouse and K. Holmes. 2000. Modeling soil–landscape and ecosystem properties using terrain attributes. **Soil Sci. Soc. Am. J.** 64(3) : 2046–2054.

Gobin, A., P. Campling and J. Feyen. 2001. Soil-landscape modelling to quantify spatial variability of soil texture. **Phys. Chem. Earth Pt. B** 26(1): 41–45.

Grunwald, S. 2006. **Environmental Soil-landscape Modeling: Geographic Information Technologies and Pedometrics.** CRC Press. New York, NY, USA. 488 pp.

Hengl, T., D.G. Rossiter and S. Husnjak. 2002. Mapping soil properties from an existing national soil data set using freely available ancillary data, pp. 1–10. *In* **17th World Congress of Soil Science Bangkok, Thailand.** 14–21 August 2002. No. 1140.

Herbst, M., B. Diekkrüger and H. Vereecken. 2006. Geostatistical co-regionalization of soil hydraulic properties in a micro-scale catchment using terrain attributes. **Geoderma** 132: 206–221.

Huggett, R.J. 1975. Soil landscape systems: A model of soil genesis. **Geoderma** 13(1): 1–22.

Jenny, H. 1980. **The Soil Resources: Origin and Behavior.** Springer-Verlag. New York, NY, USA. 377 pp.

Kheoruenromne, I. 2005. **Soil Survey.** 2nd ed. Department of Soil Science, Faculty of Agriculture, Kasetsart University. Bangkok, Thailand. 733 pp. [in Thai]

Land Development Department. 1980. **Manual of Soil Suitability Classification for Economic Crops.** Technical Paper No. 28. Soil Survey Division, Land Development Department, Ministry of Agriculture and Cooperative. Bangkok, Thailand. 70 pp. [in Thai]

Land Development Department. 1992. **Preliminary Study on Highland Development Project in Northern Thailand.** Soil Survey Division, Land Development Department, Ministry of Agriculture and Cooperatives. Bangkok, Thailand. 66 pp. [in Thai]

Landis, J. and G. Koch. 1977. The measurement of observer agreement for categorical data. **Biometrics** 33: 159–174.

Leica Geosystems. 2004. **ERDAS Imagine Software Package Version 8.7.** Atlanta, GA, USA.

McBratney, A.B., M.L.M. Santos and B. Minasny. 2003. On digital soil mapping. **Geoderma.** 117: 3–52.

McKenzie, N.J., P.E. Gessler, P.J. Ryan and D.A. O'Connell. 2000. The role of terrain analysis in soil mapping, pp. 245–266. *In* J.P. Wilson and J.C. Gallant, (eds.) **Terrain analysis: Principles and Application.** John Wiley & Sons. New York, NY, USA.

Moore, L.D., P.E. Gessler, G.A. Nielsen and G.A. Peterson. 1993. Soil attribute prediction using terrain analysis. **Soil. Sci. Soc. Am. J.** 57: 443–452.

Odeh, I.O.A., A.B. McBratney and D.J. Chittleborough. 1995. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. **Geoderma** 67: 215–226.

Pennock, D.J., B.J. Zebarth and E.D. Jong. 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. **Geoderma** 40(3-4): 297–315.

Putthapiban, P. 2002. Geology and geochronology of the igneous rocks of Thailand. *In* **Symposium on Geology of Thailand.** 26-31 August, 2002. Bangkok, Thailand.

Royal Forest Department. 2007. **Forest Area and Land Use of Thailand** [CD-ROM]. Bangkok, Thailand.

Ryan, P.J., N.J. McKenzie, D. O'Connell, A.N. Loughhead, P.M. Leppert, D. Jacquier and L. Ashton. 2000. Integrating forest soils information across scales: Spatial prediction of soil properties under Australian forests. **Forest. Ecol. Manag.** 138: 139–157.

Soil Survey Division Staff. 1993. **Soil Survey Manual.** United States Department of Agriculture. Washington, DC, USA. 437 pp.

Thai Meteorological Department. 2010. **Climatic Data of Chiang Mai.** [CD-ROM]. Thai Meteorological Department, Ministry of Information and Communication Technology. Bangkok, Thailand.

Thompson, J.A., J.C. Bell and C.A. Butler. 1997. Quantitative soil-landscape modeling for estimating the areal extent of hydromorphic soils. **Soil Sci. Soc. Am. J.** 61(3): 971–980.

Thompson J.A., E.M. Pena-Yewtukhiw and J.H. Grove. 2006. Soil-landscape modeling across a physiographic region: topographic patterns and model transportability. **Geoderma** 133: 57–70.

United States Department of Agriculture. 2004. Soil Survey Investigations Report No. 42, Version No. 4: **Soil Survey Laboratory Methods Manual.** Natural Resources Conservation

Service, United States Department of Agriculture, Washington, DC, USA. 700 pp.

United States Geological Service. 2012. **National Geochemical Database: Rock.** [Available from: http://mrdata.usgs.gov/ngdb/rock/]. [Sourced: 18 October 2012].

Weiss, A.D. 2001. Topographic position and landforms analysis (Poster Presentation). *In* **ESRI International User Conference.** 9-13 July, 2001. San Diego, CA, USA.

Wilson, J. P. and J.C. Gallant. 2000. **Terrain Analysis: Principles and Application.** John Wiley & Sons. New York, NY, USA. 479 pp.

Ziadat, F.M. 2007. Land suitability classification using different sources of information: Soil maps and predicted soil attributes in Jordan. **Geoderma** 140: 73–80.