Digital Soil Mapping for Functional Analysis of Site Characteristics in Complex Terrain

Dissertation

to attain the academic degree of Doctor of Natural Sciences (Dr. rer. nat.) of the

Bayreuth Graduate School of Mathematical and Natural Sciences (BayNAT) of the

University of Bayreuth

presented by

Gwanyong Jeong

born 20 September 1982

in Gwang-Ju, Republic of Korea

Bayreuth, April 2016

This doctoral thesis was prepared at the department of Geosciences (chair of Soil Physics) at the University of Bayreuth from September 2012 until April 2016 and was supervised by Prof. Dr. Bernd Huwe, Dr. Mareike Ließ and Prof. Dr. Soo Jin Park.

This is a full reprint of the dissertation submitted to obtain the academic degree of Doctor of Natural Sciences (Dr. rer. nat.) and approved by the Bayreuth Graduate School of Mathematical and Natural Sciences (BayNAT) of the University of Bayreuth.

Date of submission: 07.04.2016

Date of defense: 27.06.2016

Acting director: Prof. Dr. Stephan Kümmel

Doctoral committee:

Prof. Dr. Bernd Huwe   (1st reviewer)

Dr. Christina Bogner   (2nd reviewer)

Prof. Dr. Cyrus Samimi   (Chairman)

Prof. Dr. Egbert Matzner

Abstract

Information on soils such as nutrient availability is essential for sustainable mountain ecosystem management. Heterogeneous soil nutrients might determine growth, distribution, and diversity of plants. Therefore, spatial patterns of soil nutrients should be investigated in mountainous areas.

Digital soil mapping (soil landscape modelling) was used for important chemical soil parameters in the Soyang lake watershed, South Korea. Specific purposes are: (1) to develop maps of soil nutrients for ecological land potential assessment, (2) to investigate spatial patterns of various phosphorus (P) fractions, and (3) to predict nitrogen (N) to P ratios in the topsoil layer.

Firstly, vegetation indices had the highest predictive power for soil nutrients. Using selected instead of all predictors via recursive feature elimination (RFE) improved prediction results considerably. Random forest (RF) showed the best performance compared to support vector regression (SVR) and generalized additive models (GAM). Cluster analysis identified four land potential classes: fertile, medium and unfertile with an additional class dominated by high phosphorus and low carbon and nitrogen contents due to human impact. This study provides an effective approach to map ecological land potentials for sustainable mountain ecosystem management.

Secondly, surface curvature and elevation were important predictors for all P fractions. The concentrations of all P fractions changed with surface curvature and elevation. Higher values of most P fractions were found at the lower slope due to soil erosion. Especially, organic P was enriched at the lower slope, while the relative portion of residual P fractions was largest at the upper slope.

Finally, surface curvature was selected as an important predictor for P contents in organic and mineral A horizons. LiDAR derived vegetation predictors and normalized difference vegetation

index (NDVI) strongly contributed to model N in the organic layer. N to P ratios in the organic and mineral A horizons showed higher values at convex upper slopes and increased with surface curvature. This implies that spatial patterns of P and N in a mountainous catchment with steep slopes under monsoon conditions are mainly controlled by topography.

In this thesis, various methods (e.g. predictor selection and importance, uncertainty assessment, and LiDAR analysis) were applied to digital soil mapping. Important environmental predictors and processes related to spatial patterns of soil nutrients were investigated. Based on our results, it is possible to better understand soil nutrient dynamics in landscapes and identify sensitive areas under environmental changes (e.g. areas with high nitrogen deposition).

Zusammenfassung

Informationen über Böden, insbesondere deren Nährstoffverfügbarkeit, sin notwendig für ein nachhaltiges Management von Gebirgslandschaften. Die räumliche Verteilung der Bodennährstoffe hat hierbei oft einen großen Einfluss auf Wachstum, Verteilung und Diversität der Pflanzen. Der Erfassung der Raummuster von Bodennährstoffen in Berglandschaften kommt damit eine große Bedeutung zu.

Methoden der digitalen Bodenkartierung (Bodenlandschaftmodellierung) wurden im Soyang-Einzugsgebiet auf wichtige chemische Bodenparameter angewandt. Die spezifischen Ziele sind: (1) die Bereitstellung von Karten von Bodennährstoffen zur Beurteilung der Bodenfruchtbarkeit, (2) die Analyse der Raummuster der verschiedenen Phosphorfraktionen, (3) die flächige Vorhersage der N:P-Verhältnisse im Oberboden.

Es zeigte sich, dass Vegetationsindizes die beste Vorhersagegüte für Bodennährstoffe aufwiesen. Die Einschränkung auf ausgewählte Prädiktoren mit dem „recursive feature elimination"-Algorithmus (RFE) verbesserte die Vorhersagen deutlich im Vergleich zur Vorhersage mit allen Prädiktoren. Random Forest zeigte hierbei die beste Vorhersageleistung im Vergleich zu den anderen benutzten Methoden (support vector regression (SVR) und generalized additive models (GAM)). Anhand von Clusteranalysen konnten vier Klassen von Standortspotentialen unterschieden werden: fruchtbar, mittel und unfruchtbar und eine zusätzliche Klasse mit hohen Phosphorgehalten, sowie niedrigen Kohlenstoff- und Stickstoffgehalten als Folge anthropogener Einflüsse. Insgesamt stellt diese Arbeit einen Beitrag zur Kartierung ökologischer Standortspotentiale und zum nachhaltigen Management von Bergökosystemen dar.

Zweitens, Oberflächenkrümmung und Höhe waren wichtige Prädiktoren für alle Phosphorfraktionen. Die Konzentrationen aller Phosphorfraktionen änderten sich mit der Oberflächenkrümmung und der Höhe. Als Folge der Bodenerosion wurden am Unterhang oft

erhöhte Werte der meisten Phosphorfraktionen gefunden. Besonders der organisch gebundene Phosphor war am Unterhang angereichert, während die relativen Anteile der restlichen Phosphorfraktionen an der oberen Hangseite am höchsten waren.

Schließlich wurde die Oberflächenkrümmung als eine wichtige Vorhersagevariable für Phosphorgehalte in organischen und mineralischen A-Horizonten identifiziert. LiDAR basierte Vegetationsparameter und NDVI (normalized difference vegetation index) waren wesentliche Prädiktoren für die Modellierung von N in der organischen Auflage. Die N:P-Verhältnisse in organischen Auflagen und A-Horizonten zeigten höhere Werte an konvexen Oberhängen und nahmen mit der Oberflächenkrümmung zu. Das bedeutet, dass in einem gebirgigen Einzugsgebiet mit steilen Hängen unter Monsoonklima die Raummuster von Phosphor und Stickstoff hauptsächlich durch die Topografie kontrolliert werden.

In dieser Arbeit wurden unterschiedliche Methoden (z.B. Prädiktor-Selektion und -Relevanz, Unschärfeanalysen) und Analysen von LiDAR-Daten für digitale Bodenkarten angewandt. Wichtige ökologische Parameter und Prozesse wurden hinsichtlich ihrer Bedeutung für die räumliche Verteilung von Bodennährstoffen in der Landschaft analysiert. Letztlich kann man hiermit auch die Mechanismen der Nährstoffdynamik in Landschaften besser verstehen und sensitive Bereiche bei veränderten Umweltbedingungen identifizieren (z.B. Flächen mit hoher Stickstoff-Deposition).

Acknowledgements

Contents

List of figures

List of tables

Chapter 1 Synopsis

1.1 Soil nutrients in the mountainous ecosystem: background

Many studies showed essential roles of mountain soils for ecological functions (Ballabio, 2009; Roman et al., 2010; Wilcke et al., 2010). Mountain soils can serve as an important indicator of natural disasters (e.g. landslides) (Ließ et al., 2011) and also as a sensible measure of environmental changes such as climate change (Rodionov et al., 2007). Furthermore, soil nutrients such as nutrient availability determine forest growth and product (Benner et al., 2010; Osman, 2013). Hence, mountain soils should be considered as an key environmental factor for sustainable management in mountainous ecosystems (Funnell and Parish, 2005).

However, existing soil maps have limitations in providing detailed spatial information particularly concerning the soils in mountain regions (Burrough et al., 2000; Grunwald, 2006). Firstly, polygon-based soil maps couldn't describe continuous local variations of soil properties. Secondly, they don't provide information on values of soil properties required by environmental models since they were based on soil types. Lastly, traditional soil mapping based on field work requires long time, expensive costs, and many soil mappers for the update. Finally, these maps don't quite provide suitable data for environmental modelling and land management. Soil maps of Korea are available for agricultural (1:5,000) and forest (1:25,000) purposes. These maps have similar limitations. The agricultural maps don't have detailed information on mountain soils, while forest soil maps have not been surveyed in any agricultural areas. Few data was collected in the Soyang lake watershed.

Soil processes are controlled by a number of environmental factors operating over time to develop a particular soil profile with its horizontation and properties (Amundson, 2014). Jenny (1941) proposed a state factor equation and identified the principal factors as climate (c), organisms (o), relief (r), parent material (p), and time (t). Recent developments in geographic information system and remote sensing techniques made it possible to analyse the quantitative relationships between the spatial soil distribution pattern and environmental factors. There

have been many attempts to investigate these relationships (Dobos and Hengl, 2009; Grunwald, 2009; McBratney et al., 2003; Minasny and McBratney, 2015) by supervised learning methods such as support vector regression (SVR) (Smola and Schölkopf, 2004), random forest (RF) (Breiman, 2001) or artificial neural networks (ANN) (Bishop, 1995). This approach is called 'soil-landscape modelling', 'digital soil mapping' (DSM) or 'Predictive soil mapping', which can be defined as the development of a numerical or statistical model of the relationship between environmental predictors and soil properties (Scull et al., 2003).

Mountain areas provide a real challenge to any soil mapping approach as they are poor accessibility, have scarce data availability, and various slope processes like mass movement, debris flow, and severe soil erosion. Soils in mountain areas have developed by complex interactions among various environmental factors. In mountain areas, organic matter accumulation, biochemical weathering, and nutrient cycling are enhanced by vegetation in the soil (Brady and Weil, 2010). Furthermore, parent rock materials can determine the nutrient level (e.g. soil nitrogen and phosphorus) (Binkley and Fisher, 2012; Mage and Porder, 2013; Morford et al., 2011). Especially, it is difficult to quantify the spatial variability of the soil chemical properties related to soil nutrients because the chemical properties have high variations depending on time and are influenced by various soil-forming factors (Dobos and Hengl, 2009).

This study aims at predicting spatial distributions of important soil chemical properties in mountain regions using supervised learning methods including uncertainty analysis in order to determine the spatial soil fertility patterns which determine plant growth and hence forest growth and management.

1.2 Digital soil mapping

Digital soil mapping is the procedure of arranging information on individual natural resources to understand their similarities, relationships and spatial patterns to infer the spatial distribution of soil properties based on climate, geology, vegetation, soil and relief. The procedure is described by Figure 1.1. The three following questions are important issues for digital soil mapping (Lagacherie and McBratney, 2007): (1) What environmental factors are of particular importance? (2) How to select the best prediction model? (3) How better the outputs of digital soil mapping show? (e.g., uncertainty visualization)

1.2.1 Supervised learning methods: correlation modelling

McBratney et al. (2003) reviewed the recent methods for digital soil mapping and suggested the inclusion of soil (s) and spatial information (n) as predictors in addition to Jenny's environmental factors (Equation 1.1).
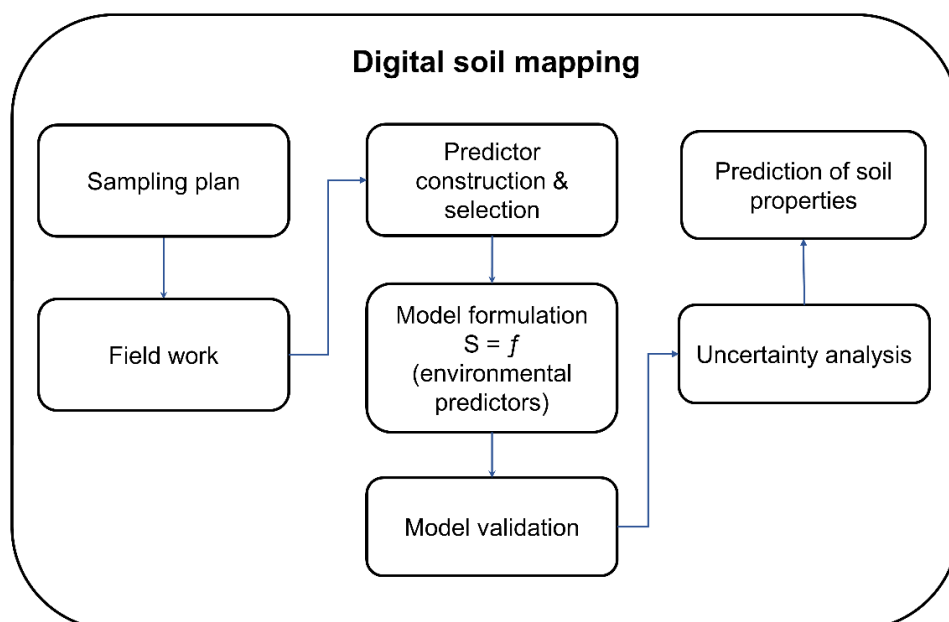
Figure 1.1 Digital soil mapping procedures. S: soil properties, f: models.

$$S = f(s, c, o, r, p, a, n) , \qquad (1.1)$$

S= soil (Sc for soil classes or Sp for soil properties); s= existing soil information; c= climate; r= relief (topographic attribute); p= parent material; a= age; n= spatial position

Several of methods for spatial soil modelling f() are available: Multiple linear regression, generalized linear models, generalized additive models (GAM), ANN, classification and regression trees (CART) and RF were used to predict soil properties. Multiple linear regression is one of the most used methods (Grunwald, 2009; Grunwald, 2006; McBratney et al., 2003). This method requires several samples for good performance and is based on linear relationships between environmental predictors and soil properties. In mountain areas, however, the number of available soil samples is limited and the relationships are often non-linear. These issues are partially resolved by non-parametric machine learning techniques (Ballabio, 2009). The machine learning techniques have a common ability to consider nonlinear relationships between responses and independent variables and the ability can be used for complex soil and environmental factor relationships in the mountain area. However, the models have some difficult characteristics, namely; they are not easy to interpret, require to tune many parameters, and might be not computationally efficient for large datasets (Kuhn and Johnson, 2013). The predictive power, ease of use and interpretability change based on model's complexity (Table 1.1).

CART, GAM, RF and SVR were used in predicting mountain soil properties and showed good results for digital soil mapping (Ballabio, 2009; Ließ et al., 2012; Tesfa et al., 2009) because the tuning of parameters may fit data from different environmental conditions well and allow the construction of appropriate models. The available machine learning models have been compared for better predictions (Table 1.1) (Manuscript 1).

Table 1.1 Summary of the strengths and weaknesses of different predictive models.

	Strengths	Weaknesses
LM	High ease of use, high interpretability, high parsimony	Low predictive power, low flexibility, low robustness to outliers
GAM	High computational efficiency, rather high flexibility	Low parsimony, low Interpretability, low robustness to outliers
RF	High ease of use, high computational efficiency, high robustness to outliers, high predictive power	Low parsimony, low interpretability
SVR	High flexibility, high predictive power	Low parsimony, low interpretability, low computational efficiency

LM: linear models, GAM: generalized additive models, RF: random forest, SVR: support vector regression. Source: adapted from Hastie et al. (2009), Minasny and Hartemink (2011), Hjort and Luoto (2013), and Elith and Franklin (2013).

In general, the prediction result is made better with more predictors but this can lead to over-fitted models which fit random noise of data instead of the underlying pattern due to the extremely complex model and too many predictors in comparison with the number of samples (Hastie et al., 2009; Hjort and Luoto, 2013; Kuhn and Johnson, 2013; Park and Vlek, 2002). Predictor selection is a critical step of regression methods, because this approach can reduce the number of predictors and produce the optimized model (Guyon and Elisseeff, 2003). Although various selection methods have been suggested, recursive feature elimination (RFE) performed improved results in digital soil mapping (Ballabio, 2009; Brungard et al., 2015). This is because RFE can avoid checking the whole combinations of predictors and effectively remove uninformative predictors.

Models represent reality as simplified patterns but might overemphasize patterns (Kuhn and Johnson, 2013). Therefore, assessing models should be required for modelers to obtain the reproducible pattern in the data. For model validation, different methods of splitting data into tuning and test datasets have been proposed. K-fold cross-validation (CV) is a resampling method partitioning the sample into k subsets, evaluating the model on one subset and training

the model by the remaining subsets, and repeating this procedure K times. Generally, modelers use the averaged performance result. They provide standard errors which can be used for uncertainty analysis of digital soil mapping. Methods to analyse the digital soil map uncertainty include monte carlo simulation, bootstrapping, and cross-validation (Minasny and Bishop, 2008).

1.2.2 Environmental predictors: terrain analysis and remote sensing

Climate and parent material are key factors for soils in the global, continental, and regional scale, while soils are locally determined by vegetation and topography over smaller areas (Gerrard, 2000). Most studies used different parameters from the Digital Elevation Model (DEM) such as elevation, slope angle, aspect, upslope contributing area (specific catchment area), profile and plan curvature, wetness index, and stream power index as predictors. Only using terrain predictors can product digital soil maps with the good accuracy (Ballabio, 2009). Moore et al. (1993) studied A horizon depth and soil texture using a DEM. McKenzie and Ryan (1999) reached relatively high predictive values for total phosphorus (R^2= 0.78) and soil carbon (R^2= 0.54). Gessler et al. (2000) accounted for between 52 and 88% of soil property variance, such as soil depth, A horizon depth, and soil carbon.

Spatial distributions of vegetation are related to the spatial heterogeneity in soil resource distributions (Binkley and Fisher, 2012). Soil development is also influenced by vegetation, which means vegetation coexists with soil as part of a feedback system (Ballabio et al., 2012). LiDAR (Light detection and ranging) metrics and normalized difference vegetation indices can be used as vegetation predictors. LiDAR is a remote sensing technology which has structural information on the illuminated surface, including 3D terrain, vegetation canopy information, and object heights. Especially, precisive DEMs and forest structure predictors from LiDAR data can improve the quality of soil nutrient models (Manuscript 3).

1.2.3 Scale: neighbourhood size effects

Scale issues are important to soil, geography, ecology and earth science (Phillips, 1999). Scale mainly consists of cartographic scale, process scale and analysis scale (Montello, 2001). Cartographic scale refers to the proportion of size of an object on a map relative to the real size in the world. Process scale is the size at which a physical earth structure or process operates irrespective of how they are researched (Montello, 2001). This scale is related with the spatial extent based on the characteristic of the natural process. Analysis scale is the size of the measure unit such as pixel size of digital elevation data (resolution) (Zhang et al., 2013).

In perspective to analysis scale, researchers should choose the suitable grid size of environmental predictors, measurement scale and the extent of research area in digital soil mapping to investigate the relationships between predictors and soil properties. Many studies found scale-dependency of soil–environmental factor relationships in various environmental conditions. However, the issue of scale has not been solved perfectly. There have been active debates associated with different methods and perspectives in the generation, analysis, and selection of DEMs (Kim and Zheng, 2011). Many studies have tried to find the right analysis scale. This right scale was simulated by constantly increasing either the cell size of DEMs or the neighborhood size (Drăguţ et al., 2009). About 15 – 30m grid sizes are recommended in many studies for digital soil mapping (Erskine et al., 2007; Kim and Zheng, 2011; Maynard and Johnson, 2014; Park et al., 2009). Cavazzi et al. (2013) investigated the interacting effects between window and grid sizes and also suggested the choice of fine scale DEMs in digital soil mapping is not always best. There has been little consideration of the effects of the neighborhood size for prediction of soil nutrients.

Effects of the neighborhood extent have an influence on values of terrain predictors (e.g. slope, aspect, and curvature) and finally on results of digital soil mapping (MacMillan and Shary, 2009; Maynard and Johnson, 2014; Wood, 2009). It is normally calculated based on 3x3 cell

neighborhood extents that can only consider the local variability. In the large extent, it is more similar to landform elements such as upper slope, linear slope, and lower slope which is critical to understand spatial patterns of soil nutrients as well as potentials of mountainous ecosystems. Therefore, effects of the neighborhood extent should be investigated for digital soil mapping (Manuscript 2).

1.3 Soil nutrient dynamics of mountainous ecosystems under nitrogen deposition

When the supply of nutrients is not enough relative to the demands of the plant, growth is limited by the availability of the most limiting nutrients, which is called as Liebig's law of the minimum (Craine, 2009). Growth can be promoted only by increasing the supply of the limiting nutrients according to this Liebig's law. Growth increases linearly with the rate of addition of limiting nutrients in experiments and also responds to more limiting nutrients in the field (Chapin and Eviner, 2013).

Nitrogen (N) and phosphorus (P) commonly limit the terrestrial primary production (Vitousek et al., 2010, 2002). N and P limitation change over the course of soil development. Primary productivity is limited by N availability in young soils but increasingly by P availability in old soils as parent materials are weathered and P is lost via leaching over millions of years (Laliberté et al., 2013). N limitation might be changed into P limitation by high anthropogenic N deposition during short-term periods in terrestrial ecosystems (Braun et al., 2010; Vitousek et al., 2010). This might also occur in Korea, since nitrogen inputs have increased rapidly in Korea due to huge industrial operations and intense agricultural activities (Jang et al., 2011; Kim et al., 2014b; Kim et al., 2011). Therefore, identifying soil nutrient dynamics such as N and P is critical to understand environmental changes of mountainous ecosystems in South Korea.

Various P fractions are important to understand P cycles in soils (Cross and Schlesinger, 1995; Yang and Post, 2011). Especially, Hedley P fractionations can be useful because P fractions

from Hedley method provide a comprehensive status of soil P pools which consist of organic and inorganic P ranging from available P to stable P according to accessibilities by plants. The biogeochemical cycle of P in soils is complex. P transformations during the course of pedogenesis are toward the P pools consisting of stable organic P and residual P (Walker and Syers, 1976). During this course, each P fraction can be distributed across the landscape, which can exert important influences on potential P availability and limitation for plants (Smeck, 1985). The spatial pattern of P fractions from Hedley procedure depending on topographic sequences was found (Agbenin and Tiessen, 1994; Araújo et al., 2004; Roberts et al., 1985; Smeck, 1985; Tiessen et al., 1994; Vitousek et al., 2003). However, relationships between various soil P fractions and environmental predictors have not yet been fully understood and spatial prediction of P Hedley fractionations was not tried (Manuscript 2).

Under low nutrient available environments such as mountainous ecosystems, an understanding of the organic layer nutrients is critical for sustainable ecosystem management. There are forest floors in mountain areas. The organic layers contain large stocks of soil nutrients which are changed organic to inorganic forms by mineralization of litter and organic matter (Wilcke et al., 2010). Many researches tried to make the spatial explicit prediction of mineral soil nutrients such as nitrogen (Peng et al., 2013; Pastick et al., 2014; Uriarte et al., 2015) and phosphorus (Agbenin and Tiessen, 1994; Araujo et al., 2004; Mage and Porder, 2013) in forest areas. Wilcke et al. (2002) reported the amount of nutrients and turnover time in organic layers of Ecuadorian tropical montane forest. Wilcke et al. (2008) also confirmed elevation gradient with decreasing contents of N and P in organic layers and the correlation between macronutrients (N and P) and tree growth. Soethe et al. (2008) found the stocks of N in the organic layer was significantly differ by elevation but N to P ratio did not response with increasing elevation in a tropical montane forest. Understanding relationships between spatial variation in organic layer nutrient contents (especially P) and variation in environmental factors was limited (Table 1.2) (Manuscript 3).

Table 1.2 Selected studies on digital soil mapping of regression approaches for N and P.

Reference	Study area (km^2)	Land use	Soil	Predictors	Models
Moore et al. (1993)	0.05	C	avP (M)	Slope, wetness index, stream power index	linear regression
McKenzie and Ryan (1999)	500	F	TP (M)	Slope, curvature, relief, gamma radiometrics	Regression tree
Johnson et al. (2000)	2.4	F	TN (O & M)	Elevation, slope, wetness index, flow accumulation	Linear regression
Ryan et al. (2000)	2.7 & 484	F	TP (M)	Aspect, curvature, gamma radiometrics, landsat TM, magnetic intensity	Regression tree, linear regression
Seibert et al. (2007)	Sweden	F	TN (O & M)	Elevation, upslope area, slope, wetness index	Correlation analysis
Sumfleth and Duttmann (2008)	10	C	TN (M)	Elevation, altitude above channel network	Regression kriging
Kunkel et al. (2011)	16	F & G	TN (M)	Vegetation index, Solar radiation	Linear regression
Kim and Zheng (2011)	0.05	SD	TP (M)	Elevation, slope, aspect, curvature, upslope area, wetness index	Spatial regression
Kim et al. (2014c)	418	W	TN & TP (M)	Various vegetation index, elevation, lithology, hydrology	Random forest
Roger et al. (2014)	1670	C & G	P fra. (M)	Elevation, slope, wetness index, relief, land use, curvature	Regression kriging

F: forest area, C: crop area, G: grassland, SD: sand dune, W: wetland, TN: total nitrogen, TP: total phosphorus, avP: available phosphorus, P frac.: P fractions, O: organic horizon, M: mineral horizon.

1.4 Complex TERRain and ECOlogical heterogeneity (TERRECO)

This PhD research is a part of complex TERRain and ECOlogical heterogeneity (TERRECO) which is the international research training group to focus on the evaluation of ecosystem services in mountainous landscapes. Especially, an assessment framework will be developed to quantify trade-offs (e.g. between crop production and water quality) to support human well-being and will be applied to determine how shifts in climate and landuse in complex terrain influence naturally derived ecosystem services (Kang and Tenhunen, 2010). Spatial information on site characteristics such as nutrient availability which determine forest growth and agricultural production is essential for ecosystem services of complex terrain region.

1.5 Overview of this thesis

1.5.1 Objectives

In order to investigate the spatial soil nutrient distribution pattern of the Soyang watershed, digital soil maps were developed, which include estimates of map uncertainty. Specific purposes were: (1) to develop maps of soil nutrients for land potential assessment, (2) to investigate spatial patterns of various P fractions, and (3) to predict N to P ratios in the topsoil layer.

Two different approaches were used for (1) less detailed soil nutrient maps of the whole Soyang watershed for land potential assessment and (2) high precision soil nutrient maps for a subarea of the watershed providing detailed environmental information in order to identify spatial patterns of soil nutrients in a mountainous watershed only vegetated with forest areas. Available LIDAR data and a high resolution remote sensing image (4 m Komsat-2) for this subarea provided a possibility for the development of these high precision maps.

Study 1: Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain

Spatial distributions of topsoil carbon (C), N and available P in mountain regions were identified using supervised learning methods, and a functional landscape analysis was performed in order to determine the spatial soil fertility pattern for the Soyang Lake watershed in South Korea. Specific research purposes were (1) to identify important predictors; (2) to develop digital soil maps; (3) to assess land potentials using digital soil maps.

Study 2: Spatial patterns of soil phosphorus fractions in a mountainous watershed

We used digital soil maps in order to investigate the spatial distribution of different P fractions. Specific research purposes were (1) to identify the important environmental predictors, and (2) to map different P fractions using the quantitative relationships between P fractions and important predictors.

Study 3: Spatial topsoil N:P ratios under monsoon conditions in a complex terrain of South Korea

In order to understand the spatial patterns of organic layer and mineral soil N and P, digital soil maps were developed using LiDAR DEM (digital elevation model) and vegetation parameters as predictors. The specific aims of our research were (1) to test the importance of LiDAR-derived vegetation and topographical parameters to understand the spatial N and P patterns, (2) to identify subareas with critical P contents, and (3) to test different validation strategies for N and P depending spatial uncertainty structures.

1.5.2 Study area

This thesis was researched in two watersheds. Firstly, the Soyang lake watershed is located in the north-eastern part of Gangwon-do province, South Korea (Figure 1.2 C) (Manuscript 1). It extends between 70 and 1700 m a.s.l. and covers an area of 2,776 km². Soyang lake, impounded by an artificial dam in 1973, is located about 10 km northeast of Chuncheon. The lake was constructed for flood control, water supply and hydroelectric power generation for downstream areas. Particularly, in the northern part of the watershed there are several no-go-areas contaminated by landmines and military facilities due to the close vicinity to North Korea.
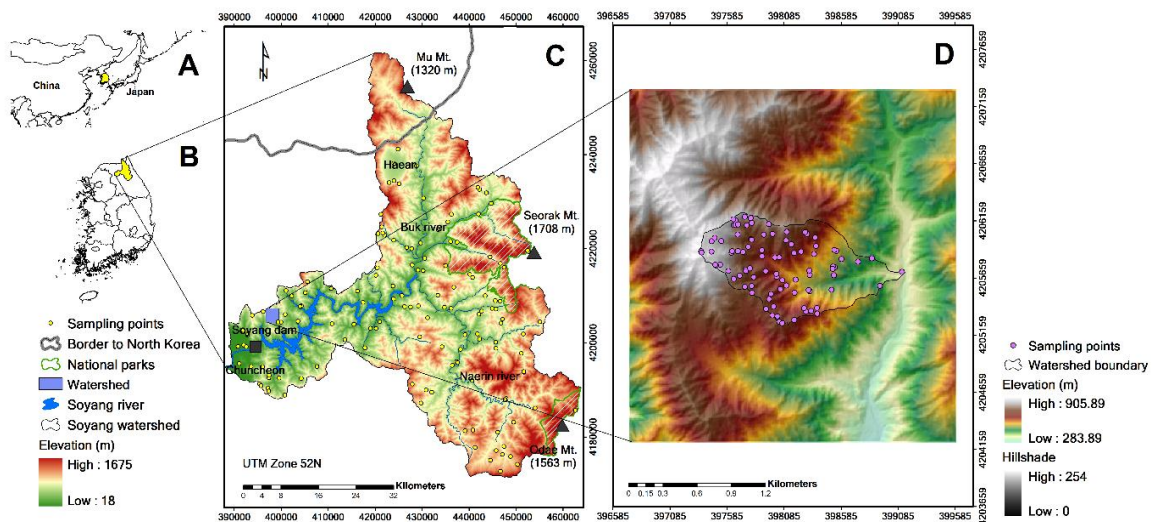
Figure 1.2 Research areas. (A) The map of the Korean Peninsula. (B) The Soyang watershed is located in the north-eastern part of South Korea. (C) The map shows the spatial distributions of sampling points for the first manuscript. The Soyang watershed is near the border to North Korea and includes two national parks (Seorak and Odae). The Soyang river originates from Mu, Seorak and Obdae Mountain. (D) The research area with the sampling points for the second and third manuscripts.

It receives an average of 1,179 mm of mean annual rain fall. About 70 % of the annual rain falls heavily in the summer monsoon season (Bartsch et al., 2014). The whole physical geographical feature is an incised meander that has a narrow river valley due to tectonic process during the Quaternary (Lee, 2004). The Soyang river originates on several mountains of the watershed. The watershed includes two national parks (Seorak and Odae).

The research area is mostly covered by forest (80%) with deciduous trees (51.8%), coniferous (25.4%), and mixed (22.8%) trees as dominant forest types. The areas geology is dominated by banded gneiss and granite (Korea Institute of Geology Mining and Material, 2001). Chuncheon and Haean consist of granite and show plain areas. Moderately coarse textured soil and clay loam soil cover around 60% of the area (National Academy of Agricultural Science, 2013).

Our second study area is located in the downstream area of the Soyang watershed (Figure 1.2 D) (Manuscript 2 and 3). During 30 years, a mean annual temperature is 11.1 °C (-4.6 – 24.6 °C) and a mean annual rainfall is 1,347 mm, with about 824.4 mm falling between June, July and August (Korea meteorological administration, 2015). The area's geology is dominated by granitic gneiss and banded gneiss. The area is 9.84 km^2 and elevation ranges between 320 and 868 m above sea level. There are various toposequences with steep slopes (over 45°). It is in a headwater catchment that has narrow depositional areas (Wohl, 2010). Its soils are mainly composed of fine gravelly sandy loam soils and fine sandy loam and gravelly loam soils (National Academy of Agricultural Science, 2013). It is a national forest mainly vegetated by Mongolian oak (*Quercus mongolica*) and Korean pine (*Pinus koraiensis*) vegetation.

1.5.3 Methods and results

1.5.3.1 Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain

Spatial distributions of topsoil C, N, and available P in mountain regions were identified using supervised learning methods, and a functional landscape analysis was performed in order to determine the spatial land potentials. 139 surface soil samples were collected by conditioned Latin Hypercube Sampling (cLHS) which is a stratified sampling design to obtain a representative dataset for the Soyang Lake watershed. cLHS was applied to guarantee for optimal coverage of the variability of environmental covariates in feature space (Minasny and McBratney, 2006). Impractical sampling designs are common in the cLHS due to difficult accessibility and sparely distributed locations within the study area (Roudier et al., 2012). Therefore, the sampling method for this research considered operational field constraints such as accessibility and no go areas contaminated by landmines as well as budget limitations. Terrain parameters and different vegetation indices were derived for predictors. We compared a generalized additive model (GAM) to random forest (RF) and support vector regression (SVR).

Predictor selection was used based on the recursive feature elimination (RFE) which calculates ranks of each predictor based on predictor importance measure, removes one lowest important predictor from the model including all predictors, and repeats this step until the only one most important predictor is left (Kuhn and Johnson, 2013). Performances of the combinations of predictors (all predictors to the most important predictor) are evaluated using root mean square error (RMSE) or R^2. A land potential assessment for soil nutrients was performed using trimmed k-mean cluster analysis which is a robust method and can handle the dataset with extreme values.

Vegetation indices showed most powerful predictabilities for soil nutrients. RF showed the best result among the three supervised learning methods. RF had better predictability and easier model construction and interpretability than SVR. GAM showed larger uncertainty concerning the variability of RMSE. Therefore, it is not easy to make the generalization of relationships between soil nutrients and environmental predictors using GAM. Finally, RF was selected for the prediction model of all three soil nutrients.

The areas dominated by high density of vegetation and deciduous forest at higher elevation showed higher C and N contents. C and N contents had strong correlations with vegetation indices. Higher contents of topsoil available P were found in the lower plain areas dominated by rice paddies. Soil P distributions were disturbed by historical landuse changes over natural processes. Spatial soil nutrient patterns were changed with effects of the vegetation and landuse and locally governed by topographical gradients in our research area.

Cluster analysis identified four land potential classes: fertile (C2), medium (C3) and low fertile (C4) forest lands with an additional class (C1) dominated by high P and low C and N contents due to human impact. C2 and C3 showed relatively high mean C and N contents based on other researches. The results showed C1 has relatively high P contents and C2, C3 and C4 had low P contents. Most of the areas (forest) showed low phosphorus contents. Table 1.3 shows the core findings of the manuscript 1.

Table 1.3 Highlights for manuscript 1.

- Vegetation indices have powerful abilities to predict soil nutrients
- Using only selected predictors via recursive feature elimination improves prediction results
- An effective approach to map land potentials for mountain ecosystem management

1.5.3.2 Spatial patterns of soil phosphorus fractions in complex terrain

Digital soil maps were developed in order to quantify the spatial variability of different P fractions. The number of 91 soil samples was collected from the A horizon in 2014. cLHS was applied to get a representative dataset. Effects of the neighborhood extent for surface curvature were investigated. Solutions with 3x3 to 35x35 window sizes (30 ~ 350 m neighborhood extents) were tested and the best solution was selected based on results of Pearson's correlation. Multiple linear regression model was applied for the spatial explicit prediction and recursive feature elimination was compared with stepAIC function. Model performances were tested using 5 repetitions of 10-fold cross validation. Finally, 50 models were provided for each P fraction. One-way analysis of variance (ANOVA) was applied to investigate significant difference in P fractions depending on landform elements. Kruskal-Wallis test and post-hoc test were used due to non-normality and heterogeneity of variances of the dataset.

Surface curvature was sensitive to changing the neighborhood extent and 19x19 window size was best for most P fractions. Values of Pearson correlation analysis between P fractions and surface curvature were largely changed based on neighborhood extents. This implied neighborhood extents strongly contributed to predictability of our P models. Surface curvature (19x19) was the best predictor for most P fractions. Total P showed the highest R-square while resin P and residual P showed lower results.

The concentrations of all P fractions changed with the gradients of surface curvature and elevation in results of ANOVA. The total soil P contents showed clear downslope increases both at the low (300-600m) and high altitude (600-900m). Especially, the proportion of organic P was enriched at the lower slope, while the proportion of residual P concentrations at the upper slope were higher. Interestingly, two proportions had an inverse relation. Additionally, resin-P, a bioavailable P for plants, showed a downslope increase. Available P increased with

increasing elevation. Soil erosion process with the steep slope and heavy monsoon rains might result in the spatial differentiation of soil P fractions. Table 1.4 shows the core findings of the manuscript 2.

Table 1.4 Highlights for manuscript 2.

- There are only very few studies that used digital soil mapping for spatial patterns of soil P fractions in complex landscapes

- Our study revealed that topography influenced the abundance of soil P fractions

- The proportion of Organic P was enriched at the lower slope, while the proportion of residual P was increased upslope

- This has important implications for soil fertility in the mountainous ecosystems with low P availability

1.5.3.3 Spatial topsoil N:P ratios under monsoon conditions in a complex terrain of South Korea

Digital soil maps of N and P in both organic and mineral A horizons were developed using LiDAR terrain and forest structural predictors. The number of 91 soil samples was used. Terrain predictors were calculated with the open source software SAGA and the CURV3 program. All analyses for a set of vegetation predictors and a DEM from LiDAR data were performed within the commercial software 'LAStools'. NDVI was constructed. RF for the prediction model and RFE for predictor selection were used. For the test of different cross validation (CV) strategies, 2, 5, 10, 20-fold and leave-one-out CV in n repetitions were explored for best validation methods. Each k-fold CV was repeated 50 (2-fold), 20 (5-fold), 10 (10-fold), and 5 (20-fold) times. The total number of external validation was set at 100. For N and P, 100 R–Squares and RMSE were calculated and the standard deviations of the 100 outputs represent uncertainty of spatial explicit predictions. All spatial predictions were done with 10 m resolution.

In the result, the 10 repetitions 10-fold method showed relatively good R^2 (0.23 – 0.52) and low standard deviation. Spatial patterns in maps of soil nutrients were not quite different depending on CV methods. However, standard deviations (uncertainty) of N and P decreased with increasing the number of calibration dataset. Therefore, repeated 10-fold CV is recommended for the model validation using small size samples like our research.

Surface curvature had the highest predictor importance for P contents in organic and mineral A horizons. LiDAR vegetation predictors and vegetation index showed a strong relationship with N in the organic layer. Models for P showed better results compared to N models. N to P ratios in the organic and mineral A horizons increased in close vicinity to the upper slope due to soil erosion process. Based on N to P ratios, the upper slope areas might be affected by higher phosphorus limitation under high nitrogen deposition. Table 1.5 shows the core findings of the manuscript 3.

Table 1.5 Highlights for manuscript 3.

- Repeated 10-fold CV is recommended for small sample sizes in digital soil mapping

- Surface curvature was the best predictor, while LiDAR metrics and vegetation index were selected for N in the organic layer

- Higher values of N to P ratios in the organic and mineral A horizons were found at the upper slope

1.6 Conclusions and discussion

1.6.1 Supervised learning methods: What is the best prediction model for digital soil mapping? (Manuscript 1 and 3)

Various supervised learning methods can be used for digital soil mapping. The methods from linear regression, generalized additive models (GAM), random forest (RF), and support vector

regression (SVR) were used for this thesis. Among these methods, RF showed powerful modelling performances as well as the ease of model construction and interpretability (Manuscript 1). However, RF don't have high interpretability compared with linear regression or classification regression trees. The model structure of "block box" machine learning can be interpreted to extract rules from the predicted model (Barakat and Diederich, 2004). Ballabio (2009) used decision trees to extract the understandable structure from the SVR result of the soil prediction. Additionally, RF has another weakness which is not good for measure data with small portions of extreme values (Kuhn and Johnson, 2013). The model tends to underestimate for samples in the high or low extreme values because RF uses the average of the data in the terminal (leaf) nodes. Alternatively, cubist uses linear models to predict the outcome in the leaf nodes. Although cubist was successful for applications of digital soil mapping (Adhikari and Hartemink, 2015; Adhikari et al., 2014), the prediction accuracy and uncertainty of cubist should be assessed for small sample sizes due to possibilities of over-fitting with other models in the next study.

1.6.2 Scale issues: What can we learn from scale for digital soil mapping? (Manuscript 1, 2, and 3)

Digital soil mapping can identify spatial patterns of soils, but may be influenced by scale parameters, including the spatial extent. We found the effects of the spatial extent. Vegetation indices were the most important predictors in the Soyang watershed (Manuscript 1), while topography predictors strongly contributed to explaining soil nutrient patterns in the small watershed (Manuscript 2 and 3). For N, most important predictors were vegetation indices (NDVI and GNDVI) and elevation in the Soyang watershed and elevation and surface curvature in the small watershed. All vegetation indices (NDWI, NDVI, and GNDVI) were selected as important predictors for available P in the Soyang watershed, while elevation and surface curvature showed powerful predictabilities in the small watershed.

Different predictors were selected for models in the both watersheds (Manuscript 1, 2 and 3). This is because key factors are different in the both watersheds. Soyang watershed is mainly covered by various landuse types including forest, deforested areas, and agricultural areas (Kim et al., 2014a). Various vegetation indices can be considered as indicators of landuse and forest types (Jones and Vaughan, 2010). Most Ah horizon was disappeared due to soil erosion in the deforested areas which had lower N contents compared to typical forest soils. Moreover, there are various forest types (deciduous, coniferous, and mixed forests) in the Soyang watershed (Kim et al., 2014a). Vegetation indices can be powerful for the spatial prediction of soil nutrients under the environmental conditions (Mulder et al., 2011). High C and N contents found at higher elevation and in forest areas with higher vegetation density which can be also measured by vegetation indices (Manuscript 1). In the small watershed, however, the variation of forest cover is not quite large. Vegetation index was not the best predictor to model N contents, while the strong environmental correlations between N and elevation were found
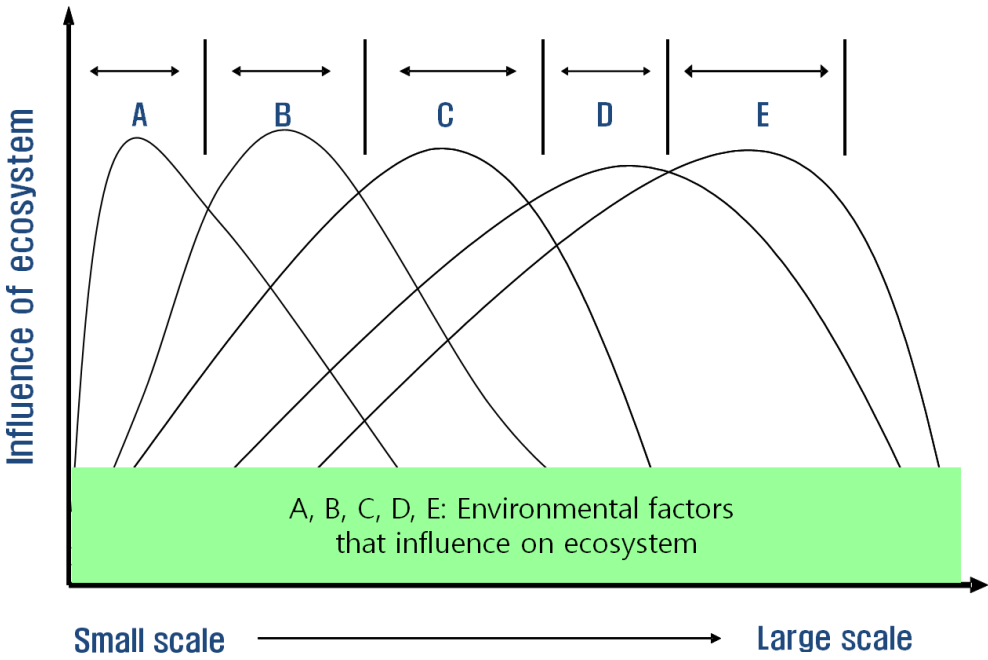
Figure 1.3 Process scale of environmental factors that influence on ecosystem (Jeong et al., 2012a).

(Manuscript 3). For available P, the spatial distribution also was related to the land use patterns due to fertilizer in the Soyang watershed (Manuscript 1). These areas with the larger amounts of phosphorus found in Chuncheon and Haean (Figure 1.2).

Depending on spatial extents, main environmental factors are exchanged because each factor is interconnected with other factors and exerts mainly influences on environmental phenomena (e.g. soils) in the optimum spatial extent scale (Gibson et al., 2000) (Figure 1.3). This implied that each main process is different based on each scale (spatial extents) (Phillips, 1999). Naturally, key predictors can be changed with increasing or decreasing the spatial extent. In our study areas, spatial patterns of soil nutrients were detected well with topography in the small watershed, while vegetation indices showed powerful predictability of the soil nutrients in the Soyang watershed.

Studies exploring the influence of the spatial extent on digital soil mapping are rare. Generally, it is expected that increasing the extent corresponds to increase complexity and hence decrease the quality of the model. However, Vasques et al. (2012) reported the quality of soil C models increased with increases in the spatial extent. In our results, the quality of soil nutrient models (R-square) slightly decreased with an increase in the spatial extent.

Effects of the neighborhood extent have an important role for digital soil mapping because these can improve the model accuracy (MacMillan and Shary, 2009; Maynard and Johnson, 2014; Wood, 2009). We investigated effects of the neighborhood extent (window size) for only surface curvature in a small watershed (Manuscript 2 and 3). Most soil nutrients were sensitive with changing neighborhood extents except N in the organic layer. P models showed results of the correlation coefficient (r) between -0.25 and -0.5. P was more sensitive with neighbourhood extents than N ($-0.15 \leq r \leq -0.1$). Among P fractions, different results showed.

The correlation coefficients of organic P showed remarkably increased -0.25 to -0.48, while resin P's slightly changed -0.1 to -0.28 with neighborhood extents. Most nutrients showed best

results in an extent of the 190 m which might adequately represent the spatial land surface configuration of our research area determining soil processes. In this thesis, the neighborhood effect of only surface curvature was explored. The effect for slope, aspect, and plan and profile curvature and various combinations should be investigated further to predict soil nutrients.

1.6.3 Soil-vegetation-topography interactions for nutrient dynamics (Manuscript 2 and 3)

Topography exerted an influence on soil P in our study area. P availability in soils is locally influenced by mineral type, clay, pH, temperature, moisture and organic matter (Brady and Weil, 2010; Negassa and Leinweber, 2009). The factors related to P availability are controlled under topographical characteristics in the landscape-scale (Camargo et al., 2012; Vincent et al., 2014; Zhou et al., 2016). We found significant differences between P fractions at different topographic positions (manuscript 2). Higher contents of various P fractions were found at the lower slope. Especially, the proportion of residual P showed the opposite trend with that of organic P at different topographic positions. This suggested that potential sources for P availability at the upper slope are relatively poorer than at the lower slope because residual P is strongly stable and highly insoluble. Therefore, productivity at the upper slope might be limited by P under high rates of nitrogen deposition.

Identifying spatial patterns of nutrients in the topsoils is useful to understand soil-vegetation-topography interactions in the landscape-scale. By surface and subsurface flows, nutrients were moved along the hillslope. Moreover, soil erosion process with heavy rains and steep slopes strongly operates in the mountain watershed (Jeong et al., 2012b; Jung et al., 2012). With soil nutrients, water availability is a key factor determining the distribution and growth of plants. Generally, soils at the convex upper slope are drier than might be expected because water is diffusing in the slopes, while soils at the concave lower slope tend to be wetter (Park and Van De Giesen, 2004). Accordingly, higher nutrient (P) and moisture availability at the

lower slope lead to a higher productivity that results in the accumulation of organic matter (Agbenin and Tiessen, 1994). Moreover, it could be that the organic matter at the lower slope has higher P contents due to high P contents in the plant litter inputs (Manuscript 2). On the other hand, lower soil P contents and P availability at the upper slope might produce P-deficient leaves which in turn result in higher N to P ratios in the organic layer. As a result, the higher plant litter nutrient contents can cause lower N to P ratios in organic layers at the lower slope (Manuscript 3). This implied that topography, soil, and vegetation might be strongly interconnected (Amundson et al., 2015), especially, under steep slopes and monsoon conditions and the relationships can allow to expand the understanding of nutrient cycles (e.g. P) (Zhou et al., 2016).

1.7 List of manuscripts and specification of individual contributions

The three studies in the thesis refer to different manuscripts. Three manuscripts were submitted.

Manuscript 1 (Chapter 2)

Authors: Gwanyong Jeong, Hannes Oeverdieck, Soo Jin Park, Bernd Huwe, Mareike Ließ

Title: Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain

Journal: Catena

Status: under review

Own and author contributions statement:

Own contribution: concept and study design 50%, data acquisition 50%, analyses of samples 80%, data analyses and figures 100%, discussion of results 80%, manuscript writing 70%

G. Jeong, H. Oeverdieck, S.J. Park, B. Huwe, and M. Ließ designed the research; **G. Jeong and H. Oeverdieck** performed the research; Samples were analysed in the BayCEER lab and the Agricultural Technology Center of Yanggu, South Korea; **G. Jeong** analyzed the data; **G. Jeong and M. Ließ** interpreted and discussed results; Figures and tables were created by **G. Jeong**; **G. Jeong** wrote the first draft of the manuscript; Revision and rewriting of the manuscript was done by **G. Jeong, B. Huwe, and M. Ließ**.

G. Jeong is the corresponding author.

Manuscript 2 (Chapter 3)

Authors: Gwanyong Jeong, Marie Spohn, Soo Jin Park, Bernd Huwe, Mareike Ließ

Title: Spatial patterns of soil phosphorus fractions in a mountainous watershed

Journal: Catena

Status: under review

Own and author contributions statement:

Own contribution: concept and study design 70%, data acquisition 90%, analyses of samples 70%, data analyses and figures 100%, discussion of results 70%, manuscript writing 70%

G. Jeong, M. Spohn, S.J. Park, B. Huwe, and M. Ließ designed the research; **G. Jeong** performed the research with support from 2 Hiwis; Samples were analysed in Eurofins, Jena and laboratory of isotope biogeochemistry, BayCEER; **G. Jeong** analyzed the data; **G. Jeong and M. Spohn** interpreted and discussed results; Figures and tables were created by **G.**

Jeong; **G. Jeong** wrote the first draft of the manuscript; Revision and rewriting of the manuscript was done by **G. Jeong, M. Spohn, B. Huwe, and M. Ließ**.

G. Jeong is the corresponding author.

Manuscript 3 (Chapter 4)

Authors: Gwanyong Jeong, Kwanghun Choi, Marie Spohn, Soo Jin Park, Bernd Huwe, Mareike Ließ

Title: Spatial topsoil N:P ratios under monsoon conditions in a complex terrain of South Korea

Status: In preparation for publication

Own and author contributions statement:

Own contribution: concept and study design 70%, data acquisition 90%, analyses of samples 70%, data analyses and figures 100%, discussion of results 70%, manuscript writing 70%

G. Jeong, K. Choi, M. Spohn, S.J. Park, B. Huwe, and M. Ließ designed the research; **G. Jeong and K. Choi** performed the research with support from 2 Hiwis; Samples were analysed in Eurofins, Jena and laboratory of isotope biogeochemistry, BayCEER; **G. Jeong** analyzed the data. **G. Jeong, M. Spohn, and B. Huwe** interpreted and discussed results; Figures and tables were created by **G. Jeong**; **G. Jeong** wrote the first draft of the manuscript; Revision and rewriting of the manuscript was done by **G. Jeong, K. Choi, M. Spohn, B. Huwe, and M. Ließ**.

G. Jeong is the corresponding author.

1.8 References

Adhikari, K., Hartemink, A.E., 2015. Digital Mapping of Topsoil Carbon Content and Changes in the Driftless Area of Wisconsin, USA. Soil Sci. Soc. Am. J. 79, 155. doi:10.2136/sssaj2014.09.0392

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

Agbenin, J.O., Tiessen, H., 1994. Phosphorus transformations in a toposequence of lithosols and cambisols from semi-arid northeastern brazil. Geoderma 62, 345–362. doi:10.1016/0016-7061(94)90098-1

Amundson, R., 2014. Soil Formation, in: Holland, H., Turekian, K. (Eds.), Treatise on Geochemistry. Academic Press, San Diego, CA, pp. 1–26. doi:10.1016/B0-08-043751-6/05073-8

Amundson, R., Heimsath, A., Owen, J., Yoo, K., Dietrich, W.E., 2015. Hillslope soils and vegetation. Geomorphology 234, 122–132. doi:10.1016/j.geomorph.2014.12.031

Araújo, M.S.B., Schaefer, C.E.R., Sampaio, E.V.S.B., 2004. Soil phosphorus fractions from toposequences of semi-arid Latosols and Luvisols in northeastern Brazil. Geoderma 119, 309–321. doi:10.1016/j.geoderma.2003.07.002

Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma 151, 338–350. doi:10.1016/j.geoderma.2009.04.022

Ballabio, C., Fava, F., Rosenmund, A., 2012. A plant ecology approach to digital soil mapping, improving the prediction of soil organic carbon content in alpine grasslands. Geoderma 187-188, 102–116. doi:10.1016/j.geoderma.2012.04.002

Barakat, N., Diederich, J., 2004. Learning-based Rule-Extraction from Support Vector Machines. Int. J. Comput. Intell. 2, 59–62. doi:10.1007/978-3-540-75390-2

Bartsch, S., Frei, S., Ruidisch, M., Shope, C.L., Peiffer, S., Kim, B., Fleckenstein, J.H., 2014. River-aquifer exchange fluxes under monsoonal climate conditions. J. Hydrol. 509, 601–614. doi:10.1016/j.jhydrol.2013.12.005

Benner, J., Vitousek, P.M., Ostertag, R., Bruijnzeel, L.A., Scatena, F.N., Hamilton, L.S., 2010. Nutrient cycling and nutrient limitation in tropical montane cloud forests, in: Bruijnzeel, L., Scatena, F., Hamilton, L. (Eds.), Tropical Montane Cloud Forests, Science for Conservation and Management. Cambridge University Press, New York, pp. 90–100.

Binkley, D., Fisher, R., 2012. Ecology and management of forest soils. John Wiley & Sons, West Sussex.

Bishop, C.M., 1995. Neural networks for pattern recognition. Clarendon Press, Oxford.

Brady, N.C., Weil, R.R., 2010. Elements of the nature and properties of soils. Prentice Hall, New Jersey.

Braun, S., Thomas, V.F.D., Quiring, R., Flückiger, W., 2010. Does nitrogen deposition increase forest production? The role of phosphorus. Environ. Pollut. 158, 2043–2052. doi:10.1016/j.envpol.2009.11.030

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239-240, 68–83. doi:10.1016/j.geoderma.2014.09.019

Burrough, P. a., van Gaans, P.F.M., MacMillan, R. a., 2000. High-resolution landform classification using fuzzy -means. Fuzzy Sets Syst. 113, 37–52. doi:10.1016/S0165-0114(99)00011-1

Camargo, L.A., Marques, J., Pereira, G.T., Alleoni, L.R.F., 2012. Spatial correlation between the composition of the clay fraction and contents of available phosphorus of an Oxisol at hillslope scale. Catena 100, 100–106. doi:10.1016/j.catena.2012.07.016

Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., Fealy, R., 2013. Are fine resolution digital elevation models always the best choice in digital soil mapping ? Geoderma 195-196, 111–121. doi:10.1016/j.geoderma.2012.11.020

Chapin, F.S., Eviner, V.T., 2013. Biogeochemical Interactions Governing Terrestrial Net Primary Production, 2nd ed, Treatise on Geochemistry. Academic Press, San Diego, CA. doi:10.1016/B978-0-08-095975-7.00806-8

Craine, J.M., 2009. Resource strategies of wild plants. Princeton university press, New Jersey.

Cross, A. F., Schlesinger, W.H., 1995. A literature review and evaluation of the Hedley fractionation: Applications to the biogeochemical cycle of soil phosphorus in natural ecosystems. Geoderma 64, 197–214. doi:10.1016/0016-7061(94)00023-4

Dobos, E., Hengl, T., 2009. Soil mapping applications, in: Hengl, T., Reuter, H. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 461–479.

Drăguţ, L., Schauppenlehner, T., Muhar, A., Strobl, J., Blaschke, T., 2009. Optimization of scale and parametrization for terrain segmentation: An application to soil-landscape modeling. Comput. Geosci. 35, 1875–1883. doi:10.1016/j.cageo.2008.10.008

Elith, J., Franklin, J., 2013. Species Distribution Modeling, in: Levin, S. (Ed.), Encyclopedia of Biodiversity. Academic Press, San Diego, CA, pp. 692–705. doi:10.1016/B978-0-12-384719-5.00318-X

Erskine, R.H., Green, T.R., Ramirez, J. a., MacDonald, L.H., 2007. Digital Elevation Accuracy and Grid Cell Size: Effects on Estimated Terrain Attributes. Soil Sci. Soc. Am. J. 71, 1371. doi:10.2136/sssaj2005.0142

Funnell, D., Parish, R., 2005. Mountain environments and communities. Routledge, London.

Gerrard, J., 2000. Fundamentals of soils. Routledge, New York.

Gessler, P.E., Chadwick, O. a., Chamran, F., Althouse, L., Holmes, K., 2000. Modeling Soil–Landscape and Ecosystem Properties Using Terrain Attributes. Soil Sci. Soc. Am. J. 64, 2046. doi:10.2136/sssaj2000.6462046x

Gibson, C.C., Ostrom, E., Ahn, T.K., 2000. The concept of scale and the human dimensions of global change: A survey. Ecol. Econ. 32, 217–239. doi:10.1016/S0921-8009(99)00092-0

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

Grunwald, S., 2006. Environmental soil-landscape modeling: Geographic information technologies and pedometrics. Taylor & Francis, Boca Raton, Boca Raton.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Hastie, T.J., Tibshirani, R.J., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer, New York.

Hjort, J., Luoto, M., 2013. Statistical Methods for Geomorphic Distribution Modeling, in: Shroder, J., Baas, A.C.W. (Eds.), Treatise on Geomorphology. Academic Press, San Diego, CA, pp. 59–73. doi:10.1016/B978-0-12-374739-6.00028-2

Jang, S.-K., Sung, M.-Y., Shin, A.-Y., Choi, J.-S., Son, J.-S., Ahn, J.-Y., Kim, J.-C., Shin, E.-S., 2011. A Study for Long-term Trend of Acid Deposition in Korea. J. Korea Soc. Environ. Adm. 17, 183–192.

Jenny, H., 1941. Factors of soil formation: A sytem of quantitative pedology. McGraw-Hill, New York.

Jeong, G.Y., Yang, H.M., Kim, S.K., Park, S.J., 2012a. Ecoregion Classification using Multi-Hierarchy of Environmental Factors. J. Korean Geogr. Soc. 47, 654–676.

Jeong, J.J., Bartsch, S., Fleckenstein, J.H., Matzner, E., Tenhunen, J.D., Lee, S.D., Park, S.K., Park, J.H., 2012b. Differential storm responses of dissolved and particulate organic carbon in a mountainous headwater stream, investigated by high-frequency, in situ optical measurements. J. Geophys. Res. Biogeosciences 117, 1–13. doi:10.1029/2012JG001999

Johnson, C.E., Ruiz-Mendez, J.J., Lawrence, G.B., 2000. Forest Soil Chemistry and Terrain Attributes in a Catskills Watershed. Soil Sci. Soc. Am. J. 64, 1804–1814.

Jones, H.G., Vaughan, R.A., 2010. Remote sensing of vegetation: Principles, techniques, and applications. Oxford University Press, Oxford.

Jung, B.J., Lee, H.J., Jeong, J.J., Owen, J., Kim, B., Meusburger, K., Alewell, C., Gebauer, G., Shope, C., Park, J.H., 2012. Storm pulses and varying sources of hydrologic carbon export from a mountainous watershed. J. Hydrol. 440-441, 90–101. doi:10.1016/j.jhydrol.2012.03.030

Kang, S., Tenhunen, J., 2010. Complex Terrain and Ecological Heterogeneity (TERRECO): Evaluating Ecosystem Services in Production Versus water Quantity/quality in Mountainous Landscapes. Korean J. Agric. For. Meteorol. 12, 307–316. doi:10.5532/KJAFM.2010.12.4.307

Kim, D., Zheng, Y., 2011. Scale-dependent predictability of DEM-based landform attributes for soil spatial variability in a coastal dune system. Geoderma 164, 181–194. doi:10.1016/j.geoderma.2011.06.002

Kim, I., Le, Q.B., Park, S.J., Tenhunen, J., Koellner, T., 2014a. Driving Forces in Archetypical Land-Use Changes in a Mountainous Watershed in East Asia. Land 3, 957–980. doi:10.3390/land3030957

Kim, I., Lee, K., Gruber, N., Karl, D.M., Bullister, J.L., Yang, S., Kim, T., 2014b. Increasing anthropogenic nitrogen in the North Pacific Ocean. Science 346, 1102–1106.

Kim, J., Grunwald, S., Rivero, R.G., 2014c. Soil Phosphorus and Nitrogen Predictions Across Spatial Escalating Scales in an Aquatic Ecosystem Using Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 52, 6724–6737.

Kim, T.-W., Lee, K., Najjar, R.G., Jeong, H.-D., Jeong, H.J., 2011. Increasing N Abundance in the Northwestern Pacific Ocean Due to Atmospheric Nitrogen Deposition. Science 334, 505–509. doi:10.1126/science.1206583

Korea Institute of Geology Mining and Material, 2001. Explanatory note of the Gangreung Sokcho sheet 1:250,000. Korea Institute of Geology, Mining and Material, Daejeon.

Korea meteorological administration, 2015. Korea weather survice. http://www.kma.go.kr/ (accessed 3.18.16).

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer, New York.

Kunkel, M.L., Flores, A.N., Smith, T.J., McNamara, J.P., Benner, S.G., 2011. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. Geoderma 165, 1–11. doi:10.1016/j.geoderma.2011.06.011

Lagacherie, P., McBratney, A.B., 2007. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping, in: Lagacherie, P., Mcbratney, A.B., Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective. Elsevier, Amsterdam, pp. 3–22.

Laliberté, E., Grace, J.B., Huston, M. a., Lambers, H., Teste, F.P., Turner, B.L., Wardle, D. a., 2013. How does pedogenesis drive plant diversity? Trends Ecol. Evol. 28, 331–340. doi:10.1016/j.tree.2013.02.008

Lee, G., 2004. Characteristics of Geomorphological Surface and Analysis of Deposits in Fluvial Terraces at Upper Reach of Soyang River. J. Korean Geogr. Soc. 39, 27–44.

Ließ, M., Glaser, B., Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and Random Forest models. Geoderma 170, 70–79. doi:10.1016/j.geoderma.2011.10.010

Ließ, M., Glaser, B., Huwe, B., 2011. Functional soil-landscape modelling to estimate slope stability in a steep Andean mountain forest region. Geomorphology 132, 287–299. doi:10.1016/j.geomorph.2011.05.015

MacMillan, R.A., Shary, P.A., 2009. Landforms and landform elements in geomorphometry, in: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 227–254.

Mage, S.M., Porder, S., 2013. Parent Material and Topography Determine Soil Phosphorus Status in the Luquillo Mountains of Puerto Rico. Ecosystems 16, 284–294. doi:10.1007/s10021-012-9612-5

Maynard, J.J., Johnson, M.G., 2014. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. Geoderma 230-231, 29–40. doi:10.1016/j.geoderma.2014.03.021

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping, Geoderma. doi:10.1016/S0016-7061(03)00223-4

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94. doi:10.1016/S0016-7061(98)00137-2

Minasny, B., Bishop, T.F.A., 2008. Analysing uncertainty, in: McKenzie, N., Grundy, M., Webster, R., Ringrose-Voase, A. (Eds.), Guidelines for Surveying Soil and Land Resources. CSIRO Publishing, Melbourne, pp. 383–393.

Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. Earth-Science Rev. 106, 52–62. doi:10.1016/j.earscirev.2011.01.005

Minasny, B., McBratney, A.B., 2015. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. doi:10.1016/j.geoderma.2015.07.017

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

Montello, D.R., 2001. Scale in geography, in: Smelser, N.J., Baltes, P.B. (Eds.), International Encyclopedia of the Social & Behavioral Sciences. Pergamon press, Oxford, pp. 13501–13504.

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Sci. Soc. Am. J. 57, 443–452.

Morford, S.L., Houlton, B.Z., Dahlgren, R.A., 2011. Increased forest ecosystem carbon and nitrogen storage from nitrogen rich bedrock. Nature 477, 78–81. doi:10.1038/nature10415

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1–19. doi:10.1016/j.geoderma.2010.12.018

National Academy of Agricultural Science, 2013. Korean Soil Information System. http://soil.rda.go.kr/soil/index.jsp (accessed 1.12.16).

Negassa, W., Leinweber, P., 2009. How does the hedley sequential phosphorus fractionation reflect impacts of land use and management on soil phosphorus: a review. J. Plant Nutr. Soil Sci. 172, 305–325. doi:10.1002/jpln.200800223

Osman, K.T., 2013. Forest Soils: Properties and Management. Springer International Publishing Switzerland.

Park, S.J., Rüecker, G.R., Agyare, W.A., Akramhanov, A., Kim, D., Vlek, P.L.G., 2009. Influence of Grid Cell Size and Flow Routing Algorithm on Soil-Landform Modeling. J. Korean Geogr. Soc. 44, 122–145.

Park, S.J., Van De Giesen, N., 2004. Soil-landscape delineation to define spatial sampling domains for hillslope hydrology. J. Hydrol. 295, 28–46. doi:10.1016/j.jhydrol.2004.02.022

Park, S.J., Vlek, P.L.G., 2002. Environmental correlation of three-dimensional soil spatial variability: A comparison of three adaptive techniques. Geoderma 109, 117–140. doi:10.1016/S0016-7061(02)00146-5

Phillips, J.D., 1999. Earth surface systems: Complexity, order, and scale. Blackwell publishers, Oxford.

Roberts, T.L., Stewart, J.W.B., Bettany, J.R., 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.

Rodionov, a., Flessa, H., Grabe, M., Kazansky, O. a., Shibistova, O., Guggenberger, G., 2007. Organic carbon and total nitrogen variability in permafrost-affected soils in a forest tundra ecotone. Eur. J. Soil Sci. 58, 1260–1272. doi:10.1111/j.1365-2389.2007.00919.x

Roger, A., Libohova, Z., Rossier, N., Joost, S., Maltas, A., Frossard, E., Sinaj, S., 2014. Spatial variability of soil phosphorus in the Fribourg canton, Switzerland. Geoderma 217-218, 26–36. doi:10.1016/j.geoderma.2013.11.001

Roman, L., Scatena, F.N., Bruijnzeel, L.A., 2010. Global and local varaitions in tropical montane cloud forest soils, in: Bruijnzeel, L., Scatena, F., Hamilton, L. (Eds.), Tropical Montane Cloud Forest: Science for Conservation and Management. Cambridge University Press, New York, pp. 77–89.

Roudier, P., Beaudette, D.E., Hewitt, A.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints, in: Minasny, B., Malone, B., McBratney, A. (Eds.), Digital Soil Assessments and beyond. CRC Press, Boca Raton, pp. 227–231.

Ryan, P.J., McKenzie, N.J., O'Connell, D., Loughhead, a. N., Leppert, P.M., Jacquier, D., Ashton, L., 2000. Integrating forest soils information across scales: Spatial prediction of soil properties under Australian forests. For. Ecol. Manage. 138, 139–157. doi:10.1016/S0378-1127(00)00393-5

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. 27, 171–197. doi:10.1191/0309133303pp366ra

Seibert, J., Stendahl, J., Sørensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma 141, 139–148. doi:10.1016/j.geoderma.2007.05.013

Smeck, N.E., 1985. Phosphorus dynamics in soils and landscapes. Geoderma 36, 185–199.

Smola, A., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. doi:Doi 10.1023/B:Stco.0000035301.49549.88

Soethe, N., Lehmann, J., Engels, C., 2008. Nutrient availability at different altitudes in a tropical montane forest in Ecuador. J. Trop. Ecol. 24, 397–406. doi:10.1017/S026646740800504X

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. doi:10.1016/j.ecolind.2007.05.005

Tesfa, T.K., Tarboton, D.G., Chandler, D.G., McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45, 1–16. doi:10.1029/2008WR007474

Tiessen, A.H., Chacon, P., Cuevas, E., 1994. Phosphorus and Nitrogen Status in Soils and Vegetation along a Toposequence of Dystrophic Rainforests on the Upper Rio Negro. Oecologia 99, 145–150.

Vasques, G.M., Grunwald, S., Myers, D.B., 2012. Influence of the spatial extent and resolution of input data on soil carbon models in Florida, USA. J. Geophys. Res. Biogeosciences 117, 1–12. doi:10.1029/2012JG001982

Vincent, A.G., Sundqvist, M.K., Wardle, D. a, Giesler, R., 2014. Bioavailable soil phosphorus decreases with increasing elevation in a subarctic tundra landscape. PLoS One 9, e92942. doi:10.1371/journal.pone.0092942

Vitousek, P., Chadwick, O., Matson, P., Allison, S., Derry, L., Kettley, L., Luers, A., Mecking, E., Monastra, V., Porder, S., 2003. Erosion and the Rejuvenation of Weathering-derived

Nutrient Supply in an Old Tropical Landscape. Ecosystems 6, 762–772. doi:10.1007/s10021-003-0199-8

Vitousek, P.M., Hättenschwiler, S., Olander, L., Allison, S., 2002. Nitrogen and Nature. Ambio A J. Hum. Environ. 31, 97–101. doi:10.1579/0044-7447-31.2.97

Vitousek, P.M., Porder, S., Houlton, B.Z., Chadwick, O. a, Applications, S.E., January, N., Applications, E., Houlton, Z., 2010. Terrestrial phosphorus limitation : mechanisms , implications , and nitrogen — phosphorus interactions. Ecol. Appl. 20, 5–15. doi:10.1890/08-0127.1

Walker, T.W., Syers, J.K., 1976. The fate of phosphorus during pedogenesis. Geoderma 15, 1–19. doi:10.1016/0016-7061(76)90066-5

Wilcke, W., Boy, J., Goller, R., Fleischbein, K., Valarezo, C., Zech, W., 2010. Effect of topography on soil fertility and water flow in an Ecuadorian lower montane forest, in: Bruijnzeel, L., Scatena, F., Hamilton, L. (Eds.), Tropical Montane Cloud Forests, Science for Conservation and Management. Cambridge University Press, New York, pp. 402–409.

Wilcke, W., Oelmann, Y., Schmitt, A., Valarezo, C., Zech, W., Homeier, J., 2008. Soil properties and tree growth along an altitudinal transect in Ecuadorian tropical montane forest. J. Plant Nutr. Soil Sci. 171, 220–230. doi:10.1002/jpln.200625210

Wilcke, W., Yasin, S., Abramowski, U., Valarezo, C., Zech, W., 2002. Nutrient storage and turnover in organic layers under tropical montane rain forest in Ecuador. Eur. J. Soil Sci. 53, 15–27. doi:10.1046/j.1365-2389.2002.00411.x

Wohl, E., 2010. Mountain rivers. American Geophysical Union, Washington, DC.

Wood, J., 2009. Geomorphometry in LandSerf, in: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 333–349.

Yang, X., Post, W.M., 2011. Phosphorus transformations as a function of pedogenesis: A synthesis of soil phosphorus data using Hedley fractionation method. Biogeosciences 8, 2907–2916. doi:10.5194/bg-8-2907-2011

Zhang, X., Drake, N.A., Wainwright, J., 2013. Spatial Modelling and Scaling Issues, in: Wainwright, J., Mulligan, M. (Eds.), Environmental Modelling: Finding Simplicity in Complexity. John Wiley & Sons Ltd, West Sussex, pp. 69–90.

Zhou, J., Wu, Y., Bing, H., Yang, Z., Wang, J., Sun, H., Sun, S., Luo, J., 2016. Variations in soil phosphorus biogeochemistry across six vegetation types along an altitudinal gradient in SW China. Catena 142, 102–111. doi:10.1016/j.catena.2016.03.004

Chapter 2 Spatial soil nutrients prediction using three supervised learning methods for assessment of land potentials in complex terrain

Gwanyong Jeong[1], Hannes Oeverdieck[1], Soo Jin Park[2], Bernd Huwe[1], Mareike Ließ[1]

[1]*Department of Geosciences/ Soil Physics Division, University of Bayreuth, Universitaetsstrasse 30, 95447 Bayreuth, Germany*

[2]*Department of Geography, Seoul National University, Shilim-Dong, San 56-1, Kwanak-Gu, Seoul 151-742, South Korea*

Corresponding author: gwanyong.jeong@uni-bayreuth.de

Abstract

Mountain soils play an essential role in ecosystem management. Assessment of land potentials can provide detailed spatial information particularly concerning nutrient availability. Spatial distributions of topsoil carbon, nitrogen and available phosphorus in mountain regions were identified using supervised learning methods, and a functional landscape analysis was performed in order to determine the spatial soil fertility pattern for the Soyang Lake watershed in South Korea. Specific research purposes were (1) to identify important predictors; (2) to develop digital soil maps; (3) to assess land potentials using digital soil maps.

Soil profiles and samples were collected by conditioned Latin Hypercube Sampling considering operational field constraints such as accessibility and no go areas contaminated by landmines as well as budget limitations. Terrain parameters and different vegetation indices were derived for the covariates. We compared a generalized additive model (GAM) to random forest (RF) and support vector regression (SVR). For the predictor selection, we used the recursive feature elimination (RFE). A land potential assessment for soil nutrients was conducted using trimmed k-mean cluster analysis.

Results suggested that vegetation indices have powerful abilities to predict soil nutrients. Using selected predictors via RFE improved prediction results. RF showed the best performance. Cluster analysis identified four land potential classes: fertile, medium and low fertile with an additional class dominated by high phosphorus and low carbon and nitrogen contents due to human impact. This study provides an effective approach to map land potentials for mountain ecosystem management.

Keywords: mountain soil, soil nutrients, digital soil mapping, land potential assessment

2.1 Introduction

Recent research has shown an increased interest in mountain soils and their essential role in ecological functions (Ballabio, 2009; Roman et al., 2010; Wilcke et al., 2013). Mountain soils are regarded as a key factor for conservation and sustainable management in mountain regions. They have effects on downstream water quality and quantity, and nutrient dynamics as an input source (Wohl, 2010). Mountain soils and particularly their organic layers contain large stocks of soil nutrients released by mineralization of the organic matter (Wilcke et al., 2010). Especially, nitrogen and phosphorus deficiencies exert influences on limitation of plant growth in mountain areas. It is an important issue to understand nutrient supply in mountain areas which influences plant productivity, diversity and compositions (Benner et al., 2010). Last but not least, spatial information on soil nutrients is required to understand and manage mountain ecosystems.

Properties of mountain soils have a high local variability (Holtmeier, 2009). This is because mountain areas underlie a high variability in climatic conditions, topography, parent material and vegetation (Funnell and Parish, 2005). Several processes such as creeping, erosion, solifluction, and landslide lead to nutrient losses in some parts and accumulations in others.

Organic matter accumulation, nutrient cycling and disturbance are enhanced by vegetation and animals (Brady and Weil, 2010). Furthermore, parent rock material influences the spatial distribution of base status and nutrient contents in the soil (Binkley and Fisher, 2012). Complex interactions among these various environmental factors generate a considerable spatial heterogeneity of physical and chemical soil properties. Mountain areas also provide various challenges to any soil mapping approach as they are difficult to access and have scarce data availability. Construction of soil dataset in mountain areas frequently requires considerable amounts of cost, time, and labor due to these reasons. Digital soil mapping can be a useful tool to reduce this kind of efforts as well as obtain reasonable results (Ballabio, 2009; Ließ et al., 2012).

Soil fertility refers to the soil's ability to provide nutrients in available forms and appropriate amounts for plant growth and reproduction (Osman, 2013). Soil fertility plays a key role in land potential or capacity (Osman, 2013; Stockdale et al., 2013). In mountain areas, comprehensive information on the soil fertility is needed. Some researches on land potentials performed spatial predictions based on digital soil mapping but most studies focused on agricultural areas (Al-Shamiri and Ziadat, 2012; Harms et al., 2015; Sun et al., 2012). Moreover, there were a few researches on spatial predictions for soil fertility in mountain areas. In this study, we tried to develop soil nutrient maps in order to investigate land potentials in the Soyang watershed, South Korea. Specific research purposes were (1) to identify the most important environmental predictors to predict soil nutrients, (2) to develop digital soil maps with uncertainty and (3) to assess land potentials.

2.2 Materials and methods

The whole procedure in this study is outlined in Figure 2.1. Our methods consisted of following three steps; predictor selection, model tuning and validation, and functional analysis.

2.2.1 Research area

The Soyang lake watershed is located in the north-eastern part of Gangwon-do province, South Korea. It extends between 70 and 1700 m a.s.l. and covers an area of 2,776 km². The Soyang lake was impounded by an artificial dam in 1973, located about 10 km northeast of Chuncheon. The dam was constructed for flood control, water supply and hydroelectric power generation for downstream areas. Being close to an even crossing the border to North Korea, parts of the corresponding watershed are not accessible due to military facilities and
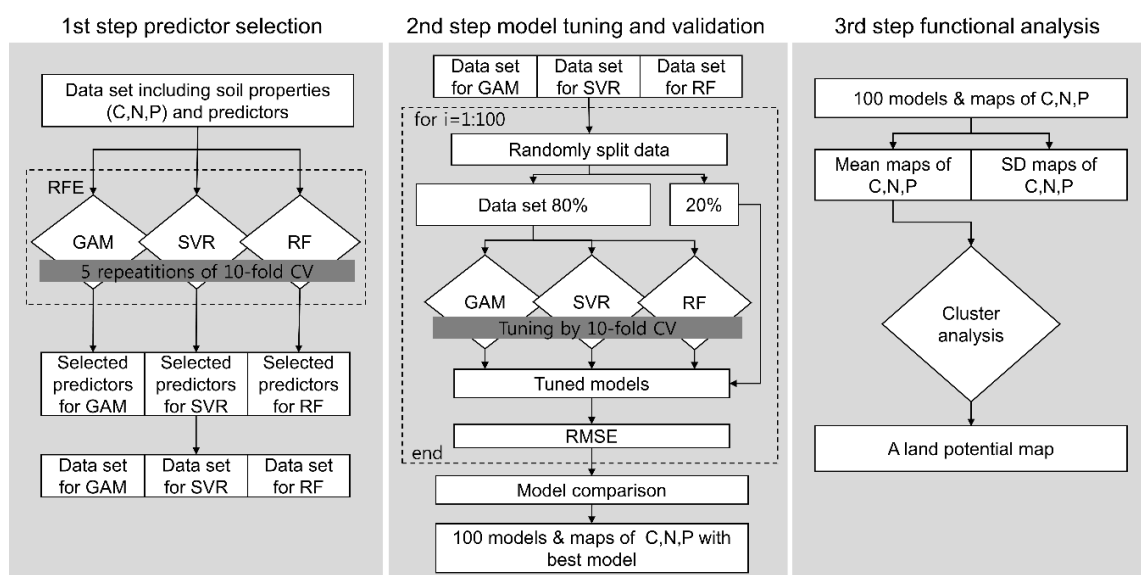
Figure 2.1 Flowchart of the proposed procedure. C = carbon, N = nitrogen, P = available phosphorus, RFE = recursive feature elimination, GAM = generalized additive model, SVR = support vector regression, RF = random forest, CV = cross validation, RMSE = root mean squared error, SD = standard deviation.

contamination by landmines. The Soyang river originates from Mu, Seorak and Obdae Mountain (Figure 2.2). The whole geomorphological feature is an incised meander that has a narrow river valley due to tectonic process during the Quaternary (Lee, 2004). The area's geology is dominated by banded gneiss and granite (Korea Institute of Geology Mining and Material, 2001). Chuncheon and Haean consist of granite and show plain areas. Moderately coarse textured soil and clay loam soil cover around 60% of the area (National Academy of Agricultural Science, 2013). The area receives 1,179 mm of mean annual rain fall. It is mostly covered by forest (80%) with deciduous trees (51.8%), coniferous (25.4%), and mixed (22.8%) trees as dominant forest types and includes two national parks (Seorak and Odae).
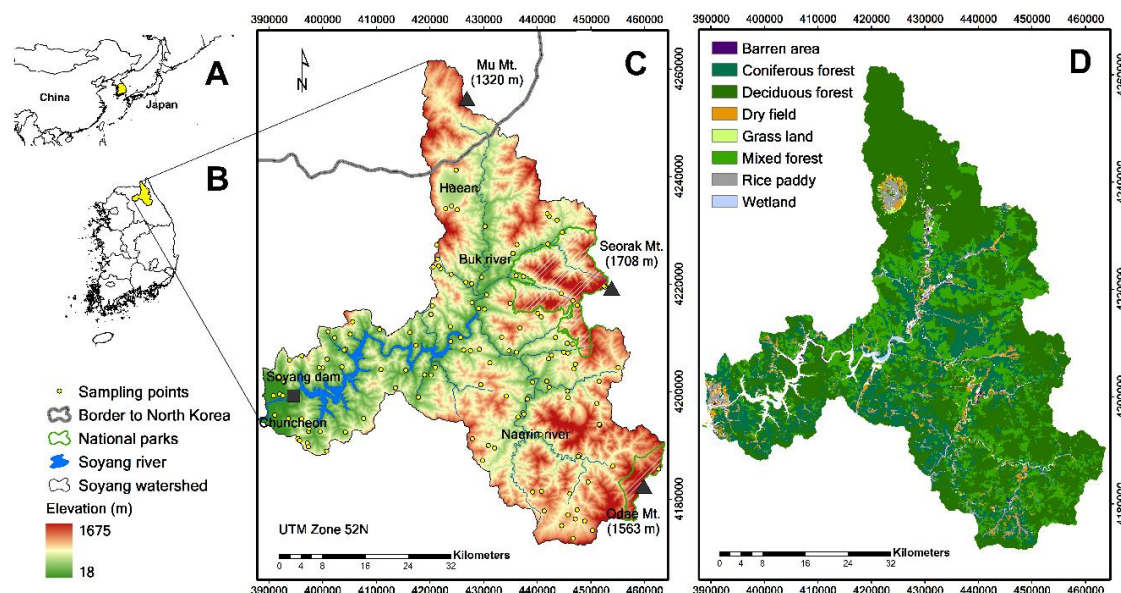
Figure 2.2 Research area. (A) The map of the Korean Peninsula. (B) The Soyang watershed is located in the north-eastern part of South Korea. (C) The map shows the spatial distributions of sampling points. The Soyang watershed is near the border to North Korea and includes two national parks (Seorak and Odae). The Soyang river originates from Mu, Seorak and Obdae Mountain. (D) The landuse map of Soyang watershed from Korean Environmental Geographic Information Service (EGIS) (http://egis.me.go.kr)

2.2.2 Soil dataset and environmental predictors

About 139 samples were collected in the research area in 2013 (Figure 2.2). To obtain a representative dataset for the Soyang Lake watershed, conditioned Latin Hypercube Sampling (cLHS) was applied to guarantee for optimal coverage of the variability of environmental covariates (Minasny and McBratney, 2006). Impractical sampling designs are common in cLHS due to difficult accessibility and sparely distributed locations within the study area (Roudier et al., 2012). The sampling method for this research, therefore, considered operational field constraints such as accessibility and no go areas contaminated by landmines as well as budget limitations. cLHS was performed within R package "clhs" (Minasny and McBratney, 2006).

Available phosphorus (P) of the topsoil horizon was extracted by the Lancaster method (pH 4.25), a P extractant for a wide range of soils (Cox, 2001; Kim et al., 2013). Alternatives are the Bray 1 method developed on acid soils and the Olsen method, originally proposed for alkaline soils but now also used for acid and neutral soils (Sims and McGrath, 2012). The P content in the extractant was then measured by inductively coupled plasma optical emission spectroscopy (Cox, 2001). The topsoil total carbon (C) and nitrogen (N) contents were determined by CNS-Analyser with conductivity detectors by high temperature combustion.

Various vegetation parameters derived from satellite images were used as predictors in addition to terrain parameters derived from a DEM of 30 m ASTER data from the U.S. Geological Survey (USGS) website (http://earthexplorer.usgs.gov/) (Table 2.1). They were calculated with the terrain analysis modules of the open source software SAGA (SAGA User Group Association, 2011). Different vegetation indices were extracted from a 30 m landsat TM image on 25th May 2009 from USGS website (http://glovis.usgs.gov/).

Table 2.1 Environmental predictors for digital soil mapping

	Predictor	Method	Reference
1	Elevation (ELEV)	-	-
2	Slope degree (SLO)	Slope, aspect, curvature saga module	(Zevenbergen et al., 1987)
3	Strahler order ≥5 Overland flow distance to channel network (OFD)	Overland low distance to channel network saga module	(Strahler, 1957)
4	Vertical overland flow distance (vOFD)	Overland low distance to channel network saga module	(Freeman, 1991)
5	Horizontal overland flow distance (hOFD)	Overland low distance to channel network saga module	(Freeman, 1991)
6	Valley depth (VD)	Valley depth saga module	-
7	Catchment area (CA)	Catchment area (Flow tracing) saga module	(Costa-Cabral and Burges, 1994)
8	Convergence index (CONV)	Convergence index saga module	(Koethe and Lehmeier, 1996)
9	Positive and negative openness (POPEN, NOPEN)	Topographic openness saga module	(Yokoyama et al., 2002)
10	Normalized difference vegetation index (NDVI)	(NIR – Red)/ (NIR+Red)	(Tucker and Sellers, 1986)
11	Green normalized difference vegetation index (GNDVI)	(NIR – Green)/ (NIR+Green)	(Gitelson et al., 1996)
12	Normalized difference water index (NDWI)	(NIR – SWIR)/ (NIR+SWIR)	(Gao, 1996)

NIR= near-infrared, SWIR= shortwave-infrared

2.2.3 Supervised learning methods

Three supervised learning methods, generalized additive model, random forest, and support vector regression were compared. These methods have the common ability to model complex nonlinear relationships between soil properties and environmental predictors. They showed good performances in mountain areas (Ballabio, 2009; Ließ et al., 2012; Tesfa et al., 2009).

1.1.1.1 Generalized additive model

The generalized additive model (GAM) is an extension of the generalized linear model (Hastie and Tibshirani, 1990). It uses a smooth function which allows identifying nonlinear relationships between the response and predictor variables. The GAM used for this study was penalized regression splines, which adds a penalty term for "overfitting" to the goodness-of-fit term to minimize the sum of squared residuals (Wood, 2006).

The GAM tries to find the optimum balance between model fit and model smoothness. This is controlled by the smoothing parameter, λ. With λ being too small, data points are fitted as good as possible resulting in a "wiggly" model. With λ being too large, the function f(x) is smoothed into a straight line. The applied R package "mgcv" uses cross-validation to determine λ (Wood, 2006).

2.2.3.1 Support vector regression

Support vector regression (SVR) is a powerful data mining technique that can be generalized to non-linear models using a kernel function (Vapnik, 1995). The kernel function is used to transform the input predictor space into a higher dimensional space. ε-SVR was used for this research (Smola and Schölkopf, 2004). For regression, SVM finds the best fit in the feature space but uses a loss function to find the regression line that minimizes model errors or residuals. The loss function in ε-SVR uses not squared but absolute residuals over the

threshold ε while the loss function ignores all errors within ± ε. This approach shows robust effects that are not sensitive to outliers on the prediction (Cherkassky and Mulier, 2007). With ε, the loss function can add a penalty (cost) for large residuals. There is a relationship between the ε and the cost parameter (Kuhn and Johnson, 2013). At the initial step, 0.1 (default value) was chosen for ε, and kept fixed while tuning the value of cost via 10-fold cross validation. We used the radial basis kernel (Equation 2.1), a general kernel function shown to be very effective. Its parameters are easier to tune than those of other kernels (Ballabio, 2009; Kuhn and Johnson, 2013).

$$\mathrm{k}(x_i, x_j) \ = \ \exp(-\sigma \parallel x_i - x_j \parallel^2), \qquad (2.1)$$

Where k is the kernel function, σ represents a scaling parameter, x is the input vectors, and ∥ ∥ is the Euclidean norm.

The radial basis function requires to tune the sigma (σ) parameter that controls its width. In the applied R package "caret" (Kuhn and Johnson, 2013), σ is automatically tuned. The SVR was performed within R package "kernlab" (Karatzoglou et al., 2004).

2.2.3.2 Random forest

Random forest (RF) is an ensemble learning method which operates by building a set of regression trees and averaging the results (Breiman, 2001). Each regression tree is constructed based on bootstrap samples of the data. A random subset of predictors of a predefined size (mtry) is used to fit each tree. Each tree is grown until the specified minimum node size (nodesize) is reached. RF uses some data (default: 1/3 of all data) as a test subset to calculate the prediction error (out of bag error). As long as the number of trees (ntree) is large, this out of bag error always converges (Breiman, 2001). RF guarantees model stability that means low bias and low variance via corrections of overfitting habit for each tree (Hastie

et al., 2009). According to Kuhn and Johnson (2013) and Strobl et al. (2009), RF does not need much tuning. The number of trees (ntree) was set to 1000. The default value five was used as minimum node size. The value of mtry was tuned using the train command in the R package "caret" (Kuhn and Johnson, 2013). RF was performed within R package "randomForest" (Breiman, 2001).

2.2.4 Recursive feature elimination

Predictor selection is a critical step in supervised learning methods, because this approach can remove uninformative and noisy predictors from the dataset and save computing time. This predictor selection, usually known as feature selection, can be placed into two categories which are filter and wrapper methods (Guyon and Elisseeff, 2003). The filter methods are a preprocessing step that evaluates the relevance of the predictors without models. This approach has some drawbacks: e.g. undesirable inputs are filtered out of the data, the selection criterion is not related to the effectiveness of the model, and important interactions between predictors cannot be quantified (Javaheri et al., 2014). The wrapper methods conduct a search in the target model itself to find an optimal combination of predictors considering the predictors' interactions (John et al., 1994). We used recursive feature elimination (RFE) which is a popular wrapper method.

RFE is a backward selection algorithm which iteratively eliminates the least important predictors from the model based on an initial predictor importance measure (Kuhn and Johnson, 2013). For RF, the random forest variable importance is used. For SVR and GAM, the predictor importance is evaluated using the locally weighted regression model (LOESS) smoother for modelling nonlinear relationships (Kuhn and Johnson, 2013). Predictor importance measures based on the R^2. Firstly, the full model is created using all predictors and a measure of predictor importance is computed to rank the predictors from most important to least. Secondly, the models are tuned and trained iteratively with the most important

predictors and without least important predictors until only one predictor remains. The RFE incorporating resampling (5 repetitions of 10-fold cross validation) was used for this. At the end, the appropriate number of predictors and the final list of selected predictors are determined. The predictors are returned in the order of the most important to the least important. The package "caret" provides the functions for RFE (Kuhn and Johnson, 2013).

2.2.5 Model validation

To compare model performance, the root mean squared error (RMSE) and pearson's correlation coefficient (r) were calculated. Soil samples were randomly divided into two datasets. The first dataset (80%) was used to tune the parameters of the models by 10-fold cross validation. The second dataset (20%) was used to estimate model performances. This procedure was repeated 100 times. At the end, 100 tuned models and RMSEs were returned for each model algorithm and each soil property.

2.2.6 Land potential assessment

Trimmed k-mean cluster analysis was applied to conduct a land potential evaluation. The soil nutrient maps for C, N, and P were used as input data. The k-mean algorithm splits the observation dataset into distinct groups (Hastie et al., 2009). It minimizes the within-class variance (homogeneous) and maximises the variance between classes (heterogeneous). Like other statistical methods, clustering methods might be influenced by few extreme values (Fritz, 2012). For robustification, the trimmed k means technique discards a proportion of the most distant observations from the centroid of each cluster (Farcomeni and Greco, 2015). The trimmed k means method requires two parameters: the number of clusters (k) and the proportion of observations with extreme values (α). Solutions with 2 to 10 clusters and α values from 0 to 20% were tested, and the best solution was selected based on the calculated

likelihood values that assess probabilities of cluster membership for the dataset. The trimmed k-means cluster analysis was performed within R package "tclust" (Fritz, 2012).

Table 2.2 Statistical summary of the collected soil dataset

	Mean	SD	MIN	Median	Max	CoV (%)	Skew	Kurt
C (mg/g)	32.78	20.25	1.10	29.35	97.60	61.78	0.73	0.27
N (mg/g)	2.29	1.33	0.30	2.02	7.00	58.08	1.02	0.82
P (mg/kg)	72.71	156.27	1.01	25.36	849.70	214.92	3.42	10.96

SD = standard deviation, MIN= minimum, MAX= maximum, CoV= coefficient of variation, Skew= Skewness, Kurt= Kurtosis.

2.3 Results and discussion

2.3.1 Soil nutrients

Table 2.2 shows descriptive statistics of the collected soil dataset. Mean topsoil C content was 33 ± 20 mg/g and mean topsoil N content was 2 ± 1 mg/g. Both C and N didn't show a strong skewness (Figure 2.3). Topsoil available P content ranged from 1.01 to 849.70 mg/ kg with a mean concentration of 72.65 ± 156.30 mg/kg. Soil P data exhibited a strong positive skewness (Figure 2.3). Soil P showed a coefficient of variation (CoV) of 214.94 % and high variance.
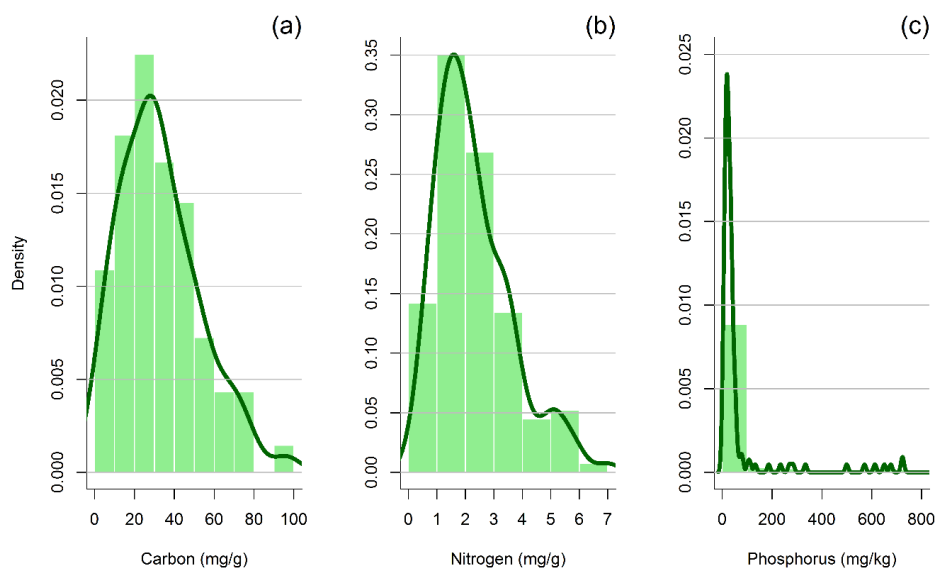
Figure 2.3 Histograms and probability functions of C, N and P for all data (a-c).

Table 2.3 Predictors selected by recursive feature elimination

	Carbon	Nitrogen	Phosphorus
GAM	NDVI, ELEV, CONV, vOFD	NDVI, CONV, GNDVI, NOPEN, vOFD	NDWI, CA, VD, OFD, CONV, hOFD, ELEV, NDVI, NOPEN, GNDVI, POPEN, vOFD
RF	NDVI, ELEV, GNDVI, vOFD, CA, OFD, NDWI, hOFD, NOPEN, VD, POPEN	NDVI, GNDVI, ELEV, VD, OFD, hOFD, vOFD, CA, POPEN, NDWI, NOPEN	NDWI, NDVI, GNDVI, NOPEN, VD, vOFD, OFD, hOFD, ELEV
SVR	NDVI, GNDVI, ELEV, POPEN, CA, VO, NDWI	POPEN, ELEV, NDVI, VD, GNDVI, vOFD, CA, CONV, NDWI, OFD, hOFD, NOPEN, SLO	NDWI, NDVI

2.3.2 Recursive feature elimination

The RFE reduced the number of predictors to those listed in Table 2.3. Using selected predictors via RFE improved the prediction results. The selected predictors were different for each soil nutrient and each supervised method (Table 2.3). This result is a commonly observed phenomenon in digital soil mapping (Brungard et al., 2015; Miller et al., 2015; Poggio et al., 2013).

NDVI had the highest overall predictor importance for C in all three algorithms. Among the terrain parameters, elevation and some hydrological parameters (vertical overland flow distance or catchment area) were selected by each of the models. The latter corresponds to accumulations of water, nutrients, and sediments from upslope (Gruber and Peckham, 2009). The convergence index is an important predictor in GAM while GNDVI is an important predictor in RF and SVR (Table 2.3).

The NDVI had the highest overall predictor importance in GAM and RF models for N but POPEN was selected in SVR model as the best predictor. Like C, similar predictors such as NDVI, GNDVI, and elevation were selected. Hydrological predictors were important for N.

The selected predictors for P prediction were distinguished with those for C and N. Here, the NDWI (instead of NDVI) had the highest overall predictor importance. For all three models, vegetation indices (NDWI and NDVI) were selected among the predictors. For GAM, catchment area, valley depth, and overland flow distance were important predictors. Results suggested the spectral predictors derived from the remote sensing image have powerful abilities to predict soil nutrients.

2.3.3 Model comparison

Figure 2.4 shows the boxplots of the RMSE and r distributions (prediction error) from the 100 models for each of the three nutrients and model algorithms. RF showed the best performance considering the median RMSE of the 100 model runs (Table 2.4). For N and P, RF also showed the lower variability of the RMSE compared to GAM and SVR. SVR and RF had the similar result for C and N content. GAM showed rather worse results based on RMSE. GAM also showed larger uncertainty concerning the variability of RMSE. Therefore, it is not easy to generalize underlying relationships between soil properties and environmental predictors using GAM. RF and SVR were better models than GAM in our results.

Tesfa et al. (2009) reported that the performances of RF were slightly better than that of GAM for the prediction of soil depth. Brungard et al. (2015) reported that RF showed a better result to predict soil classes compared to SVM. Similar results were reported by Kampichler et al.

(2010) and Hsieh et al. (2011) in ecological and medical applications. However, some studies showed contrasting results. Li et al. (2015) tested GAM, RF, and SVM for fish stock assessment and found that SVM slightly outperformed RF. Similar to our results, the performance of GAM was worse than that of RF and SVM. Compared to SVR, RF is the easier procedure, because SVR requires the tuning of more parameters. Kampichler et al. (2010) compared five machine learning methods based on modelling performance, modelling effort, classifier comprehensibility, and method intricacy. They recommended the use of RF because of its modelling performance as well as the ease of model construction and interpretability. Our results also indicated that RF can produce good model results for soil nutrient prediction.
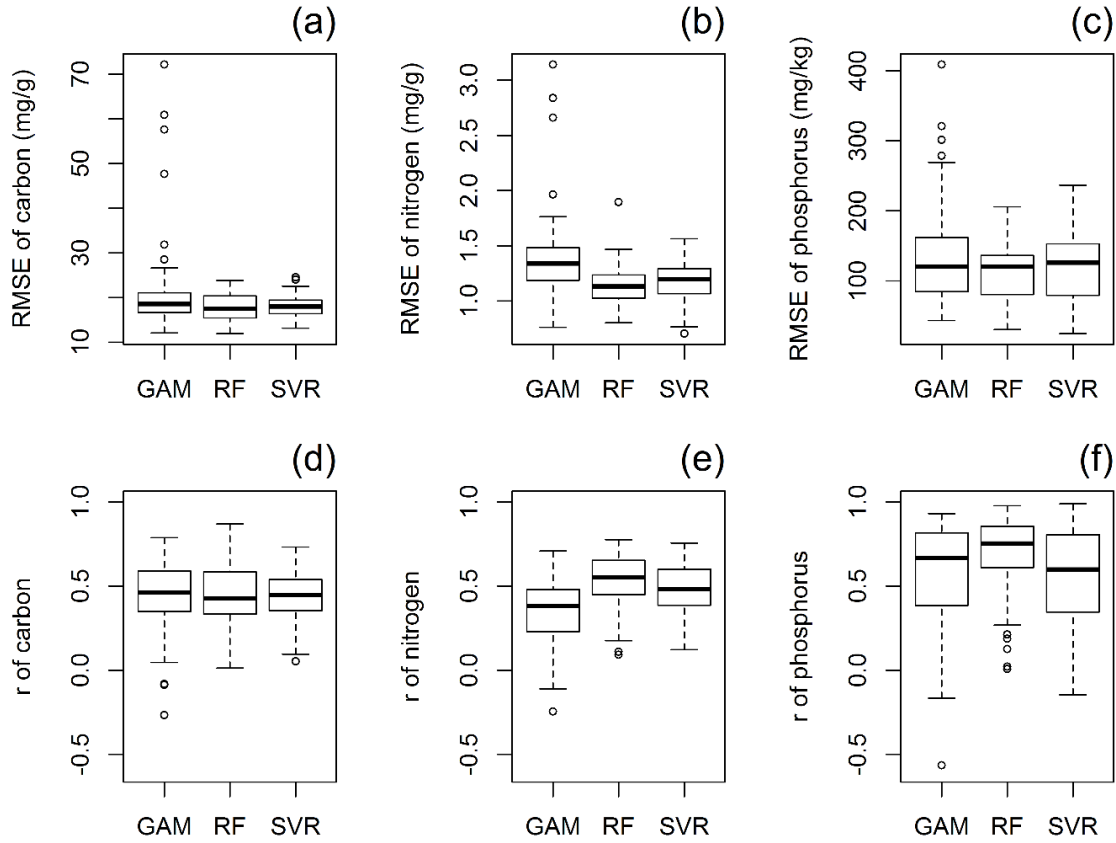
Figure 2.4 (a) – (c) RMSE of the 100 models, (d) – (f) Pearson's correlation coefficients (r) of the 100 models.

Table 2.4 Statistical Summary of supervised model performances

Model	Model estimate	Carbon	Nitrogen	Phosphorus
SVR	Median r	0.493	0.521	0.598
	Median RMSE	17.616	1.141	126.0
RF	Median r	0.427	0.552	0.754
	Median RMSE	17.544	1.131	120.4
GAM	Median r	0.461	0.382	0.667
	Median RMSE	18.588	1.338	120.6

r= Pearson's correlation coefficient

Finally, we selected the best model algorithm based on the median RMSE to predict the spatial distribution of each soil nutrient: RF for the prediction of all three soil nutrients.

2.3.4 Spatial prediction

The procedure's results are two maps for each of the three nutrients which display the mean and standard deviation of the 100 predictions. Figure 2.5 refers to the C content in the surface horizon. Figure 2.6 depicts the spatial N distribution, while Figure 2.7 shows the P content. The spatial patterns of C and N were similar, but the available P pattern was individual. Kovačević et al. (2010) reported the strong positive correlation between C and N (r=0.98) but the relatively weak relationship between C and available P (r=0.37). Similar findings are also found by other studies (Camargo et al., 2012; J. Kim et al., 2014; Kunkel et al., 2011; Sumfleth and Duttmann, 2008).

The vegetation indices made strong contributions to explain the spatial patterns of soil nutrients as noted above. Higher values of C and N contents can be found at higher altitudes (Figure 2.5 and 2.6) where vegetation density (NDVI) is high. The NDVI is related to biomass, leaf area index (LAI), and the fraction absorbed photosynthetically active radiation (FPAR) which indicate biophysical measures to model soil nutrients (Grunwald et al., 2015; Mulder et al.,

2011). The strong relationship between NDVI and soil nutrients was also reported by other studies for C and N (Kunkel et al., 2011; Sumfleth and Duttmann, 2008) and P and N (J. Kim et al., 2014). Sumfleth and Duttamann (2008) reported C (r=0.55) and N (r=0.52) contents to be significantly related to NDVI in an agricultural landscape. Kunkel et al. (2011) accounted for 54% of C stock variance and 37% of N stock variance using only NDVI in an area vegetated with grass and forest. As C and N contents are dependent on input rates from litterfall, C and N contents have significant correlations with NDVI.

Elevation might be an indicator of soil temperature and decomposition rates which result in C and N accumulations (Binkley and Fisher, 2012; de Brogniez et al., 2015; Osman, 2013). Positive relationships between C and N and elevation were reported by Kunkel et al. (2011) and Peng et al. (2013). In the Soyang watershed, the areas at higher elevation are dominated
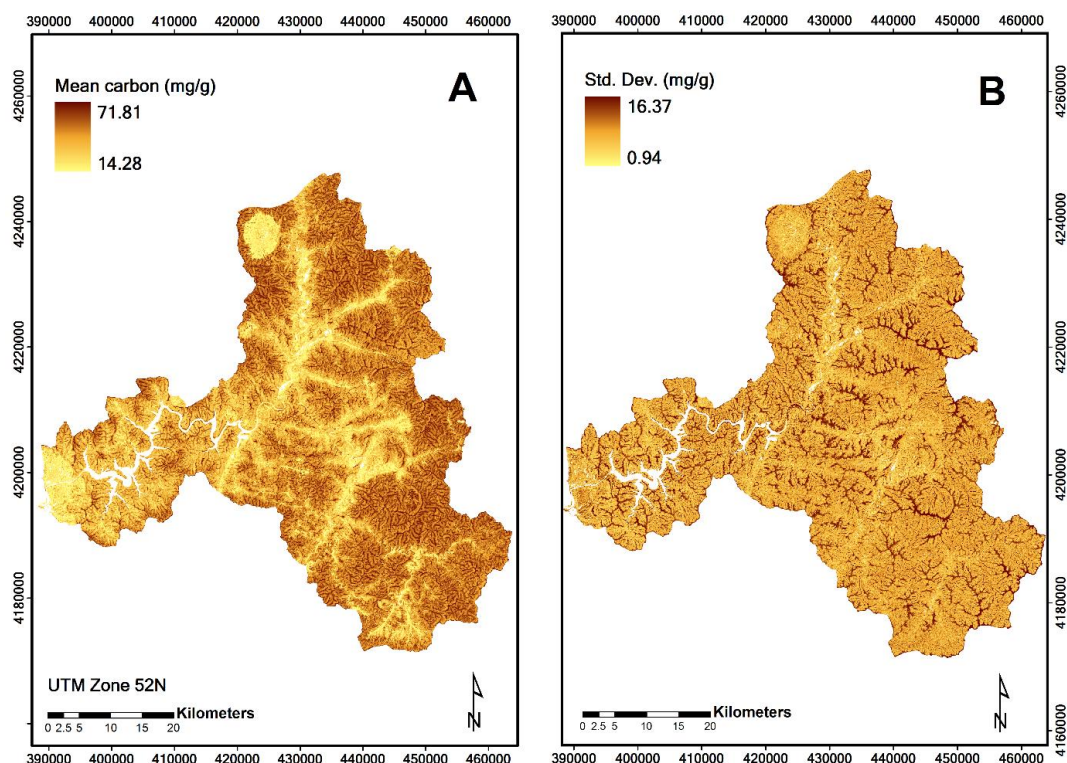
Figure 2.5 Carbon content map with random forest
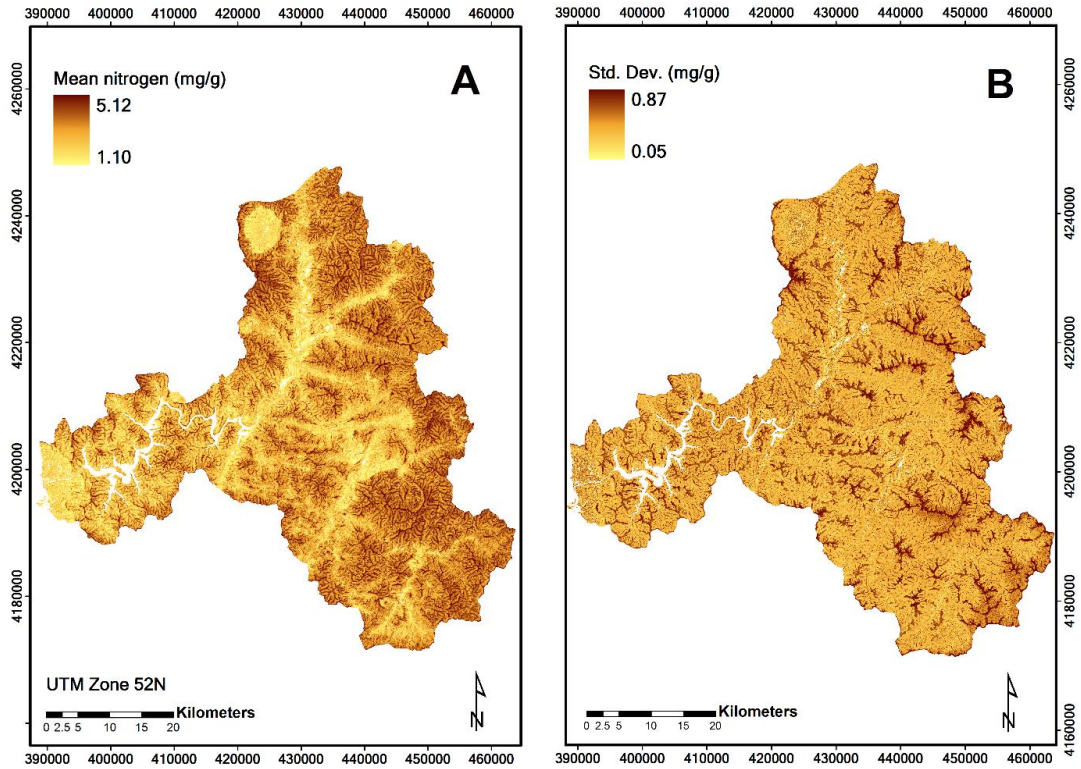
Figure 2.6 Nitrogen content map with random forest
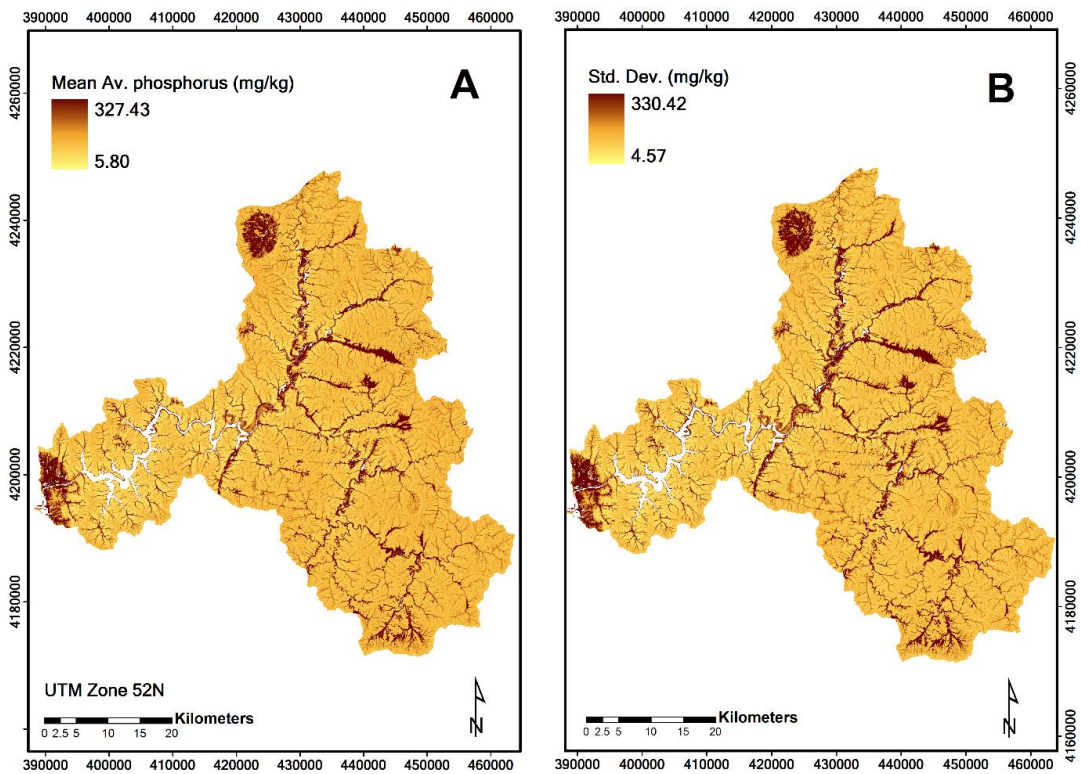
Figure 2.7 Available phosphorus content map with random forest

by forest lands which were partly designated natural conservation areas, national parks or military areas which might indicate a low level of disturbance. The lower areas are characterized by mountain farmlands and paddy areas (Kim et al., 2014). Considering the long history of landuse in the areas at lower elevation, the human impact led to a change in the spatial patterns of the soil nutrients (Figure 2.2). These land uses in the mountainous areas cause severe soil erosion, resulting in the loss of soil nutrients (Arnhold et al., 2013). The larger contents of C and N in the forest compared to the agricultural topsoils are consistent with the results found by Ruidisch et al. (2013) in the same watershed. This result is well-known but these patterns can't easily be detected by simple linear regression because of various land use types in the lower areas and the complexity of soil forming factors in high relief areas.

Among the areas at higher altitudes, the areas with the highest C content were close to the valley bottom. Lateral movement of water, sediment and nutrients is controlled by the slope configuration and the downslope hydraulic gradient which in turn result in this spatial differentiation of soil nutrients (Huggett, 1975; Moore et al., 1993; Park and Burt, 2002). Areas dominated by forest with higher vegetation density almost coincided with locations where the largest N contents were calculated (Figure 2.6). High N contents were predicted for the concave areas and for the eastern parts of the research area in the high mountain area. Usually, C and N are correlated. However, in our results, the spatial distribution of the N contents was strongly related to the channel networks (predictors OFD, vOFD and hOFD) (Figure 2.6). In the western part of the research area composed of coniferous forest and the plain agricultural area of Chuncheon (Figure 2.2), the C and N contents were generally lower than in other areas.

Areas with the highest contents of topsoil available P can be clearly related to locations near the lower plain area measured by high catchment area, high positive and low negative openness, and low vegetation indices in the western parts of the study area and related to rice

paddy and dry field areas (Figure 2.7). These areas with the larger amounts of P include Chuncheon and Haean (Figure 2.2). Historical land use changes natural processes and has effects on the soil P distribution (Roger et al., 2014). Similar to C and N, the spatial distribution of the P content corresponded to the land use patterns. However, in contrast to C and N, the hilltops and the shoulder areas showed lower P contents. Within the forest area, the P content was highest near the valley bottom and concave areas.

There were a few researches about spatial predictions of available P. Camargo et al. (2012) identified hillslope curvatures were useful to understand spatial available P patterns (Figure 2.7). In contrast to this, convergence index related to the curvature was not selected in our RF P model (Table 2.3). Moore et al. (1993) reported a similar spatial distribution of available P with slope, stream power index and wetness index as statistically significant predictors. In our research, slope was not selected but the model included hydrological predictors similar to those selected by Moore et al. (1993) (Table 2.3). McKenzie and Ryan (1999) found higher total P to be associated with lower elevation which represents a nutrient and material accumulation zone. Most of P is attached in the sediment as relatively insoluble substances and is accumulated in lower elevation or concave areas because P tends to react more readily to positively charged calcium, iron, and aluminium ions in soils (Bauer and Velde, 2014).

Figure 2.5b, 2.6b, and 2.7b show the standard deviation of C, N and P contents. Higher standard deviation of the C content was found in the coniferous forest land near the Soyang dam (Figure 2.5b) while high uncertainty for the N content was shown along the high ridges (Figure 2.6b). The spatial prediction uncertainty for available P was higher in lower and agricultural areas than in other areas (Figure 2.7b). The uncertainty for all soil nutrients was lower in mid-slope areas. Especially, higher uncertainty for P showed in areas with higher values of contents because of soil data structure. The soil data with a small proportion of extreme values shows skewed so that the data sparsity creates uncertainty (Elith and Leathwick, 2009).

2.3.5 Land potential assessment

Trimmed likelihood curves were used to choose k and α (Farcomeni and Greco, 2015). Figure 2.8 shows a high increase of the likelihood value between k=2 and K=3, with still some noticeable difference between k=3 and k=4, while no substantial increase in the likelihood between K=4 and K=5 is found. For α, we considered the degree of the likelihood increase. Considering Figure 2.8, we decided to use k=4 and α = 0.05. Class zero of Figure 2.9 included all extreme values which were trimmed from the data. It basically includes areas of agricultural land use which - due to fertilizer application - show a very different soil nutrient status, mainly indicated by the high P contents (Table 2.5). Class one had low C and N contents and high P contents. The areas also indicate agricultural land use in Haean and Chuncheon (Figure 2.2). This class occurs on lower slope and foot slope landform positions. Although this class is small areas, these areas are strongly affected by human impacts. Class two had high C and N contents. It occurs on concave mid slope positions and on the crest of high ridges vegetated with deciduous or mixed forest. Class two indicates a relatively fertile land. If these areas will
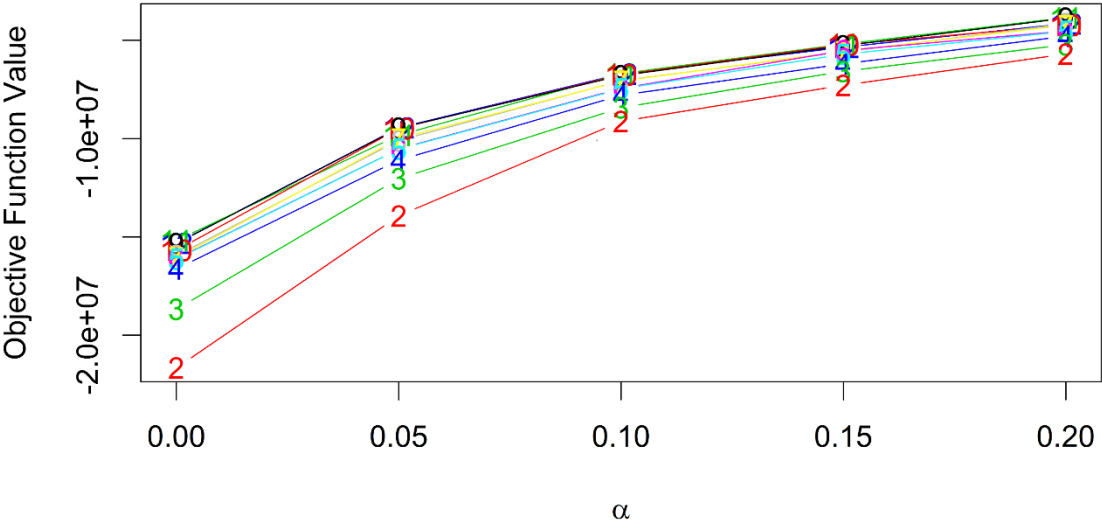
Figure 2.8 Classification trimmed likelihoods (CTL) for the land potential assessment. Numbers on the lines indicate the number of clusters for the particular setting

Table 2.5 The descriptive summary of each cluster

Cluster	Carbon (mg/g)			Nitrogen (mg/g)			Phosphorus (mg/kg)			Area (%)
	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	
0	14.28	25.07	63.95	1.12	1.95	4.55	43.60	171.8	327.43	5.00
1	16.01	25.95	52.54	1.17	1.95	4.19	29.82	49.28	76.06	4.54
2	34.75	49.18	71.81	2.06	3.46	5.12	10.16	18.95	45.48	19.00
3	33.98	40.47	48.58	1.92	2.75	3.86	7.60	13.32	31.89	44.46
4	15.38	30.00	35.58	1.10	2.05	3.21	5.80	12.62	31.60	27.01
Total	14.28	37.87	71.81	1.10	2.62	5.12	5.80	23.75	327.43	100

be changed into other landuse types, much soil C might be lost. Class three had medium contents of soil nutrients. It occurs in mid to lower slope dominated by deciduous or mixed forest. Class four was characterized by low C, N, and P contents. This cluster might indicate a vulnerable mountain area of relatively low fertility. The landscape position is typical linear mid slope positions and dominated by coniferous forest or mixed forest.

Specific carbon values as indicators of soil quality were suggested by Romig et al. (1996): healthy (40 – 60 mg/g), unhealthy (< 20 mg/g or > 80 mg/g), or impaired (20 - 40 mg/g or 60 - 80 mg/g). These values also were similar to around 40 mg/g reported by Hall (2008). Viscarra Rossel et al. (2010) and Tesfahunegn et al. (2011) suggested a lower value, 25 mg/g, for C to define high quality soil conditions. Some researches in Korea suggested the effective range ofC for the proper growth of the tree and crop is over 25 - 30 mg/g (Kang et al., 2012; Kim et al., 2010; Lee, 2012). Tesfahunegn et al. (2011) suggested 2.53 mg/g as the threshold value of N for high soil quality. Lee et al. (2012) suggested 2.5 mg/g for the effective value of the N content for the tree growth. According to these references, cluster two and three with a mean C content of 40 to 50 mg/g and a mean N content of 3 mg/g indicate healthy and high quality soil conditions (Table 2.5). Mean C and N contents of cluster one and four, 25 to 30 mg/g and

2 mg/g, are lower. According to the mentioned references, C content is still good for effective tree and crop growth, but mean N content is too small.

Sims and McGrath (2012) recommended optimum values for the plant growth by several P methods: over 10 mg/kg (Olsen) and over 30 mg/kg (Bray 1). Viscarra Rossel et al. (2010) suggested 18 - 60 mg/kg by Olsen method. Horneck et al. (2011) also suggested 40 - 100 mg/kg (Bray 1) and 25 - 50 mg/kg (Olsen) as recommendation P values for agricultural areas. Aune and Lal (1997) defined as 7 - 10 mg/kg by Bray 1 although these values are the critical P level. Lee et al. (2012) suggested 60 mg/g (Lancaster) as the effective value of the P content for forest areas. For available P, there are many values and these values are a little inconsistent because chemical analysis methods are different. Kim et al. (2013) tried to compare various methods for available P for soils in paddy and upland fields. Available P of the Lancaster method was significantly correlated with those of Bray 1 (R^2=0.94) and Olsen (R^2=0.63). The values extracted by Bray 1 overestimated the Lancaster's values while the results of Olsen method underestimated those of the Lancaster method. Therefore, 30 - 40 mg/kg are considered as the effective range of P according to the above references. Results showed class one had relatively high P contents and class two, three and four mostly covered by forest areas had low P contents (Table 2.5). More complete overviews can be found in Schoenholtz et al. (2000), Sims and McGrath (2012) and Ewing and Singer (2012).

Most of the areas, especially forest, showed low P contents possibly defining a phosphorus-limited region. The amount of N deposition has increased rapidly in Korea due to industrialization, increase in fossil fuel combustion and intensive agricultural activities (Jang et al., 2011). The total N deposition was 12.9 – 24.9 kg/ha/year from 2005 to 2010 in South Korea and the results showed an increasing trend (Jang et al., 2011). The increase in N concentrations is most due to deposition of pollutant nitrogen from atmospheric sources (Kim et al., 2011). Kim et al. (2014) also reported the rate of increase of considerable N in the North

Pacific Ocean close to the Asian continent. The nitrogen enrichment may enhance P limitation and can change mountain ecosystems in the long-term.
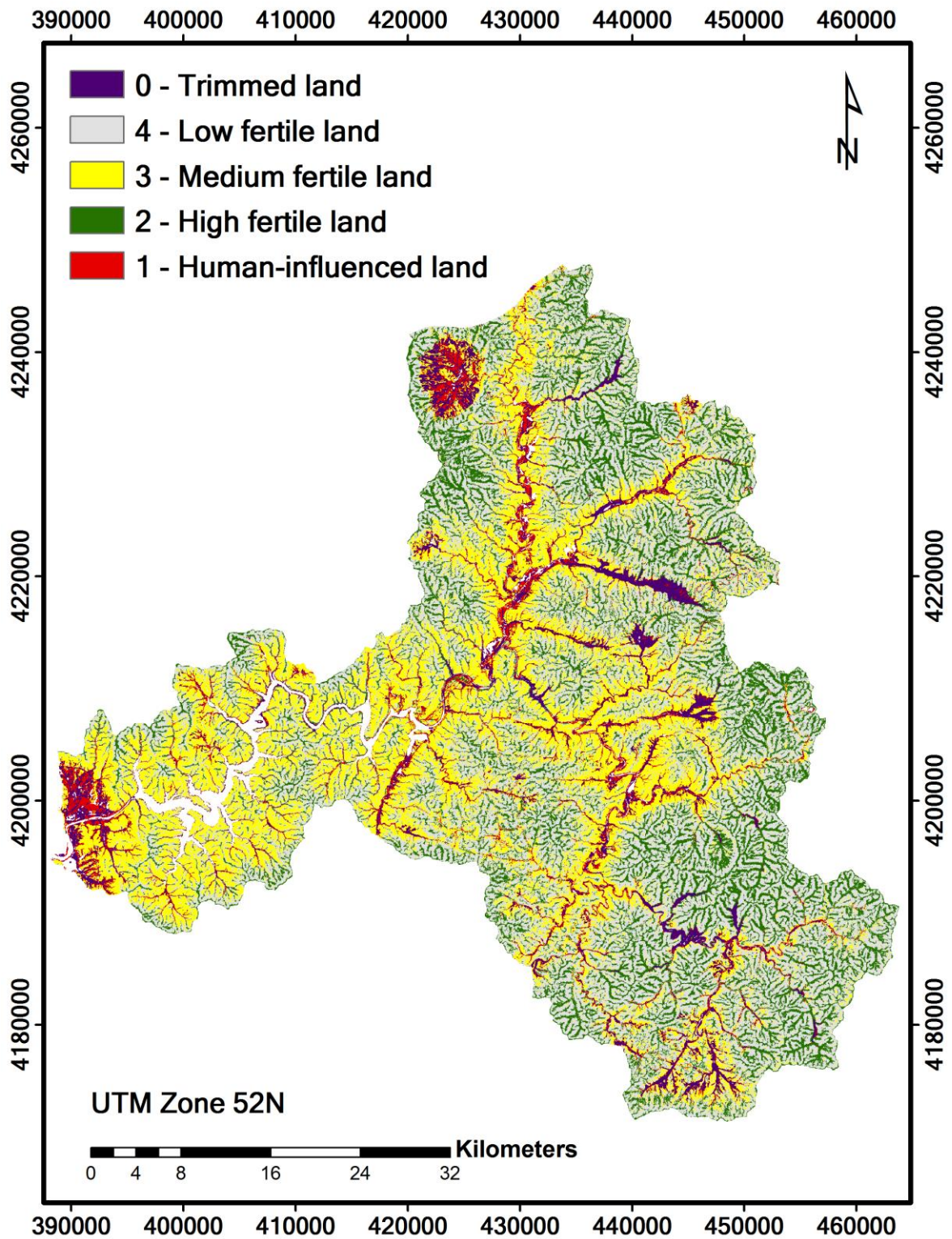
Figure 2.9 Land potential assessment map

2.4 Conclusions

In order to investigate the spatial soil fertility pattern for the Soyang Lake watershed in South Korea, soil nutrient maps were developed following specific procedures; predictor selection, model calibration and validation, and cluster analysis. Predictor selection results suggested that vegetation indices have powerful abilities to predict soil nutrients and using selected predictors via RFE improved the results. Remote sensing images indicate land use as well as vegetation density and composition which contribute to spatial patterns of soil nutrients. RF showed the best performance for all three soil nutrients.

Higher values of C and N contents can be found at higher elevation within the Soyang watershed possibly due to a high vegetation density. Areas with higher contents of topsoil available P were found in areas under agricultural land use. Spatial soil nutrient patterns were influenced by the heterogeneity of the vegetation and land use and locally governed by slope configuration. Higher uncertainty for P showed in areas with higher values because the proportion of higher values was sparsity. Higher standard deviation values of the C content were found in the coniferous forest land near the Soyang dam while high uncertainty for the N content was shown along the high ridges. Spatial prediction uncertainty for available P was higher in lower and agricultural areas than in other areas.

Cluster analysis identified four land potential classes: fertile, medium and low fertile forest lands with an additional class dominated by high P and low C and N contents due to human impacts. Cluster two and three showed relatively high mean C and N contents based on other researches. The results showed class one had relatively high P contents and class two, three and four had low P contents. Most of the areas (forest) showed low phosphorus contents possibly defining a phosphorus-limited region.

This study provides an effective approach to map land potentials. This land potential assessment could provide an integrated spatial information to manage mountain ecosystems.

2.5 Acknowledgments

2.6 References

Al-Shamiri, A., Ziadat, F.M., 2012. Soil-landscape modeling and land suitability evaluation: The case of rainwater harvesting in a dry rangeland environment. Int. J. Appl. Earth Obs. Geoinf. 18, 157–164. doi:10.1016/j.jag.2012.01.005

Arnhold, S., Ruidisch, M., Bartsch, S., Shope, C.L., Huwe, B., 2013. Simulation of runoff patterns and soil erosion on mountainous farmland with and without plastic-covered ridge-furrow cultivation in South Korea. Trans. ASABE 56, 667–679.

Aune, J.B., Lal, R., 1997. The tropical soil productivity calculator—A model for assessing effects of soil management on productivity, in: Lal, R., Stewart, B. (Eds.), Soil Management: Experimental Basis for Sustainability and Environmental Quality. Lewis Publishers, London, pp. 499–520.

Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma 151, 338–350. doi:10.1016/j.geoderma.2009.04.022

Bauer, A., Velde, B., 2014. Geochemistry at the Earth's Surface. Springer, Berlin.

Benner, J., Vitousek, P.M., Ostertag, R., Bruijnzeel, L.A., Scatena, F.N., Hamilton, L.S., 2010. Nutrient cycling and nutrient limitation in tropical montane cloud forests, in: Bruijnzeel, L., Scatena, F., Hamilton, L. (Eds.), Tropical Montane Cloud Forests, Science for Conservation and Management. Cambridge University Press, New York, pp. 90–100.

Binkley, D., Fisher, R., 2012. Ecology and management of forest soils. John Wiley & Sons, West Sussex.

Brady, N.C., Weil, R.R., 2010. Elements of the nature and properties of soils. Prentice Hall, New Jersey.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239-240, 68–83. doi:10.1016/j.geoderma.2014.09.019

Camargo, L.A., Marques, J., Pereira, G.T., Alleoni, L.R.F., 2012. Spatial correlation between the composition of the clay fraction and contents of available phosphorus of an Oxisol at hillslope scale. Catena 100, 100–106. doi:10.1016/j.catena.2012.07.016

Cherkassky, V., Mulier, F.M., 2007. Learning from data: concepts, theory, and methods. John Wiley & Sons, New Jersey.

Costa-Cabral, M.C., Burges, S.J., 1994. Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. Water Resour. Res 30, 1681–1692.

Cox, M.S., 2001. The Lancaster soil test method as an alternative to the Mehlich 3 soil test method. Soil Sci. 166, 484–489.

De Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A., Montanarella, L., van Wesemael, B., 2015. A map of the topsoil organic carbon content of Europe generated by a generalized additive mode. Eur. J. Soil Sci. 16, 121–134. doi:10.1111/ejss.12193

Elith, J., Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697. doi:10.1146/annurev.ecolsys.110308.120159

Farcomeni, A., Greco, L., 2015. Robust Methods for Data Reduction. CRC Press, Boca Raton.

Freeman, T.G., 1991. Calculating catchment area with divergent flow based on a regular grid. Comput. Geosci. 17, 413–422. doi:10.1016/0098-3004(91)90048-I

Fritz, H., 2012. tclust: An R Package for a Trimming Approach to Cluster Analysis. J. Stat. Softw. 47.

Funnell, D., Parish, R., 2005. Mountain environments and communities. Routledge, London.

Gao, B.C., 1996. NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266. doi:10.1016/S0034-4257(96)00067-3

Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298. doi:10.1016/S0034-4257(96)00072-7

Gruber, S., Peckham, S., 2009. Land-surface parameters and objects in hydrology, in: Hengl, T., Reuter, H. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 171–194.

Grunwald, S., Vasques, G.M., Rivero, R.G., 2015. Fusion of Soil and Remote Sensing Data to Model Soil Properties. Adv. Agron. 131, 1–109. doi:10.1016/bs.agron.2014.12.004

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Hall, R., 2008. Soil Essentials: Managing Your Farm's Primary Asset. Landlinks Press, Collingwood.

Harms, B., Brough, D., Philip, S., Bartley, R., Clifford, D., Thomas, M., Willis, R., Gregory, L., 2015. Digital soil assessment for regional agricultural land evaluation. Glob. Food Sec. 5, 25–36. doi:10.1016/j.gfs.2015.04.001

Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models. CRC Press, London.

Hastie, T.J., Tibshirani, R.J., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer, New York.

Holtmeier, F.-K., 2009. Mountain timberlines: ecology, patchiness, and dynamics. Springer Science & Business Media.

Horneck, D.A., Sullivan, D.M., Owen, J.S., Hart, J.M., 2011. Soil Test Interpretation Guide. http://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/22023/ec1478.pdf (accessed 8.09.2015).

Hsieh, C.H., Lu, R.H., Lee, N.H., Chiu, W.T., Hsu, M.H., Li, Y.C.J., 2011. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. Surgery 149, 87–93. doi:10.1016/j.surg.2010.03.023

Huggett, R.J., 1975. Soil landscape systems: a model of soil genesis. Geoderma 13, 1–22.

Jang, S.K., Sung, M.Y., Shin, A.Y., Choi, J.-S., Son, J.S., Ahn, J.-Y., Kim, J.C., Shin, E.S., 2011. A Study for Long-term Trend of Acid Deposition in Korea. J. Korea Soc. Environ. Adm. 17, 183–192 (in Korean).

Javaheri, S.H., Sepehri, M.M., Teimourpour, B., 2014. Response Modeling in Direct Marketing: A Data Mining-Based Approach for Target Selection, in: Zhao, Y., Cen, Y. (Eds.), Data Mining Applications with R. Academic Press, Amsterdam, pp. 153–180.

John, G., Kohavi, R., Pfleger, K., 1994. Irrelevant Features and the Subset Selection Problem, in: Cohen, W.W., Hirsh, H. (Eds.), Machine Learning: Proceeding of the Eleventh International Conference. Morgan Kaufmann Publishers, San Francisco, CA, pp. 121–129. doi:10.1145/2663598.2629474

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., Arriaga-Weiss, S., 2010. Classification in conservation biology: A comparison of five machine-learning methods. Ecol. Inform. 5, 441–450. doi:10.1016/j.ecoinf.2010.06.003

Kang, S., Roh, A., Choi, S., Kim, Y., Kim, H., Choi, M., Ahn, B., Kim, H., Kim, H., Park, J., Lee, Y., Yang, S., Ryu, J., Jang, Y., Kim, M., Sonn, Y., Lee, C., Ha, S., Lee, D., Kim, Y., 2012. Status and Changes in Chemical Properties of Paddy Soil in Korea. Korean J. Soil Sci. Fertil. 45, 968–972 (in Korean).

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. kernlab - An S4 package for kernel methods in R. J. Stat. Softw. 11, 1–20.

Kim, C., Seo, Y., Kwon, T., Park, J., 2010. Improvement in Upland Soil Management on Different Topographies and Crops. Korean J. Soil Sci. Fertil. 43, 147–152 (in Korean).

Kim, I., Le, Q.B., Park, S.J., Tenhunen, J., Koellner, T., 2014. Driving Forces in Archetypical Land-Use Changes in a Mountainous Watershed in East Asia. Land 3, 957–980. doi:10.3390/land3030957

Kim, I., Lee, K., Gruber, N., Karl, D.M., Bullister, J.L., Yang, S., Kim, T., 2014. Increasing anthropogenic nitrogen in the North Pacific Ocean. Science 346, 1102–1106.

Kim, J., Grunwald, S., Rivero, R.G., 2014. Soil Phosphorus and Nitrogen Predictions Across Spatial Escalating Scales in an Aquatic Ecosystem Using Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 52, 6724–6737.

Kim, M.S., Kwak, H.K., Kim, Y.H., Kang, S.S., Gong, M.S., 2013. Comparison of Soil Testing Methods for Plant Available Phosphate. Korean J. Soil Sci. Fertil. 46, 153–162 (in Korean).

Kim, T.W., Lee, K., Najjar, R.G., Jeong, H.-D., Jeong, H.J., 2011. Increasing N Abundance in the Northwestern Pacific Ocean Due to Atmospheric Nitrogen Deposition. Science 334, 505–509. doi:10.1126/science.1206583

Koethe, R., Lehmeier, F., 1996. SARA-System zur Automatischen Relief-Analyse: User Manual. Department of Geography, University of Goettingen.

Korea Institute of Geology Mining and Material, 2001. Explanatory note of the Gangreung Sokcho sheet 1:250,000. Korea Institute of Geology, Mining and Material, Daejeon.

Kovačević, M., Bajat, B., Gajić, B., 2010. Soil type classification and estimation of soil properties using support vector machines. Geoderma 154, 340–347. doi:10.1016/j.geoderma.2009.11.005

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer, New York.

Kunkel, M.L., Flores, A.N., Smith, T.J., McNamara, J.P., Benner, S.G., 2011. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. Geoderma 165, 1–11. doi:10.1016/j.geoderma.2011.06.011

Lee, D.K., 2012. Ecological Management of Forests. Seoul National University Press, Seoul (in Korean).

Lee, G., 2004. Characteristics of Geomorphological Surface and Analysis of Deposits in Fluvial Terraces at Upper Reach of Soyang River. J. Korean Geogr. Soc. 39, 27–44 (in Korean).

Li, Z., Ye, Z., Wan, R., Zhang, C., 2015. Model selection between traditional and popular methods for standardizing catch rates of target species : A case study of Japanese Spanish mackerel in the gillnet fishery. Fish. Res. 161, 312–319. doi:10.1016/j.fishres.2014.08.021

Ließ, M., Glaser, B., Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and Random Forest models. Geoderma 170, 70–79. doi:10.1016/j.geoderma.2011.10.010

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94. doi:10.1016/S0016-7061(98)00137-2

Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015. Impact of multi-scale predictor selection for modeling soil properties. Geoderma 239-240, 97–106. doi:10.1016/j.geoderma.2014.09.018

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Sci. Soc. Am. J. 57, 443–452.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1–19. doi:10.1016/j.geoderma.2010.12.018

National Academy of Agricultural Science, 2013. Korean Soil Information System. http://soil.rda.go.kr/eng/ (accessed 1.09.2015).

Osman, K.T., 2013. Forest Soils: Properties and Management. Springer International Publishing Switzerland.

Park, S.J., Burt, T.P., 2002. Identification and Characterization of Pedogeomorphological Processes on a Hillslope. Soil Sci. Soc. Am. J. 66, 1897–1910.

Peng, G., Bing, W., Guangpo, G., Guangcan, Z., 2013. Spatial distribution of soil organic carbon and total nitrogen based on GIS and geostatistics in a small watershed in a hilly area of northern China. PLoS One 8, 1–9. doi:10.1371/journal.pone.0083592

Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma 209-210, 1–14. doi:10.1016/j.geoderma.2013.05.029

Roger, A., Libohova, Z., Rossier, N., Joost, S., Maltas, A., Frossard, E., Sinaj, S., 2014. Spatial variability of soil phosphorus in the Fribourg canton, Switzerland. Geoderma 217-218, 26–36. doi:10.1016/j.geoderma.2013.11.001

Roman, L., Scatena, F.N., Bruijnzeel, L.A., 2010. Global and local varaitions in tropical montane cloud forest soils, in: Bruijnzeel, L., Scatena, F., Hamilton, L. (Eds.), Tropical Montane Cloud Forest: Science for Conservation and Management. Cambridge University Press, New York, pp. 77–89.

Romig, D.E., Garlynd, M.J., Harris, R.F., 1996. Farmer-based assessment of soil quality: a soil health scorecard, in: Doran, J., Jones, A. (Eds.), Methods for Assessing Soil Quality. Soil Science Society of America, pp. 39–60.

Roudier, P., Beaudette, D.E., Hewitt, A.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints, in: Minasny, B., Malone, B., McBratney, A. (Eds.), Digital Soil Assessments and beyond. CRC Press, Boca Raton, pp. 227–231.

Ruidisch, M., Arnhold, S., Huwe, B., Bogner, C., 2013. Is ridge cultivation sustainable? A case study from the haean catchment, South Korea. Appl. Environ. Soil Sci. 2013. doi:10.1155/2013/679467

Schoenholtz, S.H., Miegroet, H. Van, Burger, J.A., 2000. A review of chemical and physical properties as indicators of forest soil quality: Challenges and opportunities. For. Ecol. Manage. 138, 335–356. doi:10.1016/S0378-1127(00)00423-0

Sims, J.T., McGrath, J., 2012. Soil fertility evaluation, in: Huang, P., Li, Y., Summer, M. (Eds.), Handbook of Soil Sciences: Resource Management and Environmental Impacts. CRC Press, Boca Raton.

Singer, M.J., Ewing, S., 2012. Soil Quality, in: Huang, P., Li, Y., Summer, M. (Eds.), Handbook of Soil Sciences: Resource Management and Environmental Impacts. CRC Press, Boca Raton.

Smola, A., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. doi:Doi 10.1023/B:Stco.0000035301.49549.88

Stockdale, E.A., Goulding, K.W.T., George, T.S., Murphy, D. V, 2013. Soil fertility, in: Gregory, P., Nortcliff, S. (Eds.), Soil Conditions and Plant Growth. John Wiley & Sons, West Sussex, pp. 49–85.

Strahler, A., 1957. Quantitative analysis of watershed geomorphology. Trans. Am. Geophys. Union 38, 913–920. doi:10.1130/0016-7606

Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods 14, 323–348. doi:10.1037/a0016973

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. doi:10.1016/j.ecolind.2007.05.005

Sun, X.L., Wu, S.C., Wang, H.L., Zhao, Y.G., Zhao, Y., Zhang, G.L., Man, Y.B., Wong, M.H., 2012. Uncertainty Analysis for the Evaluation of Agricultural Soil Quality Based on Digital Soil Maps. Soil Sci. Soc. Am. J. 76, 1379. doi:10.2136/sssaj2011.0426

Tesfa, T.K., Tarboton, D.G., Chandler, D.G., McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45, 1–16. doi:10.1029/2008WR007474

Tesfahunegn, G.B., Tamene, L., Vlek, P.L.G., 2011. Evaluation of soil quality identified by local farmers in Mai-Negus catchment, northern Ethiopia. Geoderma 163, 209–218. doi:10.1016/j.geoderma.2011.04.016

Tucker, C.J., Sellers, P.J., 1986. Satellite remote sensing of primary production. Int. J. Remote Sens. 7, 1395–1416. doi:10.1080/01431168608948944

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag, New York.

Viscarra Rossel, R.A., Rizzo, R., Demattê, J.A.M., Behrens, T., 2010. Spatial Modeling of a Soil Fertility Index using Visible–Near-Infrared Spectra and Terrain Attributes. Soil Sci. Soc. Am. J. 74, 1293. doi:10.2136/sssaj2009.0130

Wilcke, W., Boy, J., Goller, R., Fleischbein, K., Valarezo, C., Zech, W., 2010. Effect of topography on soil fertility and water flow in an Ecuadorian lower montane forest, in: Bruijnzeel, L., Scatena, F., Hamilton, L. (Eds.), Tropical Montane Cloud Forests, Science for Conservation and Management. Cambridge University Press, New York, pp. 402–409.

Wilcke, W., Boy, J., Hamer, U., Potthast, K., Rollenbeck, R., Valarezo, C., 2013. Current regulating and supporting services: Nutrient cycles, in: Bendix, J., Beck, E., Bräuning, A., Makeschin, F., Mosandl, R., Scheu, S., Wilcke, W. (Eds.), Ecosystem Services,

Biodiversity and Environmental Change in a Topical Mountain Ecosystem of South Ecuador. Springer-Verlag, Berlin, pp. 141–151.

Wohl, E., 2010. Mountain rivers. American Geophysical Union, Washington, DC.

Wood, S., 2006. Generalized additive models: an introduction with R. CRC press, Boca Raton.

Yokoyama, R., Shlrasawa, M., Pike, R.J., 2002. Visualizing Topography by Openness : A New Application of Image Processing to Digital Elevation Models 68, 257–265.

Zevenbergen, L.W., Zevenbergen, L.W., Thorne, C.R., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12, 47–56.

Chapter 3 Spatial patterns of soil phosphorus fractions in a mountainous watershed

Gwanyong Jeong[1*], Marie Spohn[2], Soo Jin Park[3], Bernd Huwe[1], Mareike Ließ[4]

[1]*Department of Geosciences/ Soil Physics Division, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitaetsstrasse 30, 95447 Bayreuth, Germany*

[2]*Department of Soil Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Dr.-Hans-Frisch-Straße 1-3, 95448 Bayreuth, Germany*

[3]*Department of Geography, Seoul National University, Shilim-Dong, San 56-1, Kwanak-Gu, Seoul 151-742, South Korea*

[4]*Department of Soil Physics, Helmholtz Centre for Environmental Research - UFZ, Theodor-Lieser-Strasse 4, D-06120 Halle, Germany*

Corresponding author: gwanyong.jeong@uni-bayreuth.de

Abstract

Topography might affect the spatial patterns of soil phosphorus (P) concentrations and on the abundance of soil P fractions in mountainous landscapes. However, few attempts have been made to quantify relationships between topography and soil P fractions. We investigated the spatial patterns of various P fractions in the topsoil of a forested watershed with steep slopes ranging from 300 to 900 m above sea level in South Korea. For this purpose, different P fractions including, total P, resin Pi, bicarbonate Pi, bicarbonate Po, hydroxide Pi, hydroxide Po, apatite Pi, and residual P were determined in 91 soil samples by Hedley fractionation. We tested terrain parameters and the normalized difference vegetation index to predict spatial patterns of the P fractions.

Surface curvature contributed to 27 ~ 68 % of the total variations explained by regression models for each P fraction, and elevation contributed to 23 ~ 66% of the total variations. Based on surface curvature and elevation, we created digital soil maps depicting the spatial distribution of the P fractions. Total soil P concentration decreased strongly from lower slope to upper slope both at the low (300-600m) and high altitude (600-900m). While organic P was enriched at the lower slope likely due to erosion, the residual P fractions was largest at the upper slope. In conclusion, our results show that topography influences the spatial patterns of P fractions, allowing to predict areas which might be more sensitive to P limitation.

Keywords: phosphorus; Hedley fractionation; digital soil mapping; topography; surface curvature; uncertainty.

3.1 Introduction

Phosphorus is an essential macronutrient for all organisms on Earth. Yet, little is known about the spatial distribution of P fractions in soil. Only a small proportion of soil total P is bioavailable, and most soil P is sorbed to calcium ions in young soils or to iron and aluminium ions in older soils (Walker and Syers, 1976). Recent progress in digital soil mapping (DSM) might help to improve our understanding of the spatial pattern of soil P fractions.

Topography might be critical for the spatial patterns of soil P because it controls flows of water and materials (Huggett, 1975; Moore et al., 1993; Park and Burt, 2002). Topography proved to have effects on spatial variability of total soil P concentrations (Cheng et al., 2016; Kim and Zheng, 2011; McKenzie and Ryan, 1999; Yoo et al., 2009), available soil P (Camargo et al., 2012; Moore et al., 1993) and various P forms (total, available, and organic) (Roger et al., 2014). Camargo et al. (2012) identified curvature to be useful to understand spatial variability of available P.

Some studies identified the spatial distribution of P fractions gained by the Hedley fractionation on catenary sequences (Agbenin and Tiessen, 1994; Araújo et al., 2004; Roberts et al., 1985; Smeck, 1985; Tiessen et al., 1994; Vitousek et al., 2003). Agbenin and Tiessen (1994) reported that P contents decreased downslope in semi-arid regions of Brazil. Araújo et al. (2004) studied spatial variations of P fractions in toposequences of semi-arid soils and observed a spatial trend of organic P to increase downslope. Vitousek et al. (2003) found residual P decreased and organic P increased downslope. Mage and Porder (2013) confirmed parent material and topographical positions (ridge, slope, and valley) explained the variances of P status. These studies confirmed topographic effects on a variety of P fractions. However, quantitative relationships between soil P fractions and topography have not yet been fully investigated, and there are only very few studies that used DSM to better understand the spatial patterns of P fractions in complex landscapes.

The great progress in DSM makes it possible to develop numerical or statistical models of the relationship between environmental predictors and soil properties (Grunwald, 2009; McBratney et al., 2003; Minasny and McBratney, 2015; Scull et al., 2003). A characteristic of DSM is that the soil of interest at "unknown" locations can be predicted quantitatively onto a digital soil map with uncertainty using the relationship between soil and environmental predictors at "known" locations under similar environmental conditions. Moreover, terrain analysis with digital elevation model (DEM) provides a variety of predictors which might represent the environment and indicate hydrological, geomorphological and pedological processes (Sabine Grunwald, 2006; McBratney et al., 2003). These are useful tools to identify and understand spatial variability of soil properties.

In the present study, we analyzed the spatial pattern of several soil P fractions in the topsoil of a mountainous watershed in South Korea using DSM. The purpose of this study was (1) to identify the environmental predictors of the spatial distribution of several P fractions, and (2) to

map different P fractions using the quantitative relationships between P fractions and environmental predictors.

3.2 Materials and methods

3.2.1 Research area

This study area is a small headwater catchment with confined depositional areas (Wohl, 2010) located in the downstream area of the Soyang watershed, Gangwon-do province, South Korea (37° 59' 21" - 37° 59' 52" N and 127° 49' 51" - 127° 52' 03" E) (Figure 3.1). During 30 years, the average temperature of the air has been 11.1 °C (-4.6 – 24.6 °C) and the area has received a mean of 1,347 mm rain annually, with about 824 mm falling between June and August (Korea
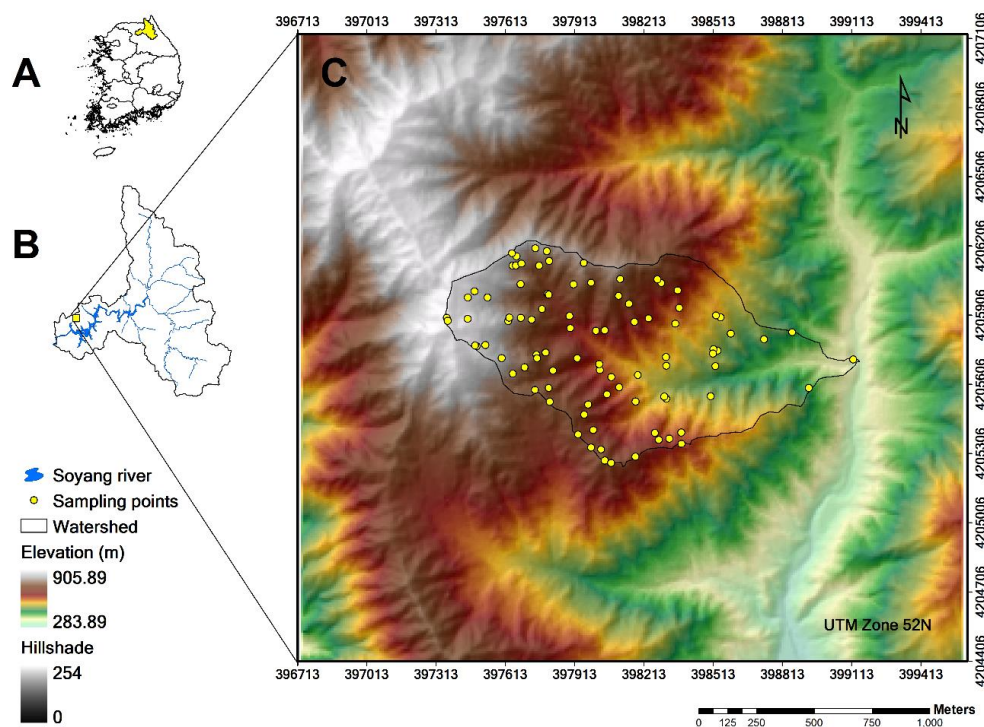
Figure 3.1 Research area. (A) South Korea. (B) Soyang watershed is located in the north-eastern part of South Korea. (C) The map shows elevation and hillshade of our research area and a spatial pattern of sampling points.

meteorological administration, 2015). About 70 % of the annual rain falls heavily in the summer monsoon season (Bartsch et al., 2014). The area's geology is dominated by granitic gneiss and banded gneiss (Korea Institute of Geology Mining and Material, 2001). It extends between 320 and 868 m a.s.l. and has an area of 9.84 km^2 and comprises steep slopes (over 45°). The soil texture is gravelly fine sandy loam, fine sandy loam (National Academy of Agricultural Science, 2013). It is covered by the national forest formed by Mongolian oak (*Quercus mongolica*) (40 - 50 years) and Korean pine (*Pinus koraiensis*) (30 - 35 years), with local occurrence of Japanese red pine (*Pinus densiflora*) and Japanese larch (*Larix kaempferi*).

3.2.2 Soil sampling and chemical analyses

Conditioned Latin Hypercube Sampling (cLHS) was applied to represent the density functions of the environmental predictor space and get a good dataset for regression modelling (Minasny and McBratney, 2006). The cLHS was performed with R package "clhs" (Roudier et al., 2012). A Qmini GNSS (global navigation satellite system) GPS was used to reduce the positioning in accuracy during field work to less than 5 m. A total number of 91 soil samples was collected from the A horizon in 2014 (Figure 3.1).

The soil samples were air-dried and sieved (< 2 mm). Several P fractions were obtained according to Hedley method (Hedley et al., 1982) The sequential extractions method first removes phosphate ions by anion exchange resins and then available and the more stable inorganic P and organic P fractions using a series of sequentially stronger extracting reagents. We used 0.5 g soils for the extraction. Resin Pi, bicarbonate Pi (NaHCO$_3$ Pi), bicarbonate Po (NaHCO$_3$ Po), hydroxide Pi (NaOH Pi), hydroxide Po (NaOH Po), apatite Pi (HCl Pi), and residual P (H$_2$SO$_4$/H$_2$O$_2$ P) were measured according to DIN EN ISO 6878 (DEV, 2002) using UV/VIS-spectroscopy (Perkin Elmer, Lambda 2, USA) in EUROFINS, Jena, Germany. The P fractions were reclassified into total P, resin Pi, available P (Resin Pi, bicarbonate Pi, and bicarbonate Po), organic P (bicarbonate Po and hydroxide Po), apatite Pi, and residual P (Yang

Table 3.1 Environmental predictors for digital soil mapping.

	Predictor	Method	Reference
1	Elevation (ELEV)	-	-
2	Slope degree (SLO)	Slope, aspect, curvature saga module	(Zevenbergen et al., 1987)
3	Strahler order ≥5 Overland flow distance to channel network (OFD)	Overland low distance to channel network saga module	(Strahler, 1957)
4	Solar radiation (SOL)	Potential incoming solar radiation saga module	(Böhner and Antonic, 2009)
5	Surface curvature (CUR)	CURV3 program	(Park et al., 2001)
6	Normalized difference vegetation index (NDVI)	(NIR – Red)/ (NIR+Red)	(Tucker and Sellers, 1986)

NIR= near-infrared

and Post, 2011). After grinding the samples, total carbon was measured by an elemental analyser (NA 1108, CE Instruments, Milano, Italy).

3.2.3 Environmental predictors

The 10 m digital elevation model (DEM) was produced using LiDAR (Light Detection and Ranging) point data surveyed by the National Geographic Information Institute (NGII) in South Korea (National Geographic Information Institute, 2015). The point observation data had a vertical accuracy of below 10 cm and an average of 4.08 points/ m^2. 10 – 30 m grid sizes are recommended for digital soil mapping (Cavazzi et al., 2013; Erskine et al., 2007; Kim and Zheng, 2011; Maynard and Johnson, 2014; Park et al., 2009). This is because topographical details decrease with grid size over 30 m, while accuracy of DEM below 10 m can be easily affected by temporal and measurement errors. Topographical predictors were calculated with the terrain analysis modules of the open source software SAGA (Conrad et al., 2015) (Table

3.1). The normalized difference vegetation index (NDVI) was extracted from a 4 m Kompsat-2 image from 11th October 2014. The NDVI was changed into 10 m cell size for DSM.

Curvature is sensitive to changing neighborhood extent (Wood, 2009). It is normally calculated based on 3x3 cell neighborhood extents that can only consider the local variability. Effects of the neighborhood extent have an influence on values of terrain predictors (e.g. slope, aspect, and curvature) and finally on results of digital soil mapping (MacMillan and Shary, 2009; Maynard and Johnson, 2014; Wood, 2009). We investigated effects of the neighborhood extent for surface curvature. One advantage of this approach is an ability to consider the variability of slope configuration from fine scale to larger scale without decreasing DEM accuracy. Changing the cell size (over 30 m) can have the same effect but can lose original topographic details of the DEM.

The surface curvature (CUR) value reflects the degree of bending three-dimensional surface morphology (Park et al., 2001). Positive curvature value indicates a convex morphology, while the negative curvature value represents a concave slope (Equation 3.1). The degree of bending surface morphology increases with increasing the value. The value will be close to zero at the linear slope. In the large extent, negative surface curvature value indicates predominantly concave lower slope, while positive surface curvature value indicates convex upper slope.

$$CUR = (\sum_{i=1}^{n}(Zc_i - Zs_i)/d)/n, \qquad (3.1)$$

Where Zc is the elevation of the current cell, Zs is the elevation of surrounding points, d is the horizontal distance between the two points, and n is the total number of surrounding points.

Solutions with 3x3 to 35x35 window sizes (30 ~ 350 m neighborhood extents) were tested and the best solution was selected based on results of Pearson's correlation (Park et al., 2001). To

check variations of surface curvature with changing window size, the coefficient of variation of surface curvature was calculated.

3.2.4 Linear regression model and ANOVA analysis

A linear regression model was used to examine the influence of topography on P fractions. Spatial patterns of P fractions can be predicted quantitatively using the relationships terrain predictors and different P fractions.

Predictor selection improves model results (Brungard et al., 2015; Miller et al., 2015; Poggio et al., 2013). Two methods were used for predictor selection. We used the stepAIC function (backward stepwise) which starts a model including all predictors and deletes one at a time until removing predictors would degrade the quality of the model based on akaike information criterion (AIC). The R package "MASS" was used for the stepAIC function (Venables and Ripley, 2002). Additionally, we used recursive feature elimination (RFE) which is a backward selection algorithm and iteratively eliminates the least important variables from the model based on an initial variable importance measure (Kuhn and Johnson, 2013). However, the RFE was incorporated with resampling (5 repetition 10-fold cross validation) to select the final list of predictors. The R package "caret" provides the functions for RFE (Kuhn and Johnson, 2013).

Relative weights, measure the importance of predictors, approximates relative contributions of each predictor to R-square (R^2) in the linear regression (Johnson, 2000). It is a useful mean to interpret the strength of each predictor in statistical relationships with the other predictors. Kabacoff (2015) provided a r code for relative weights.

Model performance was tested by 5 repetitions of a 10-fold cross. As a consequence, 50 models were adapted for each P fraction. All predictions should be accompanied by estimates of uncertainty which can give users or planners information on some constraint on the digital

soil maps. The standard deviations of 50 model results were considered as predictions of uncertainty (Ließ et al., 2014).

One-way analysis of variance (ANOVA) was used to identify differences in P fractions between topographic positions. The assumptions required for ANOVA were assessed using Shapiro for normality and Bartlett test for homogeneity of variances. In case that the assumptions of the normality and the equality of variance was not met, Kruskal-Wallis test was used. The R package "pgirmess" was used for Kruskal-Wallis test (Siegel and Castellan, 1988).

3.3 Results

3.3.1 Phosphorus fractions

The mean total P of the A horizons was 389 ± 171 mg kg^{-1} (Appendix 3.1). The total P concentrations in the A horizons showed a high variance based on values of standard deviation and coefficient of variation (CoV). Organic P revealed the highest mean content (158 ± 86 mg kg^{-1}) compared to other P fractions, while Resin Pi had the smallest mean content (3.42 ± 1.63 mg kg^{-1}). Apatite Pi had the highest value of CoV (87.15 %) and exhibited a strong skewness and kurtosis with 4.15 and 27.71. A log transformation applied to resolve right-skewed distribution of apatite Pi. Apatite P ranged from 0.94 to 12.05% of total P with a mean of 2.48%, irrespective of the topographic position and altitude.

3.3.2 Models and predictors of soil phosphorus fractions

Surface curvature was sensitive to changing the neighborhood extent (Figure 3.2). For example, values of Pearson correlation between total P and curvature ranged from -0.25 to -0.50 (r). The 19 X 19 window size (CUR19) showed best results for most P fractions. This indicates that surface curvature was measured within 95 m radius at each cell. The window size increased with highlighting much more trends in curvature that indicates more distinct differences between the upper slope and the lower slope. Coefficient of variations reached the zero value almost at the plateau because the variations of curvature decrease with window size (Appendix 3.2).

Surface curvature and elevation were selected for all P fractions by both predictor selection methods (Table 3.2). Slope degree was also an important predictor for all P fractions. The stepAIC function showed better results in almost all P fractions than recursive feature elimination but differences of performances were negligible (Figure 3.3). Selected predictors for total P, resin Pi, and apatite Pi were the same and similar predictors for the others were
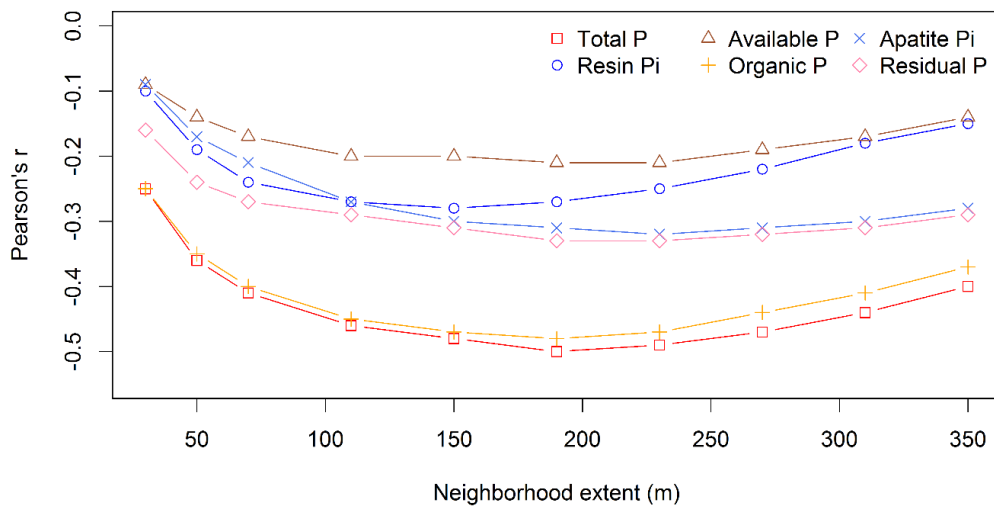
Figure 3.2 Correlations (r) between soil phosphorus fractions and surface curvature with varying the neighbourhood extent.

Table 3.2 Selected predictors.

	stepAIC	RFE
Total P	CUR19, ELEV, SOL, OFD, SLO	CUR19, ELEV, SLO, OFD, SOL
Resin Pi	CUR19, ELEV, OFD, SLO	CUR19, ELEV, SLO, OFD
Available P	CUR19, ELEV, OFD, SLO	ELEV, CUR19
Organic P	CUR19, ELEV, OFD, SLO, SOL	CUR19, ELEV, SLO, OFD, SOL, NDVI
Apatite Pi	CUR19, ELEV, SLO	CUR19, ELEV, SLO
Residual P	CUR19, ELEV, SLO, SOL, NDVI	CUR19, ELEV, SLO, SOL

RFE: recursive feature elimination, CUR: surface curvature, ELEV: elevation, SOL: solar radiation, OFD: overland flow distance to channel network, SLO: slope degree, NDVI: normalized difference vegetation index.
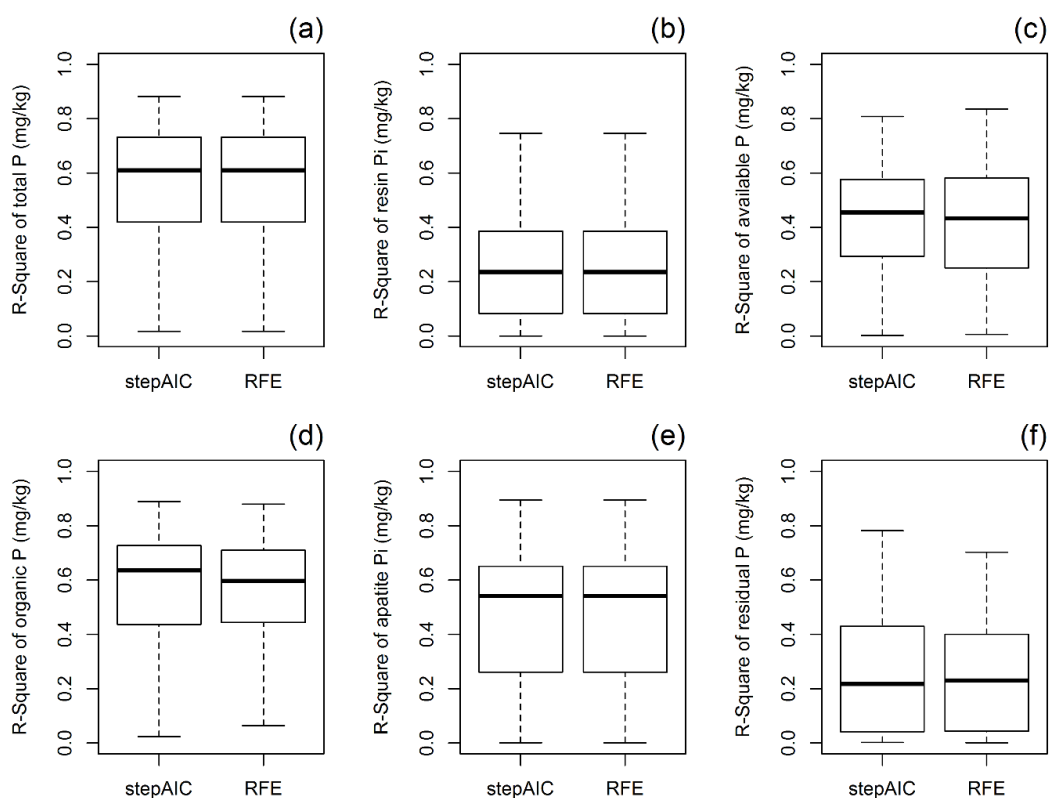
Figure 3.3 Model validation tested by 5 repetitions of a 10-fold cross with the stepAIC function and recursive feature elimination (RFE).

Table 3.3 Relative importance of predictors.

Predictors	Total P (%)	Resin Pi (%)	Available P (%)	Organic Po (%)	Apatite Pi (%)	Residual P (%)
CUR19	57.56	55.06	27.48	54.18	67.91	58.76
	(3.69)	(5.62)	(3.92)	(4.00)	(4.08)	(6.58)
ELEV	31.71	28.98	65.95	36.05	29.24	23.22
	(3.83)	(5.99)	(3.66)	(3.70)	(4.03)	(5.97)
SOL	2.00			1.44		11.45
	(0.95)			(0.86)		(3.79)
OFD	3.95	7.35	4.11	3.76		
	(0.52)	(2.01)	(1.11)	(0.52)		
SLO	4.78	8.61	2.46	4.56	2.84	6.57
	(1.26)	(3.64)	(1.18)	(1.36)	(0.68)	(2.47)
NDVI						

Mean of 50 models, standard deviation of 50 models in parentheses.

selected by both methods. Predictors for P fractions except residual P were selected by the stepAIC function.

Our linear model showed good performances for all P fractions except for resin Pi and residual P based on R^2. Our results indicate relationships between environmental predictors and P fractions might be linear in our research area. Organic P showed the best result of the external validation with a mean of 0.58 (R^2), while resin Pi ($R^2 = 0.26$) and residual P ($R^2 = 0.25$) fitted least according to R^2. Total P fraction ($R^2 = 0.56$) showed a better performance.

Terrain predictors account for 26 – 58% of the total variations of P fractions in linear regression models. In the results of predictor importance, surface curvature is one of the best predictors for all P fractions except for available P (Table 3.3). Surface curvature contributed to 28 ~ 68 % of the total variations explained by regression models for each P fraction. Elevation contributed to 23 ~ 66% of the total variations explained by regression models. Considering a low relative

weight of vegetation index (NDVI), terrain predictors showed promising results for prediction of P fractions.

3.3.3 Spatial patterns of soil P fractions

Differences of P fractions were identified based on elevation and curvature which were selected as the best predictors. Although the topographic classification criteria were rather arbitrary, surface curvature values were set at -1 and 1 to classify lower slope (< -1), mid slope (-1 to 1), and upper slope (> 1). The total P contents, and the proportions of organic P and residual P were significantly different between topographical positions (Figure 3.4). Total P was 2.2 times higher at the lower slope than on the upper slope at an altitude of 300-600 m, and 1.8 times at an altitude of 600-900 m (Figure 3.4a and b). Total P did not differ between the corresponding topographic positions at 300-600 and 600-900 m (Figure 3.4a and b). The proportion of organic P was significantly higher at the lower slope than on the upper slope at an altitude of 600-900 m (Figure 3.4c and d). The proportion of residual P showed the opposite trend than the organic P (Figure 3.4c and d). The proportion of residual P was highly correlated with the proportion of organic P (r=-0.97, p<0.001). The carbon-to-organic P ratios decreased with topographical positions (Figure 3.5c). Carbon contents had no significant relationship with surface curvature (r=-0.02, p=0.93) but carbon-to-organic P ratios strongly correlated with curvature (r=0.56, p<0.001).

The maps for each P fraction display the mean of 50 predictions (Figure 3.6). We found evident differences between P fractions at different topographic positions in our maps. Higher values of total P contents were found at the lower slope. Total P concentrations increased with elevation (Figure 3.6a). Resin Pi showed higher contents at concave lower slopes than at convex upper slopes. Concentrations of available P were predicted to increase with elevation. Organic P, apatite Pi and residual P showed catenary differentiations across the landscape.
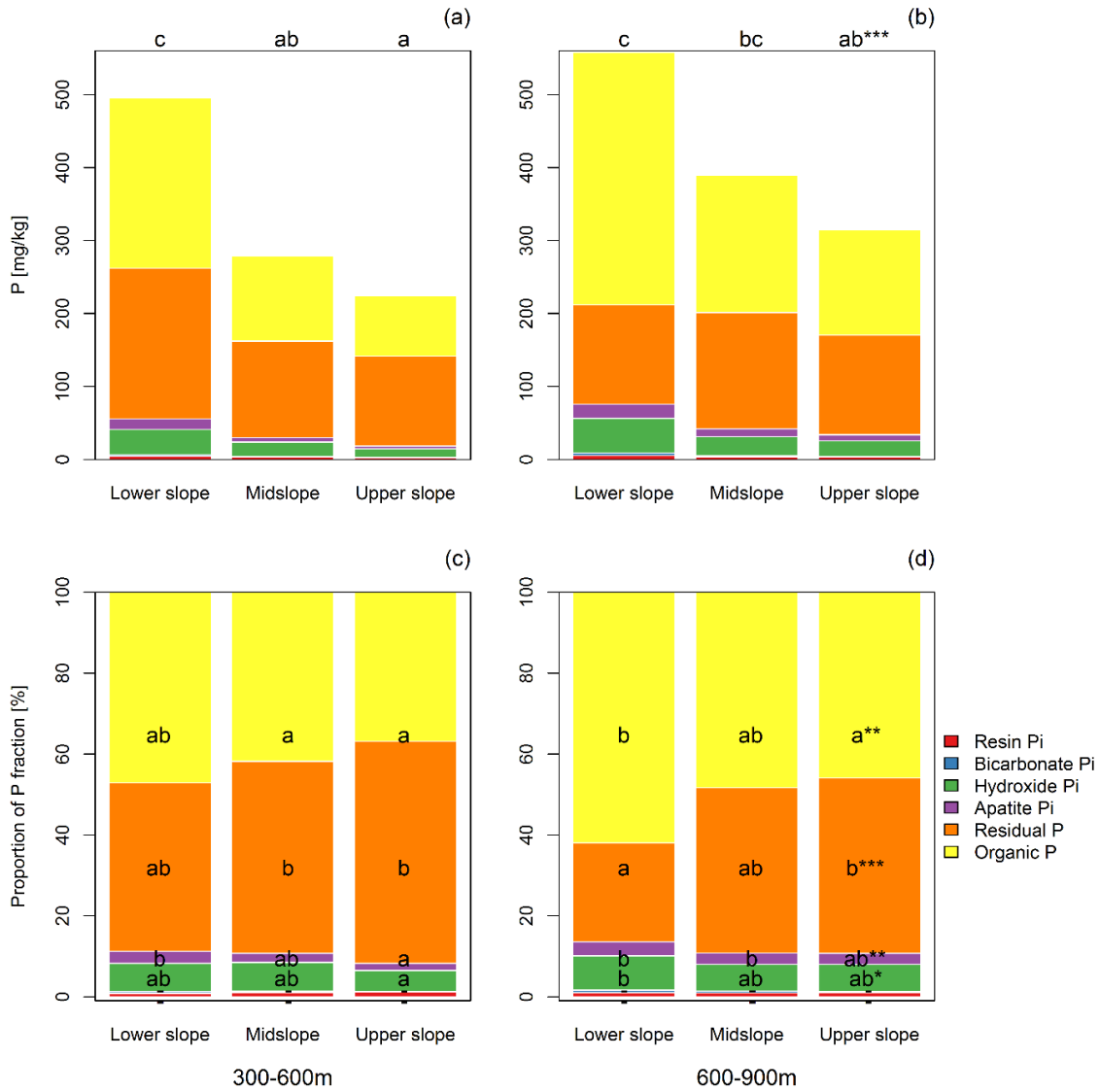
Figure 3.4 Concentrations of the phosphorus fractions at sites at an altitude of 300-600 m (a) and 600-900 m (b), and percentages of phosphorus fractions at sites at an altitude of 300-600 m (c) and 600-900 m (d). Different letters indicate significant differences significantly different at $p<0.05$ with Kruskal-Wallis test. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, and – stands for not significant.
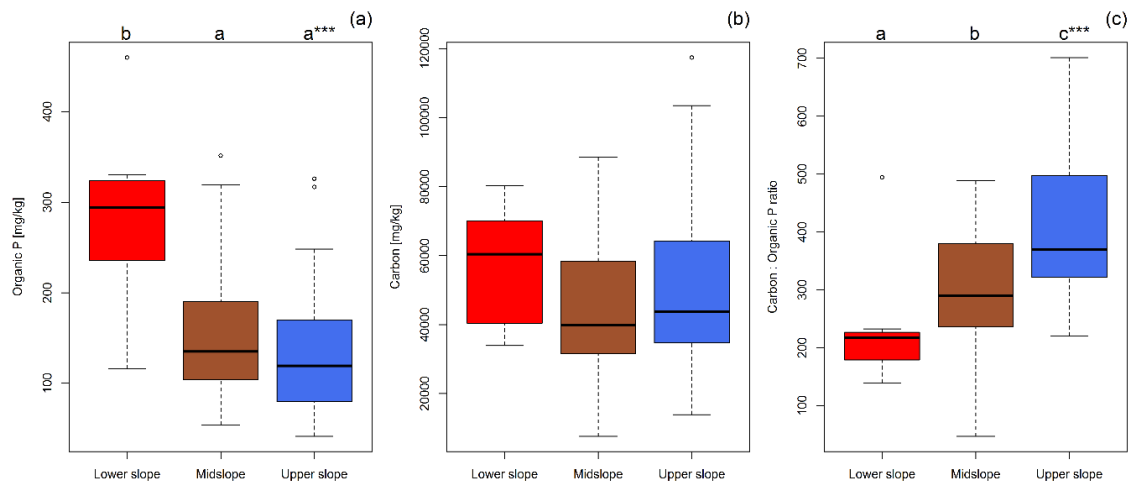
Figure 3.5 Organic phosphorus (a), carbon (b), and carbon: organic phosphorus ratio (c) at the different topographical positions. Different letters indicate significant differences at p<0.05 with Kruskal-Wallis test. * p<0.05, ** p<0.01, and *** p<0.001.

Higher standard deviation of total P contents, resin P, available P, and organic P were found in the valley floor (Appendix 3.3). The spatial uncertainty for apatite P was high in the hillslope summit (Appendix 3.3e), while high uncertainty of residual P contents was found in the more extended areas compared to other fractions (Appendix 3.3f).
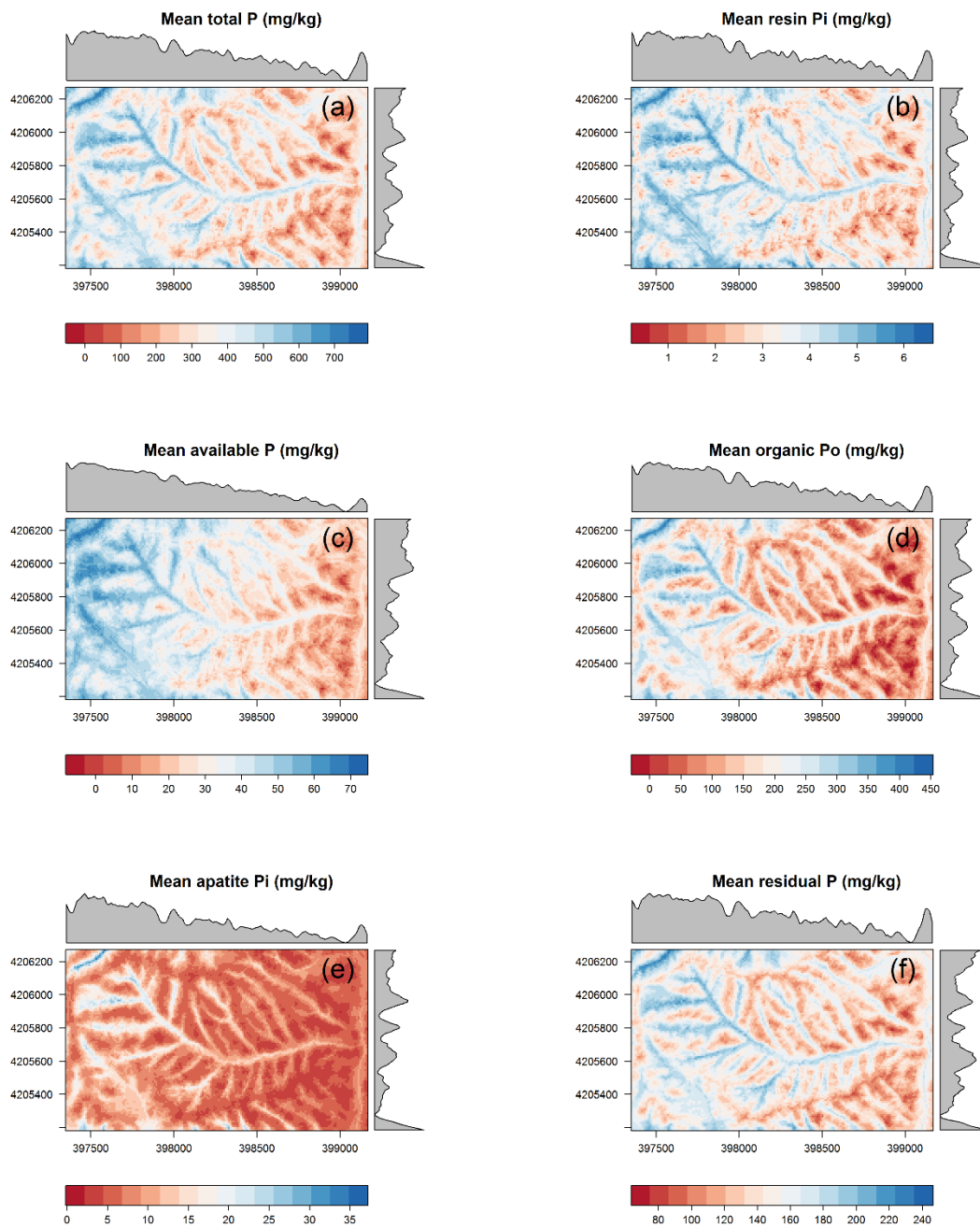
Figure 3.6 Predicted mean phosphorus fractions with summaries of the column and row.

3.4 Discussion

Our results show substantial effects of topography on soil total P and on P fractions (Figure 3.6). This pattern can most likely be explained by erosion and soil creep at the upper slope, and colluvial deposition and accumulation at the lower slope.

Concentrations of total and of organic P were higher at the lower slope than at the upper slope, while the residual P concentration was higher at the upper slope (Figure 3.6d). Similarly, McKenzie and Ryan (1999) identified higher total P at the depositional lower slope. The reason for this pattern is most likely erosion especially of organic matter, leading of an accumulation of organic P at the lower slope. Concentration of total P and organic P were significantly lower at the upper than at the lower slope (Figure 3.4).

The different ratios of C-to-organic P are likely due to two processes (Figure 3.5b). First, higher total P contents at the lower slope might reduce organic P mineralization, leading to low ratios of C-to-organic P at the lower slope. Second, the low ratios of C-to-organic P might also be due to litter inputs to the soil with low C-to-P ratios at the lower slope caused by the high total P contents at this topographic position. The high total P contents likely lead to low C-to-P ratios in biomass which affects the C-to-P ratios of the soil organic matter through litter inputs (Huang and Spohn, 2015).

Resin P is bioavailable for plants, and its concentration strongly depends on the activity of microorganisms that mobilize P (Richardson et al., 2009). Vincent et al. (2014) identified decreases of resin P contents with increasing elevation in a subarctic tundra. They explained this result by a decline in temperature that influences soil microbial activity. The curvature may indicate soil moisture, since soils on convex slopes are usually drier because water is moving down the slopes, while soils on convergent lower slope tend to be wetter. For this reason, slope curvature might have an effect on microbial activity. Moore et al. (1993) reported a similar result, showing a relationship between available P and lower slope degree, stream power index

and higher wetness index which indicate the lower slope. C contents had significant relationships with available P contents (r=0.67, p<0.001) and elevation (r=0.51, p<0.001). This indicates that the spatial pattern of P availability in our research area was related to organic matter.

Several ratios of P fractions are used to qualitatively estimate the degrees of soil weathering. The apatite Pi-to-total P ratio is useful for estimating the long term apatite P reserve of a soil and the extent to which soil minerals have undergone P transformations. Weathering results in a progressive change of apatite P to residual P (Walker and Syers, 1976). Lightly weathered soils have a high percentage of P in apatite (47%), while in strongly weathered soils, apatite P only accounts for about 3% of total soil P (Yang and Post, 2011). Residual P increase with the weathering stage of soils so in highly weathered soils, residual P-to-total P ratio showed 0.50 ~ 0.59 (Yang and Post, 2011). Residual P-to-total P ratios ranged from 20.3 to 73.3% with a mean of 45.00%, showing a high variability of ratios compared to other results (Agbenin and Tiessen, 1994; Araújo et al., 2004). Total P contents in the soils under study ranged from 160 to 920 mg kg^{-1} with a mean of 389 mg kg^{-1}. They corresponded to an average P content of a global soil P dataset (Yang and Post, 2011), although the variation was high.

Our maps of soil P fractions were produced with predictabilities from 25% to 58%. Topography explained 26% of resin P variance and 25% of residual P variance in the regression model (Figure 3.3). Moore et al. (1993) accounted for 48% of available P variance. McKenzie and Ryan (1999) reached high predictive values for total P (78%) but Kim and Zheng (2011) reported relatively lower R^2 (0.11) of total P in a coastal dune. Based on these results, total P (56%), organic P (58%), available P (44%), and apatite P (46%) reached high predictive values only using terrain variables, but resin Pi and residual P showed relatively low predictive values. This implies that there are other factors that shape the distribution of resin P and residual P that were not included in our model.

The highest uncertainties of spatial prediction for all P fractions were found in the adjacent valley floor, while the uncertainties for P fractions were lower in the ridge to mid-slope areas (Appendix 3.3). The uncertainties were related to the data structure. Less samples were collected in the adjacent valley than others and the unbalanced samples might have effects on the uncertainty. However, this is because of a geomorphological characteristic of our research area located in the upstream that has relatively small areas of lower slope where soils form and deposit.

3.5 Conclusions

This is one of the first studies that applied DSM to better understand and visualize spatial distributions of soil P fractions. Our study revealed that topography influenced the spatial variability of soil P fractions across a mountainous watershed. Especially organic P was strongly affected by surface curvature. This has important implications for soil fertility in mountainous ecosystems with low P availability and it might influence the productivity, biodiversity, and community composition of plants.

3.6 Acknowledgments

3.7 References

Agbenin, J.O., Tiessen, H., 1994. Phosphorus transformations in a toposequence of lithosols and cambisols from semi-arid northeastern brazil. Geoderma 62, 345–362. doi:10.1016/0016-7061(94)90098-1

Araújo, M.S.B., Schaefer, C.E.R., Sampaio, E.V.S.B., 2004. Soil phosphorus fractions from toposequences of semi-arid Latosols and Luvisols in northeastern Brazil. Geoderma 119, 309–321. doi:10.1016/j.geoderma.2003.07.002

Bartsch, S., Frei, S., Ruidisch, M., Shope, C.L., Peiffer, S., Kim, B., Fleckenstein, J.H., 2014. River-aquifer exchange fluxes under monsoonal climate conditions. J. Hydrol. 509, 601–614. doi:10.1016/j.jhydrol.2013.12.005

Böhner, J., Antonic, O., 2009. Land-surface paremeters specific to topo-climatology, in: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 195–226.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239-240, 68–83. doi:10.1016/j.geoderma.2014.09.019

Camargo, L.A., Marques, J., Pereira, G.T., Alleoni, L.R.F., 2012. Spatial correlation between the composition of the clay fraction and contents of available phosphorus of an Oxisol at hillslope scale. Catena 100, 100–106. doi:10.1016/j.catena.2012.07.016

Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., Fealy, R., 2013. Are fine resolution digital elevation models always the best choice in digital soil mapping ? Geoderma 195-196, 111–121. doi:10.1016/j.geoderma.2012.11.020

Cheng, Y., Li, P., Xu, G., Li, Z., Cheng, S., Gao, H., 2016. Spatial distribution of soil total phosphorus in Yingwugou watershed of the Dan River, China. Catena 136, 175–181. doi:10.1016/j.catena.2015.02.015

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. doi:10.5194/gmd-8-1991-2015

DEV, 2002. German Standard Methods for the Examination of Water, Wastewater and Sludge. WILEY-VCH, Berlin.

Erskine, R.H., Green, T.R., Ramirez, J. a., MacDonald, L.H., 2007. Digital Elevation Accuracy and Grid Cell Size: Effects on Estimated Terrain Attributes. Soil Sci. Soc. Am. J. 71, 1371. doi:10.2136/sssaj2005.0142

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

Grunwald, S., 2006. What do we really know about the space-time continuum of soil-landscapes, in: Grunwald, S. (Ed.), Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics. CRC Press, New York, pp. 3–36.

Hedley, M.J., Stewart, J.W.B., Chauhan, B.S., 1982. Changes in Inorganic and Organic Soil Phosphorus Fractions Induced by Cultivation Practices and by Laboratory Incubations1. Soil Sci. Soc. Am. J. 46, 970–976. doi:10.2136/sssaj1982.03615995004600050017x

Huang, W., Spohn, M., 2015. Effects of long-term litter manipulation on soil carbon, nitrogen, and phosphorus in a temperate deciduous forest. Soil Biol. Biochem. 83, 12–18. doi:10.1016/j.soilbio.2015.01.011

Huggett, R.J., 1975. Soil landscape systems: a model of soil genesis. Geoderma 13, 1–22.

Johnson, J.W., 2000. A heuristic method for estimating the relative weight of predictor variables in multiple regression. Multivariate Behav. Res. 35, 1–19. doi:10.1207/S15327906MBR3501_1

Kabacoff, R.I., 2015. R in action: Data analysis and graphics with R. Minning, Shelter Island.

Kim, D., Zheng, Y., 2011. Scale-dependent predictability of DEM-based landform attributes for soil spatial variability in a coastal dune system. Geoderma 164, 181–194. doi:10.1016/j.geoderma.2011.06.002

Korea Institute of Geology Mining and Material, 2001. Explanatory note of the Gangreung Sokcho sheet 1:250,000. Korea Institute of Geology, Mining and Material, Daejeon.

Korea meteorological administration, 2015. Korea weather survice. http://www.kma.go.kr/ (accessed 18.3.2016).

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer, New York.

Ließ, M., Hitziger, M., Huwe, B., 2014. The Sloping Mire Soil-Landscape of Southern Ecuador : Influence of Predictor Resolution and Model Tuning on Random Forest Predictions. Appl. Environ. Soil Sci. 2014. doi:10.1155/2014/603132

MacMillan, R.A., Shary, P.A., 2009. Landforms and landform elements in geomorphometry, in: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 227–254.

Mage, S.M., Porder, S., 2013. Parent Material and Topography Determine Soil Phosphorus Status in the Luquillo Mountains of Puerto Rico. Ecosystems 16, 284–294. doi:10.1007/s10021-012-9612-5

Maynard, J.J., Johnson, M.G., 2014. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. Geoderma 230-231, 29–40. doi:10.1016/j.geoderma.2014.03.021

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping, Geoderma. doi:10.1016/S0016-7061(03)00223-4

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94. doi:10.1016/S0016-7061(98)00137-2

Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015. Impact of multi-scale predictor selection for modeling soil properties. Geoderma 239-240, 97–106. doi:10.1016/j.geoderma.2014.09.018

Minasny, B., McBratney, A.B., 2015. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. doi:10.1016/j.geoderma.2015.07.017

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil Attribute Prediction Using Terrain Analysis. Soil Sci. Soc. Am. J. 57, 443–452.

National Academy of Agricultural Science, 2013. Korean Soil Information System. http://soil.rda.go.kr/soil/index.jsp (accessed 12.3.2016).

National Geographic Information Institute, 2015. National Spatial Information Clearinghouse. https://www.nsic.go.kr (accessed 21.3.2016).

Park, S.J., Burt, T.P., 2002. Identification and Characterization of Pedogeomorphological Processes on a Hillslope. Soil Sci. Soc. Am. J. 66, 1897–1910.

Park, S.J., McSweeney, K.K., Lowery, B.B., 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103, 249–272. doi:10.1016/S0016-7061(01)00042-8

Park, S.J., Rüecker, G.R., Agyare, W.A., Akramhanov, A., Kim, D., Vlek, P.L.G., 2009. Influence of Grid Cell Size and Flow Routing Algorithm on Soil-Landform Modeling. J. Korean Geogr. Soc. 44, 122–145.

Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma 209-210, 1–14. doi:10.1016/j.geoderma.2013.05.029

Richardson, A.E., Barea, J.M., McNeill, A.M., Prigent-Combaret, C., 2009. Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. Plant Soil 321, 305–339. doi:10.1007/s11104-009-9895-2

Roberts, T.L., Stewart, J.W.B., Bettany, J.R., 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.

Roger, A., Libohova, Z., Rossier, N., Joost, S., Maltas, A., Frossard, E., Sinaj, S., 2014. Spatial variability of soil phosphorus in the Fribourg canton, Switzerland. Geoderma 217-218, 26–36. doi:10.1016/j.geoderma.2013.11.001

Roudier, P., Beaudette, D.E., Hewitt, A.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints, in: Minasny, B., Malone, B., McBratney, A. (Eds.), Digital Soil Assessments and beyond. CRC Press, Boca Raton, pp. 227–231.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. 27, 171–197. doi:10.1191/0309133303pp366ra

Siegel, S., Castellan, N.J., 1988. Nonparametric statistics for the behavioral sciences. McGraw-hill, New York.

Smeck, N.E., 1985. Phosphorus dynamics in soils and landscapes. Geoderma 36, 185–199.

Strahler, A., 1957. Quantitative analysis of watershed geomorphology. Trans. Am. Geophys. Union 38, 913–920. doi:10.1130/0016-7606

Tiessen, A.H., Chacon, P., Cuevas, E., 1994. Phosphorus and Nitrogen Status in Soils and Vegetation along a Toposequence of Dystrophic Rainforests on the Upper Rio Negro. Oecologia 99, 145–150.

Tucker, C.J., Sellers, P.J., 1986. Satellite remote sensing of primary production. Int. J. Remote Sens. 7, 1395–1416. doi:10.1080/01431168608948944

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer.

Vincent, A.G., Sundqvist, M.K., Wardle, D. a, Giesler, R., 2014. Bioavailable soil phosphorus decreases with increasing elevation in a subarctic tundra landscape. PLoS One 9, e92942. doi:10.1371/journal.pone.0092942

Vitousek, P., Chadwick, O., Matson, P., Allison, S., Derry, L., Kettley, L., Luers, A., Mecking, E., Monastra, V., Porder, S., 2003. Erosion and the Rejuvenation of Weathering-derived Nutrient Supply in an Old Tropical Landscape. Ecosystems 6, 762–772. doi:10.1007/s10021-003-0199-8

Walker, T.W., Syers, J.K., 1976. The fate of phosphorus during pedogenesis. Geoderma 15, 1–19. doi:10.1016/0016-7061(76)90066-5

Wohl, E., 2010. Mountain rivers. American Geophysical Union, Washington, DC.

Wood, J., 2009. Geomorphometry in LandSerf, in: Hengl, T., Reuter, H.I. (Eds.), Geomorphometry: Concepts, Software, Applications. Elsevier, Amsterdam, pp. 333–349.

Yang, X., Post, W.M., 2011. Phosphorus transformations as a function of pedogenesis: A synthesis of soil phosphorus data using Hedley fractionation method. Biogeosciences 8, 2907–2916. doi:10.5194/bg-8-2907-2011

Yoo, K., Mudd, S.M., Sanderman, J., Amundson, R., Blum, A., 2009. Spatial patterns and controls of soil chemical weathering rates along a transient hillslope. Earth Planet. Sci. Lett. 288, 184–193. doi:10.1016/j.epsl.2009.09.021

Zevenbergen, L.W., Zevenbergen, L.W., Thorne, C.R., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12, 47–56.

Appendix 3.1 Statistical summary of phosphorus fractions.

	Mean	SD	MIN	Median	Max	CoV (%)	Skew	Kurt
Total P	389.07	171.04	160.00	330.00	920.00	43.96	1.40	4.52
Resin Pi	3.42	1.63	0.40	3.10	9.00	47.80	1.02	4.32
Available P	35.71	17.03	8.40	32.50	83.50	47.69	0.61	-0.33
Organic Po	158.00	85.55	41.40	129.20	459.70	54.15	1.02	0.58
Apatite Pi	9.07	7.91	2.20	6.20	64.40	87.15	4.15	27.71
Residual P	144.39	41.33	80.70	136.80	285.70	28.63	1.15	4.51

SD = standard deviation, MIN= minimum, MAX= maximum, CoV= coefficient of variation, Skew= Skewness, Kurt= Kurtosis
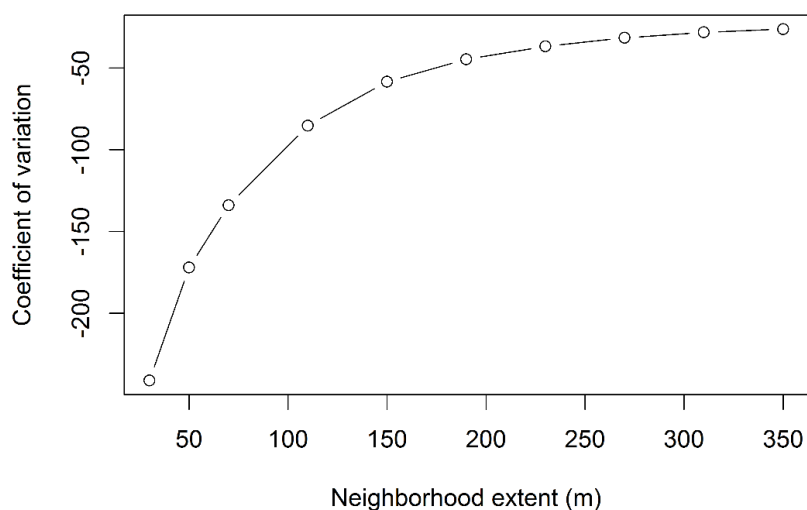
Appendix 3.2 Coefficient of variations of surface curvature with varying the neighbourhood extent.
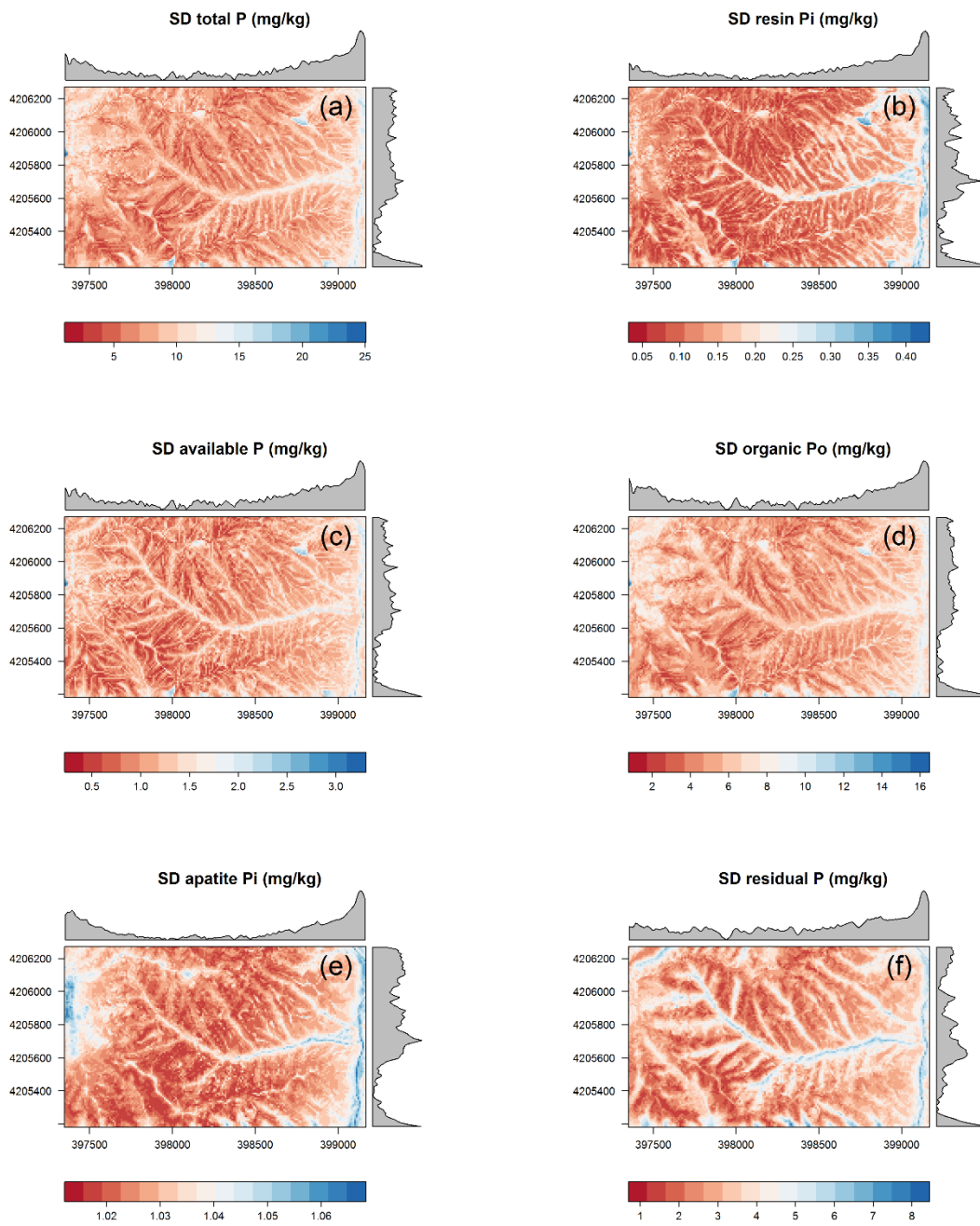
Appendix 3.3 Predicted SD phosphorus fractions with summaries of the column and row. SD: standard deviation.

Chapter 4 Spatial topsoil N:P ratios under monsoon conditions in a complex terrain of South Korea

Gwanyong Jeong[1], Kwanghun Choi[2], Marie Spohn[3], Soo Jin Park[4], Bernd Huwe[1], Mareike Ließ[5]

[1]*Department of Geosciences/ Soil Physics Division, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitaetsstrasse 30, 95447 Bayreuth, Germany*

[2]*Biogeographical Modelling, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitätsstraße 30, D-95440 Bayreuth, Germany*

[3]*Department of Soil Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Dr.-Hans-Frisch-Straße 1-3, Bayreuth 95448, Germany*

[4]*Department of Geography, Seoul National University, Shilim-Dong, San 56-1, Kwanak-Gu, Seoul 151-742, South Korea*

[5] *Department of Soil Physics, Helmholtz Centre for Environmental Research - UFZ, Theodor-Lieser-Strasse 4, D-06120 Halle, Germany*

Corresponding author: gwanyong.jeong@uni-bayreuth.de

Abstract

Spatial patterns of nitrogen (N) and phosphorus (P) in topsoils are critical for plant nutrition. Relatively little is known about spatial patterns of N and P in the organic layer in mountainous landscapes. Therefore, digital soil maps of N and P in both the organic layer and the A horizon were developed using a LiDAR (light detection and ranging) DEM (digital elevation model) and LiDAR derived vegetation parameters. Soil samples were collected in a small watershed covered by forest in South Korea. LiDAR metrics, the normalized difference vegetation index (NDVI), and terrain parameters were derived as predictors. Spatial explicit predictions of N to P ratios were done using random forest with uncertainty analysis.

Surface curvature showed the highest predictor importance for P contents in organic layers and A horizons, and LiDAR vegetation predictors and NDVI were important predictors for N in the organic layer. N:P ratios increased with surface curvature and were elevated at the convex upper slope compared to the concave lower slope. This was due to P enrichment of the soil at the lower slope and a more even distribution of N. Our digital soil maps showed that the topsoils at the upper slopes contained relatively little P. The findings are critical to understand dynamics of N and P in mountainous ecosystems with monsoon climate and steep slopes.

Keywords: digital soil mapping, random forest, LiDAR, organic layer, nitrogen, phosphorus, complex terrain, South Korea.

4.1 Introduction

Nitrogen (N) and phosphorus (P) are the most limiting nutrients for the primary productivity in terrestrial ecosystems (Vitousek et al., 2010, 2002). Soil nutrient contents vary during long-term soil development, in the way that N increase, while P declines during the course of pedogenesis. This is because N enters the ecosystem via N–fixing microorganisms, whereas P is derived from the weathering of phosphate minerals. As a result, primary productivity is initially N-limited in lightly weathered soils but increasingly P-limited in highly weathered soils over millions of years (Laliberté et al., 2013).

P limitation might be further enhanced by atmospheric N deposition during short-term periods (years to decades) (Braun et al., 2010; Vitousek et al., 2010). Furthermore, enormous increases in N depositions were also found in East Asia where population and economy are growing rapidly (Manning, 2012). In Korea, atmospheric N inputs have rapidly increased due to huge industrial operations and intense agricultural activities (Jang et al., 2011; Il-nam Kim et al., 2014; Kim et al., 2011). The annual average wet input of N ranged 12.9 - 24.9 kg ha^{-1}year^{-1} from 2005 to 2010 at 33 sites in Korea (Jang et al., 2011). The increase in N

concentrations was caused by atmospheric N deposition (Kim et al., 2011). This might have effects on the productivity, biodiversity, and community composition of plants (Turner, 2008).

An understanding of nutrient contents in the organic layer is critical for mountainous ecosystem management. Organic layers are made up of relatively freshly fallen organic matter, including whole leaves, twigs, and fruits. Nutrients that are returned to soil by litterfall are important for plant nutrition (Huang and Spohn, 2015). The N:P ratio in topsoil is used as indicators of potential growth limitation (Cleveland and Liptzin, 2007). The investigation of the spatial patterns of nutrients in the organic layer and the A horizon can be a useful tool to gain insight into soil-vegetation linkages if nutrients in the organic layer are outputs from vegetation (e.g. Uriarte et al. (2015)).

Many researchers tried to assess spatial patterns of soil N (Kunkel et al., 2011; Liu et al., 2013; Peng et al., 2013) and P (J. Kim et al., 2014; McKenzie and Ryan, 1999; Roger et al., 2014). Little is known about the spatial pattern of N:P ratios in the organic layer that are caused by environmental factors such as topography and vegetation. Previous studies found environmental correlations between N contents in the O layer and topographic parameters in a temperate forested watershed (Johnson et al., 2000) and in boreal forests (Seibert et al., 2007). Wilcke et al. (2008) reported an elevation gradient of decreasing contents of N and P in organic layers, while Soethe et al. (2008) found that the N stocks of the organic layer differ significantly with elevation in tropical mountainous forests. Our understanding of quantitative relationships between nutrient contents (especially P) in the organic layer and environmental factors is still limited.

The factors of soil formation are generally used as predictors for regression-based digital soil mapping (McBratney et al., 2003; Scull et al., 2003). For spatial predictions, most studies used various topographical predictors derived from digital elevation models (DEM) such as elevation, slope angle, curvature, and wetness index. According to Ballabio (2009), maps of soil

properties can be produced with an good accuracy by only using terrain predictors in mountainous areas. On the other hand, the spatial vegetation pattern is related to the spatial heterogeneity of soils (Binkley and Fisher, 2012). Soil development is influenced by vegetation (Ballabio et al., 2012). Therefore, various vegetation predictors derived from satellite images have contributed to explaining the spatial variability of soil nutrients (Grunwald et al., 2015; Mulder et al., 2011). However, to our knowledge, no attempt to use LiDAR (Light detection and ranging) derived vegetation parameters for the spatial predictions of soil properties has been made yet.

The vegetation LiDAR predictors could allow to extend our understanding of spatial soil data with better insight into the relationship between soils and vegetation. The LiDAR-derived vegetation predictors are related to the vertical variability of the vegetation within each cell which reflects forest structure metrics (Jones and Vaughan, 2010). Canopy cover percentage and maximum height can indicate the above ground biomass and the forest productivity (Zellweger et al., 2015). LiDAR predictors also can play a role as an ecological indicator such as light condition in the forest floor (Zellweger et al., 2015). LiDAR intensity varies with land cover and forest types (Ørka et al., 2009). Additionally, LiDAR predictors are high resolution data which have detailed spatial information compared to other remote sensing data (e.g. Spot (15 m) or Landsat images (30 m)). It is expected that NDVI and LiDAR data are important for N predictions related to forest biomass, but probably not for P since P is mainly originating from bedrock.

In order to understand the spatial patterns of N and P in the organic layer and mineral soil, digital soil maps were developed using LiDAR DEM (digital elevation model) and vegetation parameters as predictors. The specific aims of our research were (1) to test the importance of LiDAR-derived vegetation and topographical parameters to understand the spatial N and P patterns, (2) to identify subareas with critical P contents, and (3) to test different validation strategies for N and P depending spatial uncertainty structures.

4.2 Materials and methods

4.2.1 Research area

The study area is 9.84 km^2 and located in the downstream area of the Soyang watershed, Gangwon-do province, South Korea (Figure 4.1). Soyang lake watershed is a major drinking water source for Seoul. The mean annual temperature is 11.1 °C and a mean annual rainfall

of 1,347 mm (Korea meteorological administration, 2015). About 70 % of the annual rain (824.4 mm) falls heavily in the summer monsoon season (June, July and August) (Korea meteorological administration, 2015). The area's geology is Gyeonggi gneiss complex which consists of granitic gneiss and banded gneiss (metasedimentary rock) (Korea Institute of Geology Mining and Material, 2001). The parent materials formed in Paleoproterozoic and is oldest basement rocks in the Korean Peninsula (Chough, 2013). The elevation ranges between 320 and 868 m above sea level and the area consists of various steep slopes (over 45°) because of tectonic uplift during the Quaternary (Lee, 2004).
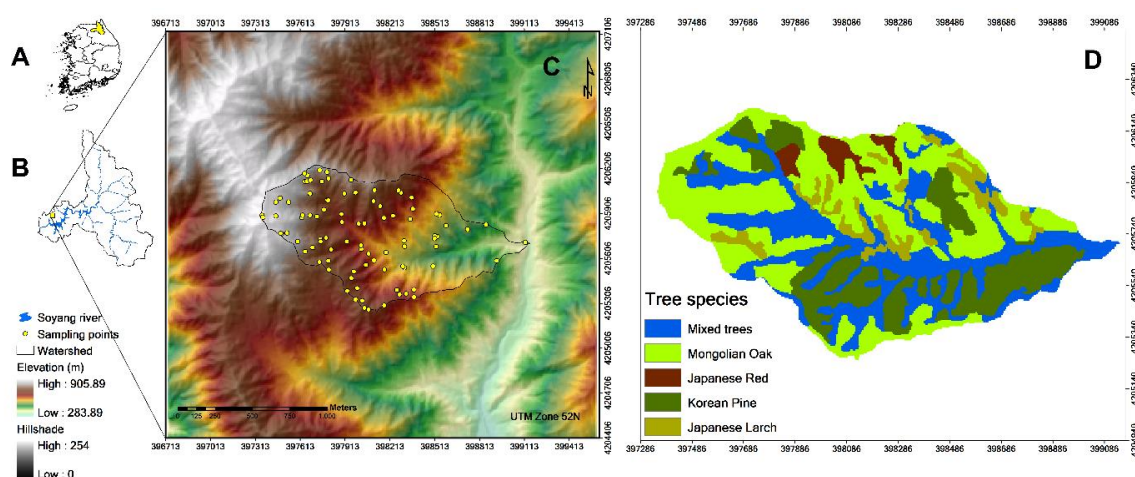

Figure 4.1 Research area. (A) The Soyang watershed within South Korea. (B) The research area within the Soyang watershed. (C) The research area with the sampling points. (D) The tree species map (fgis.forest.go.kr/).

It is in a headwater catchment that has narrow depositional areas and valley and plays important roles for biogeochemical cycles and downstream hydrological systems due to a key source of nutrients (Wohl, 2010). Its soils are mainly composed of fine gravelly sandy loam soils and fine sandy loam and gravelly loam soils (National Academy of Agricultural Science, 2013). It is a national forest and main tree species are Mongolian oak (*Quercus mongolica*) (40 - 50 years) and Korean pine (*Pinus koraiensis*) (30 - 35 years), locally vegetated with Japanese red pine (*Pinus densiflora*) and Japanese larch (*Larix kaempferi*) (Figure 4.1D).

4.2.2 Soil sampling and chemical analyses

A total number of 91 soil samples was collected from the organic layer and the A horizon in 2014. Conditioned Latin Hypercube Sampling (cLHS) was applied to optimize the density functions of the covariate space for the regression models (Minasny and McBratney, 2006). cLHS is a kind of stratified random sampling that divides empirical density functions based on the number of samples, draws one random sample within each interval, and then matches random samples across multiple predictors with conditioning to draw samples in the real world. cLHS was performed within R package "clhs" (Roudier et al., 2012). Spatial position information of sampling points was recorded with a Qmini H3 GNSS (global navigation satellite system) GPS (accuracy is within 5m) during field work.

The samples of the organic layer were extracted by a metal frame of 0.3 m x 0.3 m. Mineral soil samples were air-dried and sieved (< 2 mm). The organic layer samples were oven-dried.

Total P was extracted with HNO_3 and HF and measured according to DIN EN ISO 11885 / 22036 (DEV, 2002) by ICP-OES (Perkin Elmer, 2100 ZL, USA). After grinding to fine powder, total N was measured by an elemental analyser NA 1108 (CE Instruments, Milano, Italy). N:P ratios were calculated based on mass.

4.2.3 Environmental predictors

LiDAR is a remote sensing technology which has structural information on the illuminated surface, including 3D terrain, vegetation canopy information, and object heights (Franklin, 2010). LiDAR provides massive amounts of point data which include x, y, and z coordinates of the objects and can be changed into digital terrain model and digital surface model (Hyyppä et al., 2008). The laser emits short pulses of light and the sensors records echoes from leaves, branches, and underlying soil. The initial reflections are from the canopy top, next reflections come from lower canopy, and finally the last is from the ground (Jones and Vaughan, 2010). By separating the above-ground reflections, the ground reflections will enable to produce models of the terrain and finally vegetation using differences between above-ground and ground heights (Jones and Vaughan, 2010). With height data, LiDAR provides intensity data that reflect characteristics of objects and can be useful information on forest types and tree species (Ørka et al., 2009).

We used LiDAR point data which had a vertical accuracy of below 10 cm and an average of 4.08 points/ m^2, and surveyed by the National Geographic Information Institute (NGII) in South Korea (National Geographic Information Institute, 2015). The point data were pre-processed to identify ground returns, classify all returns and calculate the normalized vegetation heights. We calculated a set of forest structural predictors using LAStools program which provides a wide variety of processing methods on LiDAR data (Isenburg, 2014) (Table 4.1). The LiDAR predictors were considered to indicate important ecological factors such as light conditions close to the forest floor and the variation and distribution of vegetation along the vertical profile. The normalized difference vegetation index (NDVI) was derived from a 4 m Kompsat-2 image on 11th October 2014.

Most topographical predictors were calculated with the terrain analysis modules of the open source software SAGA based on the LiDAR DEM (Conrad et al., 2015). In addition, surface

Table 4.1 Environmental predictors for digital soil mapping

	Predictor	Method	Reference
1	Elevation (ELEV)	-	-
2	Slope degree (SLO)	Slope, aspect, curvature SAGA module	Zevenbergen et al. (1987)
3	Catchment area (CA)	Catchment area (Parallel) SAGA module	Costa-Cabral and Burges (1994)
4	SAGA topographical wetness index (STWI)	SAGA wetness index SAGA module	Böhner et al. (2002)
5	Surface curvature (CUR19)	CURV3 program	Park et al. (2001)
6	Normalized difference vegetation index (NDVI)	(NIR – Red)/ (NIR+Red)	Tucker and Sellers (1986)
7	Maximum height (Hmax)	Lascanopy LAStools module	Isenburg (2014)
8	Canopy cover percentage (Hccp)	Lascanopy LAStools module	Isenburg (2014)
9	Standard deviations of heights (Hstd)	Lascanopy LAStools module	Isenburg (2014)
10	Forest canopy and height (Hch)	Canopy cover percentage (Hccp) x maximum height (Hmax)	-
11	First return intensity average (Hfiravg)	Lasgrid LAStools module	Isenburg (2014)

NIR= near-infrared

curvature that reflects the degree of bending three-dimensional surface morphology was calculated with the CURV3 program (Park et al., 2001). Surface curvature (CUR19) was measured within a 19 x 19 local window at each cell based on the best results of the correlation analyses. All predictors were changed into 10 m cell size via the nearest neighbor resampling method.

4.2.4 Random forest and ANOVA analysis

Random forest (RF) is an ensemble learning method which operates by building a set of regression trees averaging the results (Breiman, 2001). Each tree is built by bootstrap samples of the data and a subset of predictors. As long as the number of trees is large, RF has low bias and low variance of prediction (Breiman, 2001). Therefore, the number of trees was set to 1000. The size of the predictor subset (mtry) was tuned by R package "caret" (Kuhn and Johnson, 2013). RF was performed within R package "randomForest" (Breiman, 2001).

RF is able to model complex nonlinear relationships between soil properties and environmental predictors. Compared to other supervised learning methods (e.g. neural networks and support vector regression), RF is easier to apply, because it does not require much tuning (Kuhn and Johnson, 2013; Strobl et al., 2009). Although RF does not show the structure of each tree, it has better interpretability because it provides a predictor importance measure. For this measure, predictor values are permuted. The importance is determined by the difference of mean square error before and after permuting (Strobl et al., 2009). RF showed good performance in digital soil mapping (Grimm et al., 2008; J. Kim et al., 2014; Ließ et al., 2012; Tesfa et al., 2009; Wiesmeier et al., 2011). Kampichler et al. (2010) recommended the use of RF because of its modelling performance as well as the ease of model construction and interpretability.

Predictor selection influences model performance (Brungard et al., 2015; Miller et al., 2015; Poggio et al., 2013). Recursive feature elimination (RFE), a backward predictor selection method, starts with all predictors and iteratively eliminates the least important predictor one by one based on an initial RF predictor importance measure until the best predictor remains (Kuhn and Johnson, 2013). At the end, the best number of predictors and the final list of selected predictors are returned. The package "caret" provides the functions for RFE (Kuhn and Johnson, 2013).

To assess the model performance, R² and RMSE (root mean square error) was calculated. For model validation, we used k-fold cross-validation (CV) where the sample is partitioned randomly into k subsets; one subset is left out for model validation while the remaining subsets are used for model training; the process is repeated k times (once for each fold) and the k estimates of performance are summarized. In the k-fold CV, the choice of k choice was based on the partitioning in test and validation sets. In the case of 10-fold CV, for example, 10% data set are used for the validation and the remaining 90% data set are used for the calibration. The choice of k is usually 5 or 10, but there is no formal rule (Kuhn and Johnson, 2013). Although the subsets are generated randomly, the subdivision will still have an impact on model validation results. This can be acknowledged by repetitions of the n-fold CV. The number of repetitions (n) might also affect the estimation of model performance: The more repetitions, the better the results (Molinaro et al., 2005). We explored 2, 5, 10, 20-fold and leave-one-out CV in n repetitions to account for a total of 100 validation measures: $n \times k = 100$. At the end, 100 R–Squares and RMSEs were returned for each soil property. The standard deviation of the corresponding 100 predictions give a spatial uncertainty estimate.

One-way analysis of variance (ANOVA) was used to test for differences in N and P contents in each topographical position. The assumptions for ANOVA were tested based on Shapiro test for normality and Bartlett test for homogeneity of variances. Kruskal-Wallis test was applied in case that the assumptions were not met. R package "pgirmess" was used for Kruskal-Wallis test (Siegel and Castellan, 1988).

4.3 Results

4.3.1 Descriptive statistics of soil nutrients

The mean total N in the O layer (No) was the highest value (12245 ± 1986 mg kg^{-1}). No had the lowest variance, while total P in the O horizons (Po) showed a relatively higher variance based on values of standard deviation and coefficient of variation (CoV) (Table 4.2). Therefore,

Table 4.2 Statistical summary of N and P contents (mg kg^{-1}) and ratios.

	Mean	SD	MIN	Median	Max	CoV (%)	Skew	Kurt
No	12245	1986	8000	12200	17800	16.22	0.35	2.92
Po	624	190	310	610	1240	30.39	0.44	2.97
Na	2990	1348	700	2600	7300	45.07	0.81	3.52
Pa	389	171	160	330	920	43.96	1.40	4.52
No/Po	20.83	4.82	12.16	20.17	38.06	23.12	0.76	3.77
Na/Pa	7.91	2.42	1.89	7.78	13.85	30.55	0.21	3.06

SD = standard deviation, MIN= minimum, MAX= maximum, CoV= coefficient of variation, Skew= Skewness, Kurt= Kurtosis, N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.

the variability of N:P ratios in the O horizons (No/Po) depended on Po contents. The total N and P ratios in the A horizons (Na/Pa) showed higher variances compared to the O horizons. The mean No/Po was 20.83 ± 4.82 and the mean Na/Pa was 7.91 ± 2.42.

4.3.2 Predictors and models

For the choice of k, we selected 10 repeated 10-fold CV. This CV showed the relatively good result of R-square and lower mean RMSE and the lower uncertainty for all soil nutrients (Figure 4.2 and 4.3). Especially, spatial patterns of mean Po were not changed based on CV strategies, while uncertainty decreased prominently with increasing portions of calibration dataset (Appendix 4.2). The values of CoV (coefficient of variation) decreased sharply in the lower elevation areas with higher uncertainties based on CV strategies (Appendix 4.2b and h).

RFE found the optimal combination of predictors to improve prediction results (Table 4.3). Surface curvature and elevation were selected for all soil nutrients. Additionally, selected predictors for Po and Pa were curvature and elevation. Normalized difference vegetation index (NDVI) and LiDAR vegetation predictors (Hfiravg, Hstd, and Hmax) were selected for No.

Our RF model showed good performances for all soil nutrients based on R^2 (Figure 4.2). Mean R-Square values ranged from 0.23 to 0.52. Pa showed the best result of the validation while the R-square for Na/Pa was lowest. Models for P showed better results compared to N models.

Terrain predictors accounted for 5.37 – 54.54 % of the total variations of soil nutrients in our RF models. Surface curvature was one of the best predictors for soil nutrients with the exception of No (Figure 4.4). Surface curvature contributed 6.50 - 53.07 % of the total variations explained by RF models for each soil nutrient. Elevation also showed larger predictor
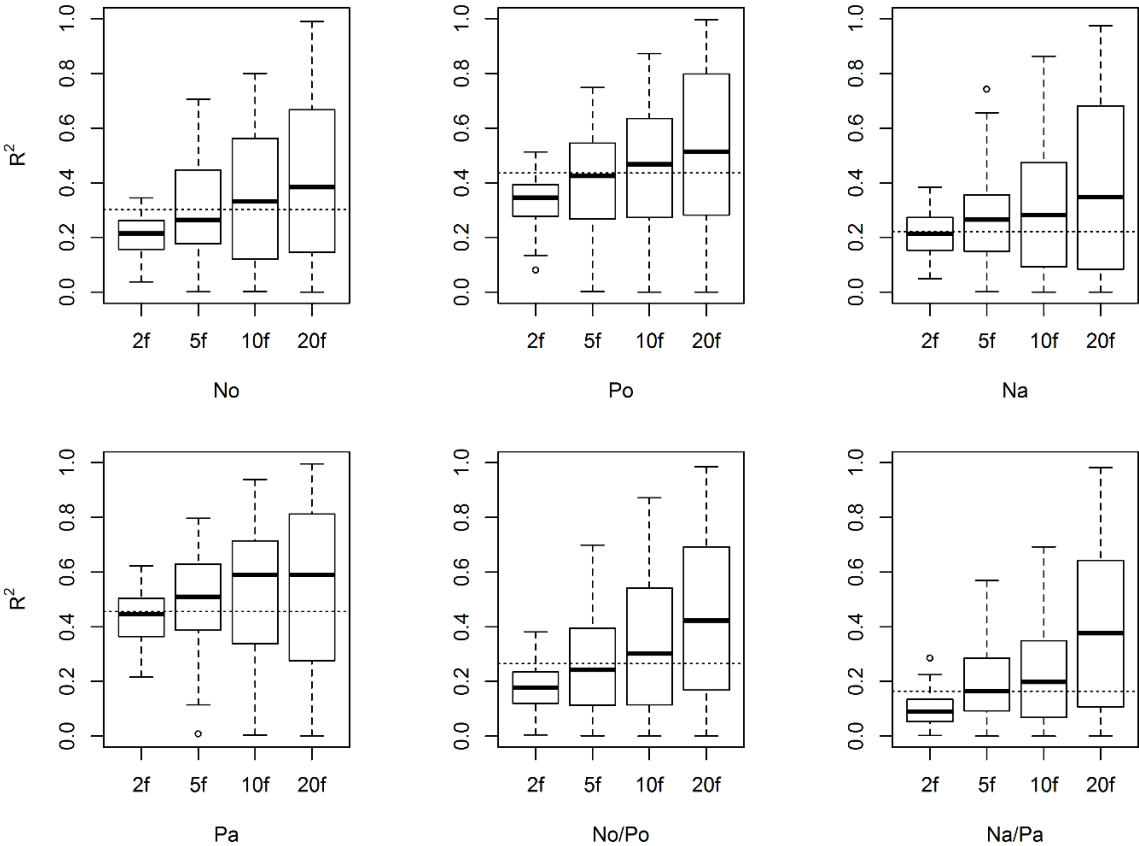

Figure 4.2 Model validation based on R-Square with cross validation methods. The dot lines refer to the leave-one-out cross-validated result. 2f: 2-fold 50 repetitions, 5f: 5-fold 20 repetitions, 10f: 10-fold 10 repetitions, 20f: 20-fold 5 repetitions, N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.
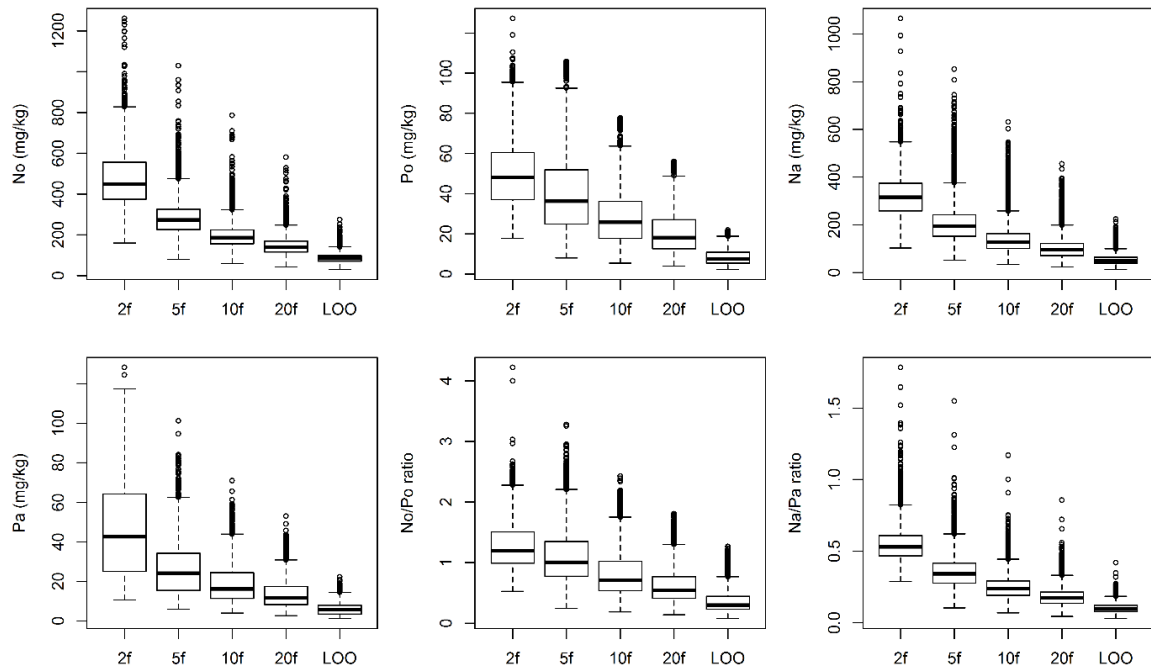
Figure 4.3 Boxplots of standard deviations of 100 predicted values for each raster cell with cross validation methods. 2f: 2-fold 50 repetitions, 5f: 5-fold 20 repetitions, 10f: 10-fold 10 repetitions, 20f: 20-fold 5 repetitions, LOO: leave-one-out, N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.

importance (9.55 – 39.22 %). NDVI and LiDAR-derived vegetation predictors (Hstd, Hmax, Hpdy, and Hfiravg) were also important for the nutrients. Results of RF predictor importance were not the same with RFE´s, but two results are quite similar and most important predictors were not different.

Table 4.3 Selected predictors using recursive feature elimination (RFE) based on 10-fold 10 repetitions.

Soil properties	Predictors
No	ELEV, NDVI, Hfiravg, CUR19, STWI, Hstd, Hmax
Po	CUR19, ELEV
Na	ELEV, CUR19
Pa	CUR19, ELEV
N/Po	CUR19, CA, Hstd, ELEV, Hmax, Hch
N/Pa	CUR19, CA, NDVI, ELEV, STWI

ELEV: elevation, CUR19: surface curvature (19 x 19 window), STWI: SAGA topographical wetness index, CA: Catchment area, SLO: slope degree, NDVI: normalized difference vegetation index, Hfiavg: First return intensity average, Hstd: Standard deviations of heights, Hmax: Maximum height, Hch: Forest canopy and height (Hmax X Hccp).
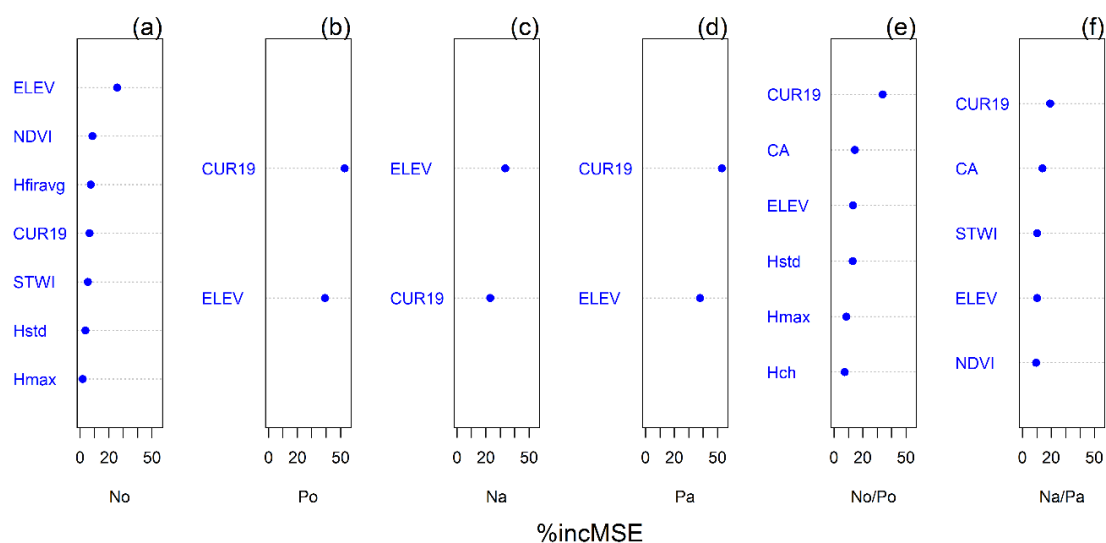

Figure 4.4 Mean relative importance of predictors for N and P based on the increased mean square error (%incMSE) from random forest. N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.

4.3.3 Environmental relationships and spatial patterns of nutrients

Differences in N and P were identified based on surface curvature which was the best predictor (Figure 4.4). Although the topographic classification criteria were rather arbitrary, surface curvature values were set at -1 and 1 to classify lower slope (< -1), mid slope (-1 to 1), and upper slope (> 1). No and Na didn't change depending on topographic positions, while Po and Pa were significantly different. The No/Po and Na/Pa were significantly different and widened from lower to upper slope positions (Figure 4.5). This implied that the significant differences were caused by P. The No/Po differed between the topographic positions combined with elevation (Appendix 4.3a), while Na/Pa did not show significant difference.
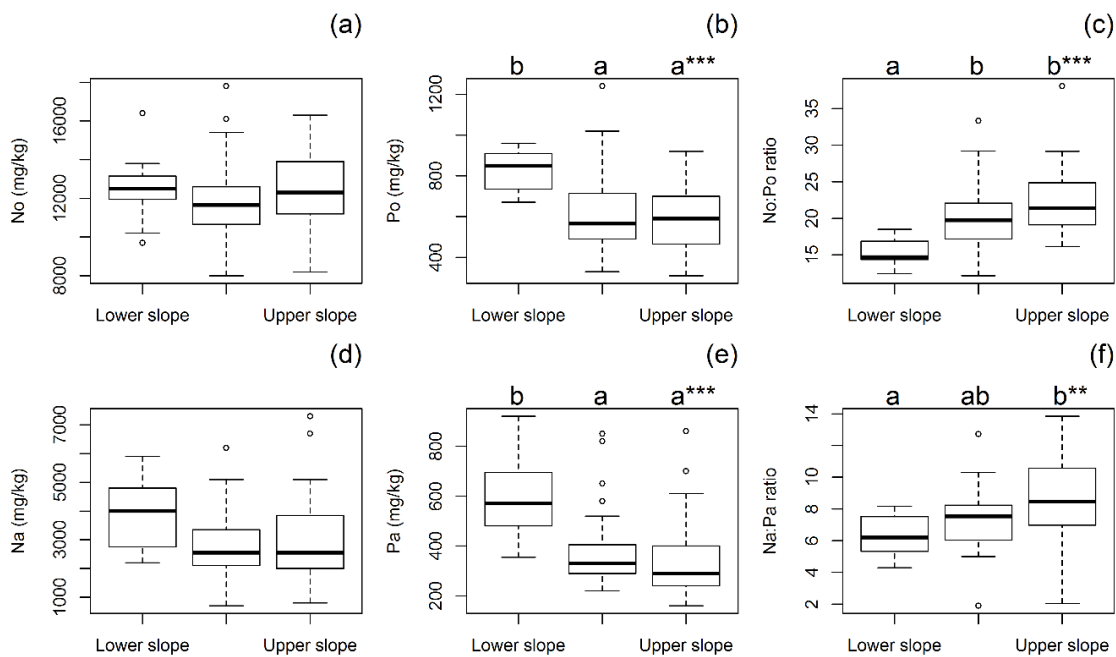

Figure 4.5 Box plots of nitrogen, phosphorus, and the ratios in the organic horizon and in the mineral soil A horizon based on topographical positions (a-f). Different letters indicate significant differences at $p < 0.05$ with Kruskal-Wallis test. ** $p < 0.01$, and *** $p < 0.001$. N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.

The map of each nutrient displays the mean of 100 predictions (Figure 4.6). Contents of No and Na increased with elevation. We found remarkably different contents of P between the upper slope and the lower slope. No/Po and Na/Pa showed higher values at convex upper slope.

Higher standard deviations of Po and No/Po were found at lower elevation and in the valley floor (Appendix 4.4). The spatial uncertainties of Na/Pa were high in the higher elevation (Appendix 4.4), while uncertainties of No and Na/Pa were complex and it is difficult to understand spatial patterns of uncertainty (Appendix 4.4).
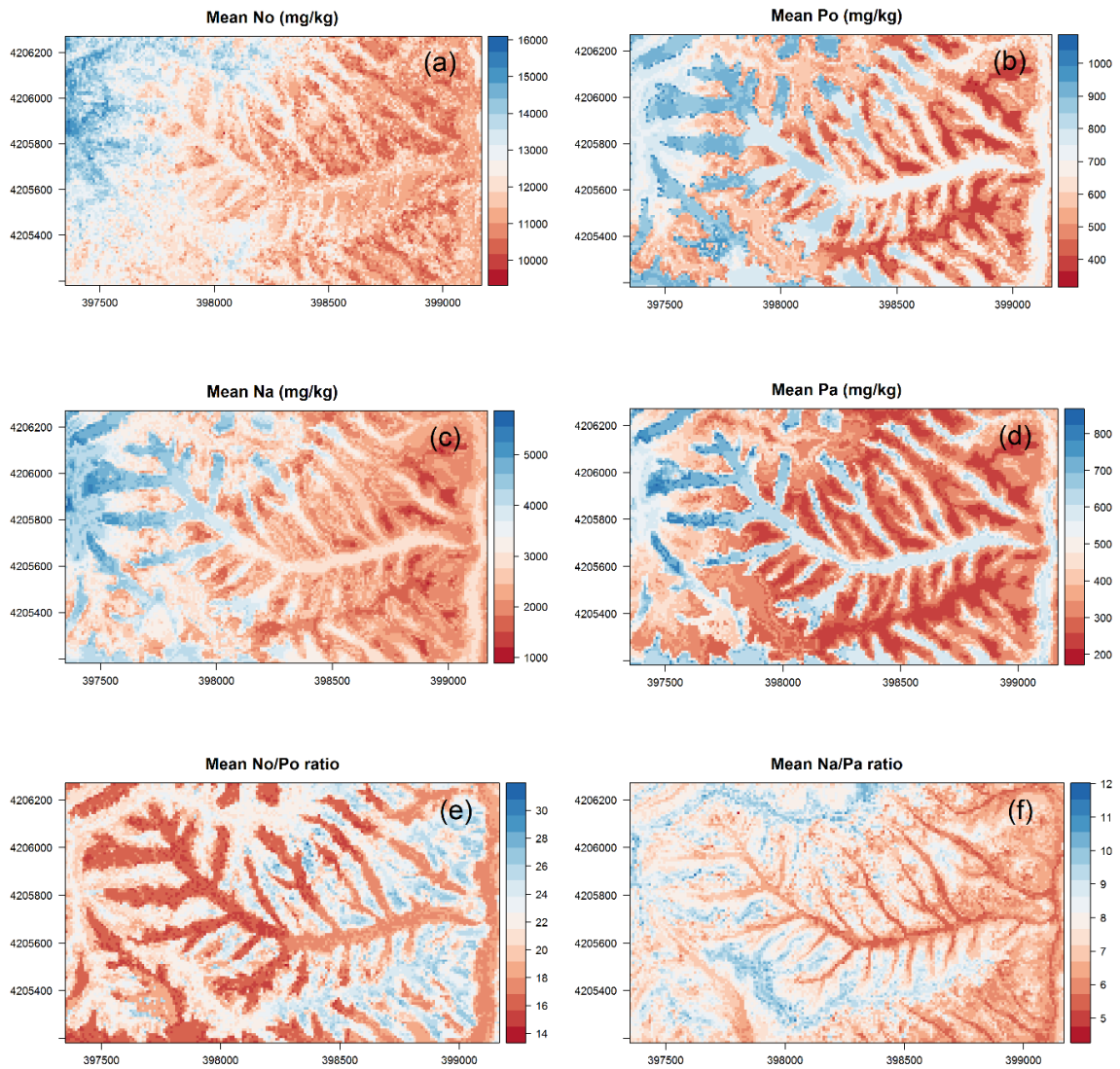
Figure 4.6 Predicted mean soil N and P contents and ratios. N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.

4.4 Discussion

4.4.1 Model performances based on different cross validation strategies

For small sample sizes, model calibration requires every possible dataset to improve the model result, while validation results can be different severely depending on which samples were included for the validation (Kuhn and Johnson, 2013). If we want to have the better trained model, portions of calibration dataset should be increased. However, portions of validation datasets should be increased to acquire a representative model to fit the underlying relationship (Remesan and Mathew, 2015). In our results, 20-fold CV showed the best result of R-square and RMSE and the lowest uncertainty of predictions (Figure 4.2 and 4.3), while 2-fold CV showed the lowest R-Square and RMSE and the highest mean value of the uncertainty. Variations of standard deviations of 100 predictions decreased with increasing the number of samples for the calibration (2-fold to 20-fold) (Figure 4.3). This implied that in case of small sample sizes, bigger calibration datasets improved the estimates and reduced the uncertainty of the spatial prediction. Park and Vlek (2002) tested the change of the prediction error with different numbers of training data sets and confirmed when more samples were used for tuning dataset, the prediction accuracy becomes larger. Similar decreasing changes of the prediction error were also found in various prediction methods (Ballabio, 2009). However, these studies did not find systematic changes and thus not include suggestions of optimal validation methods for soil predictions. In this study, uncertainty structures were explored and might be used as criteria for optimal validation methods in the future. Remesan and Mathew (2015) required to investigate the optimal number of the calibration datasets because using very few datasets might result in poorly calibrated model, while lots of data for calibration might cause to overfitting. So, they recommended 10-fold CV. Specifically, Kuhn and Johnson (2013) suggested repeated 10-fold CV for small sample sizes because the bias and variance are not high and the computational efficiency is good. In our results, repeated 10-fold CV is also

recommended to predict soil nutrients in the case of small sample sizes due to relatively high predictability and low uncertainty.

4.4.2 Important predictors of N and P

Topography was the most important factor for spatial patterns of N in our study area. Elevation might reflect soil temperature and the rate of decomposition which result in contents of soil N (Binkley and Fisher, 2012; Huggett, 1995; Osman, 2013). In the results, elevation has strongly positive relationships with No ($r=0.58$, $p<0.001$) and Na ($r=0.49$, $p<0.001$). Bedison and Johnson confirm the relationship with elevation and No ($R^2= 0.41$, $P<0.001$) in mountainous forested areas, USA. Additionally, positive relationships between Na and elevation were reported by Kunkel et al. (2011) and Wang et al. (2013). Peng et al. (2013) also showed that Na increased with elevation due to a low level of human disturbance. Catchment area (CA) and topographical wetness index (TWI) related to soil moisture were important predictors for No (Johnson et al., 2000; Seibert et al., 2007). Although Na showed a correlation with TWI ($r=0.26$, $p<0.05$), CA and TWI were not significant in our results. According to Aandahl (1948), higher nitrogen contents found at the lower slope. Na was not significantly related to topographic positions (Figure 4.5d), while surface curvature was selected as an important predictor for the Na model with elevation (Figure 4.4c). This is because higher Na was found in the areas of higher elevation and close to the lower slope (Figure 4.6c) which might have a higher productivity (plants and microbes) and therefore higher nitrogen fixation.

The spatial pattern of N is determined by vegetation (Bedison and Johnson, 2009; Zhang et al., 2010). Although NDVI and LiDAR predictors were not selected as predictors for Na, maximum height ($r=0.24$, $p<0.05$) and standard deviations of heights ($r=0.23$, $p<0.05$) were correlated with Na. Other studies found significant relationships between Na and NDVI which can measure vegetation density and aboveground biomass (Kim et al., 2014; Kunkel et al., 2011; Sumfleth and Duttmann, 2008). For No, NDVI ranked as the second important predictor

and the LiDAR intensity of first returns (Hfiravg), which is often used as an indicator of forest type (Ørka et al., 2009), was an important predictor. This implies that the density of forest cover and forest types have an effect on No contents and No/Po ratios. No relationship between P and LiDAR predictors was found.

The LiDAR-derived predictors are promising because of various advantages for DSM, as above mentioned. In future studies, the vegetation predictors should be applied to forest areas where the variation of forest cover is quite different. Forest structure (LiDAR metrics) can have an effect on erosion and deposition of materials which in turn might result in changed soil nutrient contents. Hahm et al. (2014) confirmed that differences in erosion rates were affected by tree canopy cover. However, there is no research that investigated the relationship between soil erosion, forest structures, and nutrient status using LiDAR data yet to our knowledge.

4.4.3 Spatial patterns of N:P ratios

Here we found that N:P ratios increased with surface curvature, and were elevated at the upper slope compared to the lower slope. This was due to P enrichment of the soil at the lower slope and a more even distribution of N (Figure 4.5). No/Po and Na/Pa were strongly related with surface curvature (Figure 4.4) and the topographical positions (Figure 4.5), which implies that P dynamics were affected strongly by topography. P carried from the upper slope was accumulated in the lower slope by surface and subsurface flows as previously reviewed by Smeck (1985). Soil erosion is strong in the watershed under study due to storm events and steep slopes (Jeong et al., 2012; Jung et al., 2012). As a consequences, the higher soil P contents at the lower slope compared to the upper slope can lead to a higher plant litter P contents, finally causing lower No/Po. Especially, No/Po compared with Na/Pa showed strong correlations with elevation, surface curvature and vegetation predictors (Figure 4.4 and Appendix 4.3), which means spatial patterns of No/Po might be caused by interconnected relationships with soil, topography and vegetation. Similarly, Uriarte et al. (2015) found that soil

N:P was correlated with leaf litter N:P and was determined by topography in a tropical mountainous forest with heavy rainfalls and steep slopes.

4.5 Conclusions

Here we created for the first time digital soil maps showing the spatial pattern of N:P ratios using terrain and LiDAR-derived vegetation predictors. In validation strategies, the repeated 10-fold CV method showed relatively good predictability and low uncertainty. Both No/Po and Na/Pa ratios related to nutrient conditions showed higher values at the upper slope. These maps help to identify areas with low nutrient availability. Our analyses showed that topographic characteristics may help to predict spatial patterns of nutrients in mountainous regions.

4.6 Acknowledgments

4.7 References

Aandahl, A., 1948. The characterisation of slope positions and their influence on the total nitrogen content of a few virgin soils in Western Iowa. Soil Sci. Soc. Am. J. 13, 449–454.

Ballabio, C., 2009. Spatial prediction of soil properties in temperate mountain regions using support vector regression. Geoderma 151, 338–350. doi:10.1016/j.geoderma.2009.04.022

Ballabio, C., Fava, F., Rosenmund, A., 2012. A plant ecology approach to digital soil mapping, improving the prediction of soil organic carbon content in alpine grasslands. Geoderma 187-188, 102–116. doi:10.1016/j.geoderma.2012.04.002

Bedison, J.E., Johnson, A.H., 2009. Controls on the Spatial Patterns of Carbon and Nitrogen in Adirondack Forest Soils along a Gradient of Nitrogen Deposition. Soil Sci. Soc. Am. J. 73, 2105. doi:10.2136/sssaj2008.0336

Binkley, D., Fisher, R., 2012. Ecology and management of forest soils. John Wiley & Sons, West Sussex.

Böhner, J., Köthe, R., Conrad, O., Gross, J., Ringeler, A., Selige, T., 2002. Soil regionalisation by means of terrain analysis and process parameterisation, in: Micheli, E., Nachtergaele, F., Montanarella, L. (Eds.), Soil Classification 2001. The European Soil Bureau, Joint Research Centre, Luxembourg, pp. 213–222.

Braun, S., Thomas, V.F.D., Quiring, R., Flückiger, W., 2010. Does nitrogen deposition increase forest production? The role of phosphorus. Environ. Pollut. 158, 2043–2052. doi:10.1016/j.envpol.2009.11.030

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239-240, 68–83. doi:10.1016/j.geoderma.2014.09.019

Chough, S.K., 2013. Geology and sedimentology of the korean peninsula. Elsevier, London.

Cleveland, C.C., Liptzin, D., 2007. C:N:P stoichiometry in soil: Is there a "Redfield ratio" for the microbial biomass? Biogeochemistry 85, 235–252. doi:10.1007/s10533-007-9132-0

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. doi:10.5194/gmd-8-1991-2015

Costa-Cabral, M.C., Burges, S.J., 1994. Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. Water Resour. Res 30, 1681–1692.

DEV, 2002. German Standard Methods for the Examination of Water, Wastewater and Sludge. WILEY-VCH, Berlin.

Franklin, S.E., 2010. Remote sensing for biodiversity and wildlife management: Synthesis and applications. McGrawHill, New York.

Grimm, R., Behrens, T., Märker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis. Geoderma 146, 102–113. doi:10.1016/j.geoderma.2008.05.008

Grunwald, S., Vasques, G.M., Rivero, R.G., 2015. Fusion of Soil and Remote Sensing Data to Model Soil Properties. Adv. Agron. 131, 1–109. doi:10.1016/bs.agron.2014.12.004

Hahm, W.J., Riebe, C.S., Lukens, C.E., Araki, S., 2014. Bedrock composition regulates mountain ecosystems and landscape evolution. Proc. Natl. Acad. Sci. doi:10.1073/pnas.1315667111

Huang, W., Spohn, M., 2015. Effects of long-term litter manipulation on soil carbon, nitrogen, and phosphorus in a temperate deciduous forest. Soil Biol. Biochem. 83, 12–18. doi:10.1016/j.soilbio.2015.01.011

Huggett, R.J., 1995. Geoecology. Routledge, New York.

Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. Int. J. Remote Sens. 29, 1339–1366. doi:10.1080/01431160701736489

Isenburg, M., 2014. LAStools - efficient tools for LiDAR processing, version 2.1. http://lastools.org (accessed 1.1.16).

Jang, S.-K., Sung, M.-Y., Shin, A.-Y., Choi, J.-S., Son, J.-S., Ahn, J.-Y., Kim, J.-C., Shin, E.-S., 2011. A Study for Long-term Trend of Acid Deposition in Korea. J. Korea Soc. Environ. Adm. 17, 183–192.

Jeong, J.J., Bartsch, S., Fleckenstein, J.H., Matzner, E., Tenhunen, J.D., Lee, S.D., Park, S.K., Park, J.H., 2012. Differential storm responses of dissolved and particulate organic

carbon in a mountainous headwater stream, investigated by high-frequency, in situ optical measurements. J. Geophys. Res. Biogeosciences 117, 1–13. doi:10.1029/2012JG001999

Johnson, C.E., Ruiz-Mendez, J.J., Lawrence, G.B., 2000. Forest Soil Chemistry and Terrain Attributes in a Catskills Watershed. Soil Sci. Soc. Am. J. 64, 1804–1814.

Jones, H.G., Vaughan, R.A., 2010. Remote sensing of vegetation: Principles, techniques, and applications. Oxford University Press, Oxford.

Jung, B.J., Lee, H.J., Jeong, J.J., Owen, J., Kim, B., Meusburger, K., Alewell, C., Gebauer, G., Shope, C., Park, J.H., 2012. Storm pulses and varying sources of hydrologic carbon export from a mountainous watershed. J. Hydrol. 440-441, 90–101. doi:10.1016/j.jhydrol.2012.03.030

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., Arriaga-Weiss, S., 2010. Classification in conservation biology: A comparison of five machine-learning methods. Ecol. Inform. 5, 441–450. doi:10.1016/j.ecoinf.2010.06.003

Kim, I., Lee, K., Gruber, N., Karl, D.M., Bullister, J.L., Yang, S., Kim, T., 2014. Increasing anthropogenic nitrogen in the North Pacific Ocean. Science 346, 1102–1106.

Kim, J., Grunwald, S., Rivero, R.G., 2014. Soil Phosphorus and Nitrogen Predictions Across Spatial Escalating Scales in an Aquatic Ecosystem Using Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 52, 6724–6737.

Kim, T.-W., Lee, K., Najjar, R.G., Jeong, H.-D., Jeong, H.J., 2011. Increasing N Abundance in the Northwestern Pacific Ocean Due to Atmospheric Nitrogen Deposition. Science 334, 505–509. doi:10.1126/science.1206583

Korea Institute of Geology Mining and Material, 2001. Explanatory note of the Gangreung Sokcho sheet 1:250,000. Korea Institute of Geology, Mining and Material, Daejeon.

Korea meteorological administration, 2015. Korea weather survice. http://www.kma.go.kr/ (accessed 3.18.16).

Kuhn, M., Johnson, K., 2013. Applied predictive modeling. Springer, New York.

Kunkel, M.L., Flores, A.N., Smith, T.J., McNamara, J.P., Benner, S.G., 2011. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. Geoderma 165, 1–11. doi:10.1016/j.geoderma.2011.06.011

Laliberté, E., Grace, J.B., Huston, M. a., Lambers, H., Teste, F.P., Turner, B.L., Wardle, D. a., 2013. How does pedogenesis drive plant diversity? Trends Ecol. Evol. 28, 331–340. doi:10.1016/j.tree.2013.02.008

Lee, G., 2004. Characteristics of Geomorphological Surface and Analysis of Deposits in Fluvial Terraces at Upper Reach of Soyang River. J. Korean Geogr. Soc. 39, 27–44.

Ließ, M., Glaser, B., Huwe, B., 2012. Uncertainty in the spatial prediction of soil texture. Comparison of regression tree and Random Forest models. Geoderma 170, 70–79. doi:10.1016/j.geoderma.2011.10.010

Liu, Z.-P., Shao, M.-A., Wang, Y.-Q., 2013. Spatial patterns of soil total nitrogen and soil total phosphorus across the entire Loess Plateau region of China. Geoderma 197-198, 67–78. doi:10.1016/j.geoderma.2012.12.011

Manning, P., 2012. The impact of nitrogen enrichment on ecosystems and their services, in: Wall, D.H. (Ed.), Soil Ecology and Ecosystem Services. Oxford University Press, pp. 256–269.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping, Geoderma. doi:10.1016/S0016-7061(03)00223-4

McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94. doi:10.1016/S0016-7061(98)00137-2

Miller, B.A., Koszinski, S., Wehrhan, M., Sommer, M., 2015. Impact of multi-scale predictor selection for modeling soil properties. Geoderma 239-240, 97–106. doi:10.1016/j.geoderma.2014.09.018

Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388. doi:10.1016/j.cageo.2005.12.009

Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction Error Estimation : A Comparison of Resampling Methods. Bioinformatics 21, 3301–3307.

Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1–19. doi:10.1016/j.geoderma.2010.12.018

National Academy of Agricultural Science, 2013. Korean Soil Information System. http://soil.rda.go.kr/soil/index.jsp (accessed 1.12.16).

National Geographic Information Institute, 2015. National Spatial Information Clearinghouse. https://www.nsic.go.kr (accessed 3.21.16).

Ørka, H.O., Næsset, E., Bollandsås, O.M., 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sens. Environ. 113, 1163–1174. doi:10.1016/j.rse.2009.02.002

Osman, K.T., 2013. Forest Soils: Properties and Management. Springer International Publishing Switzerland.

Park, S.J., McSweeney, K.K., Lowery, B.B., 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103, 249–272. doi:10.1016/S0016-7061(01)00042-8

Park, S.J., Vlek, P.L.G., 2002. Environmental correlation of three-dimensional soil spatial variability: A comparison of three adaptive techniques. Geoderma 109, 117–140. doi:10.1016/S0016-7061(02)00146-5

Peng, G., Bing, W., Guangpo, G., Guangcan, Z., 2013. Spatial distribution of soil organic carbon and total nitrogen based on GIS and geostatistics in a small watershed in a hilly area of northern China. PLoS One 8, 1–9. doi:10.1371/journal.pone.0083592

Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma 209-210, 1–14. doi:10.1016/j.geoderma.2013.05.029

Remesan, R., Mathew, J., 2015. Hydrological data driven modelling: A case study approach. Springer International Publishing Switzerland.

Roger, A., Libohova, Z., Rossier, N., Joost, S., Maltas, A., Frossard, E., Sinaj, S., 2014. Spatial variability of soil phosphorus in the Fribourg canton, Switzerland. Geoderma 217-218, 26–36. doi:10.1016/j.geoderma.2013.11.001

Roudier, P., Beaudette, D.E., Hewitt, A.E., 2012. A conditioned Latin hypercube sampling algorithm incorporating operational constraints, in: Minasny, B., Malone, B., McBratney, A. (Eds.), Digital Soil Assessments and beyond. CRC Press, Boca Raton, pp. 227–231.

Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. Prog. Phys. Geogr. 27, 171–197. doi:10.1191/0309133303pp366ra

Seibert, J., Stendahl, J., Sørensen, R., 2007. Topographical influences on soil properties in boreal forests. Geoderma 141, 139–148. doi:10.1016/j.geoderma.2007.05.013

Siegel, S., Castellan, N.J., 1988. Nonparametric statistics for the behavioral sciences. McGraw-hill, New York.

Smeck, N.E., 1985. Phosphorus dynamics in soils and landscapes. Geoderma 36, 185–199.

Soethe, N., Lehmann, J., Engels, C., 2008. Nutrient availability at different altitudes in a tropical montane forest in Ecuador. J. Trop. Ecol. 24, 397–406. doi:10.1017/S026646740800504X

Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods 14, 323–348. doi:10.1037/a0016973

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. doi:10.1016/j.ecolind.2007.05.005

Tesfa, T.K., Tarboton, D.G., Chandler, D.G., McNamara, J.P., 2009. Modeling soil depth from topographic and land cover attributes. Water Resour. Res. 45, 1–16. doi:10.1029/2008WR007474

Tucker, C.J., Sellers, P.J., 1986. Satellite remote sensing of primary production. Int. J. Remote Sens. 7, 1395–1416. doi:10.1080/01431168608948944

Turner, B.L., 2008. Resource partitioning for soil phosphorus: A hypothesis. J. Ecol. 96, 698–702. doi:10.1111/j.1365-2745.2008.01384.x

Uriarte, M., Turner, B.L., Thompson, J., Zimmerman, J.K., 2015. Linking Spatial Patterns of Leaf Litterfall and Soil Nutrients in a Tropical Forest: a Neighborhood Approach. Ecol. Appl. 25, 150313143409001. doi:10.1890/15-0112.1

Vitousek, P.M., Hättenschwiler, S., Olander, L., Allison, S., 2002. Nitrogen and Nature. Ambio A J. Hum. Environ. 31, 97–101. doi:10.1579/0044-7447-31.2.97

Vitousek, P.M., Porder, S., Houlton, B.Z., Chadwick, O. a, Applications, S.E., January, N., Applications, E., Houlton, Z., 2010. Terrestrial phosphorus limitation : mechanisms , implications , and nitrogen — phosphorus interactions. Ecol. Appl. 20, 5–15. doi:10.1890/08-0127.1

Wang, K., Zhang, C., Li, W., 2013. Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. Appl. Geogr. 42, 73–85. doi:10.1016/j.apgeog.2013.04.002

Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. Plant Soil 340, 7–24. doi:10.1007/s11104-010-0425-z

Wilcke, W., Oelmann, Y., Schmitt, A., Valarezo, C., Zech, W., Homeier, J., 2008. Soil properties and tree growth along an altitudinal transect in Ecuadorian tropical montane forest. J. Plant Nutr. Soil Sci. 171, 220–230. doi:10.1002/jpln.200625210

Wohl, E., 2010. Mountain rivers. American Geophysical Union, Washington, DC.

Zellweger, F., Braunisch, V., Morsdorf, F., Baltensweiler, A., Abegg, M., Roth, T., Bugmann, H., Bollmann, K., 2015. Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests. For. Ecol. Manage. 349, 36–44. doi:10.1016/j.foreco.2015.04.008

Zevenbergen, L.W., Zevenbergen, L.W., Thorne, C.R., Thorne, C.R., 1987. Quantitative analysis of land surface topography. Earth Surf. Process. Landforms 12, 47–56.

Zhang, Z.M., Yu, X.X., Qian, S., Li, J.W., 2010. Spatial variability of soil nitrogen and phosphorus of a mixed forest ecosystem in Beijing, China. Environ. Earth Sci. 60, 1783–1792. doi:10.1007/s12665-009-0314-z
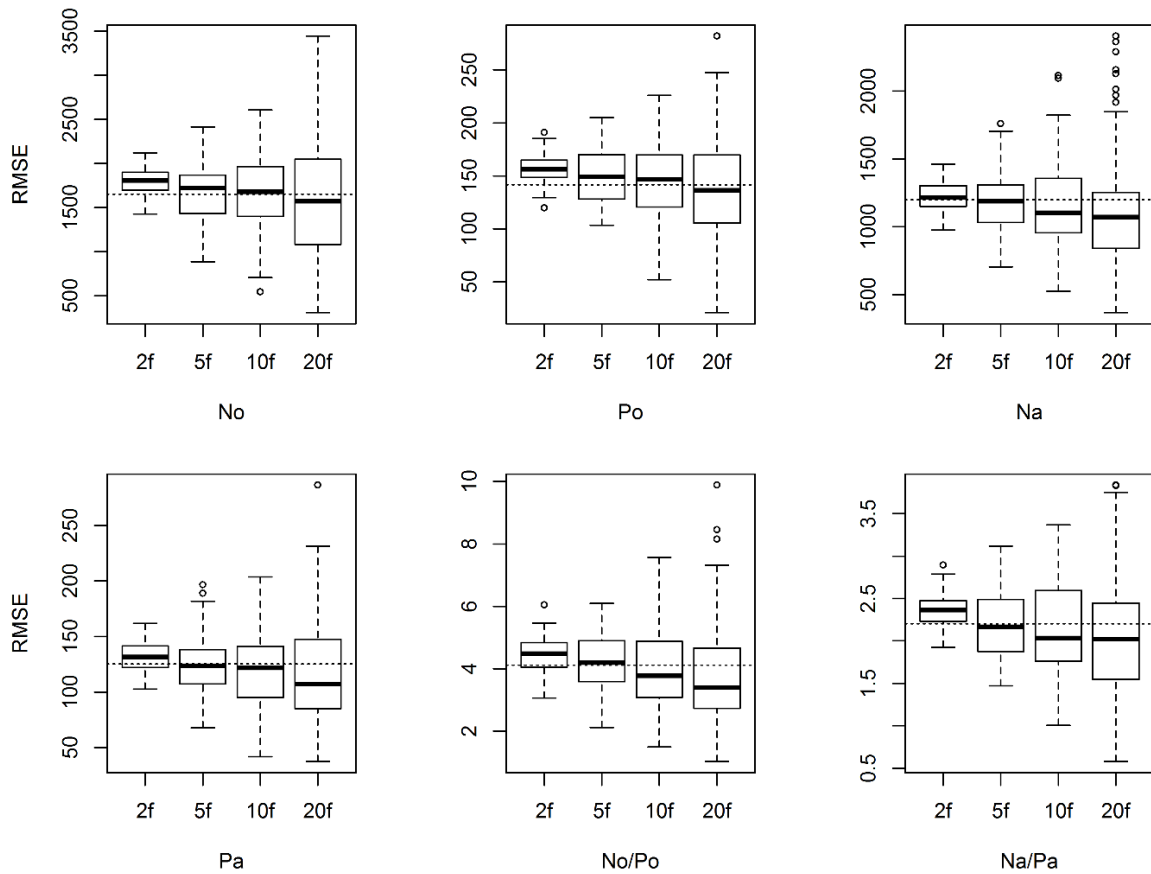
Appendix 4.1 Model validation based on RMSE with cross validation methods. The dot lines refer to the leave-one-out cross-validated result. 2f: 2-fold 50 repetitions, 5f: 5-fold 20 repetitions, 10f: 10-fold 10 repetitions, 20f: 20-fold 5 repetitions, N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.
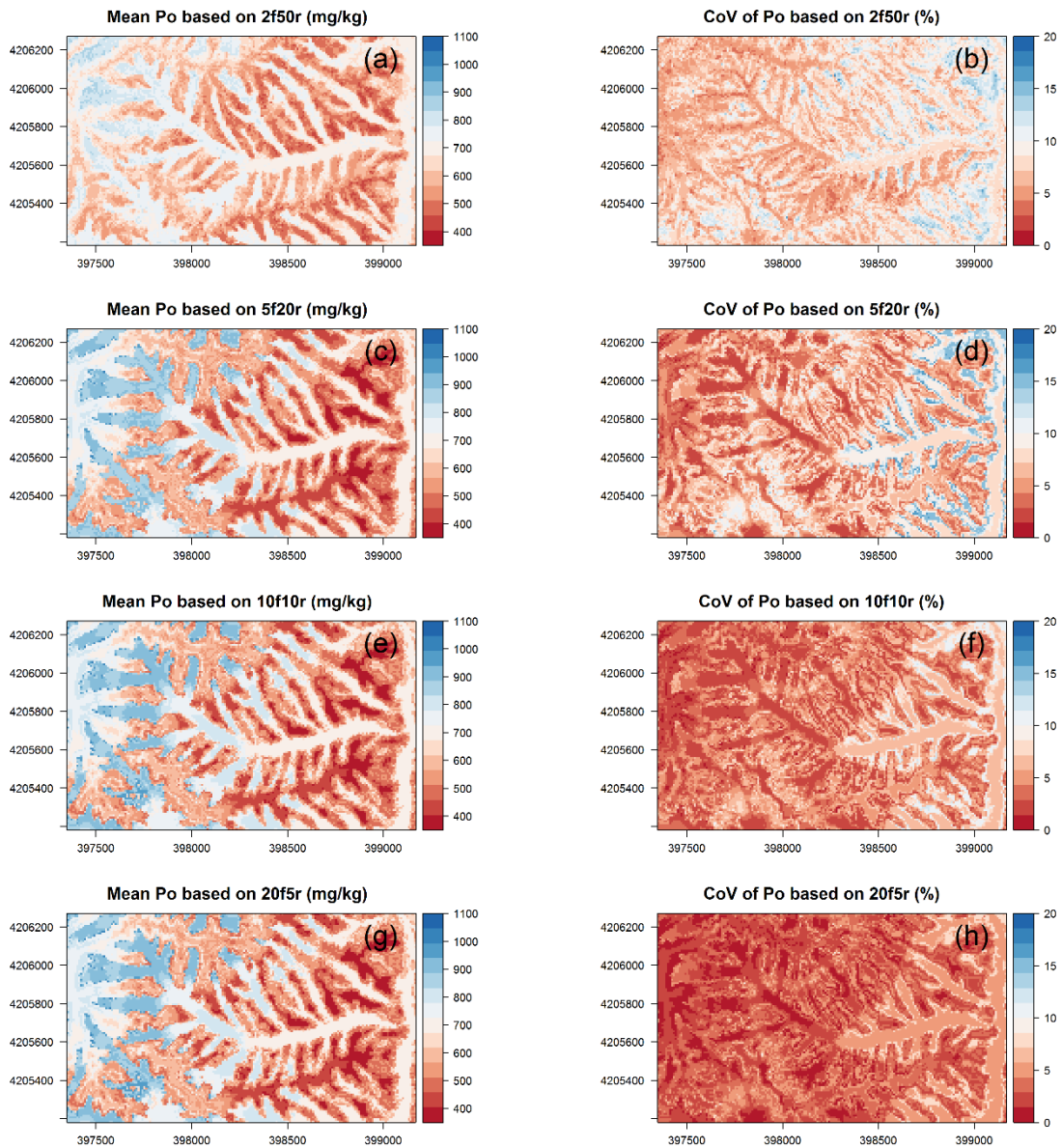
Appendix 4.2 Maps of mean and coefficient of variation (CoV) of 100 models of phosphorus in organic layers (Po) with cross validation methods. 2f50r: 2-fold 50 repetitions, 5f20r: 5-fold 20 repetitions, 10f10r: 10-fold 10 repetitions, 20f5r: 20-fold 5 repetitions, N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.
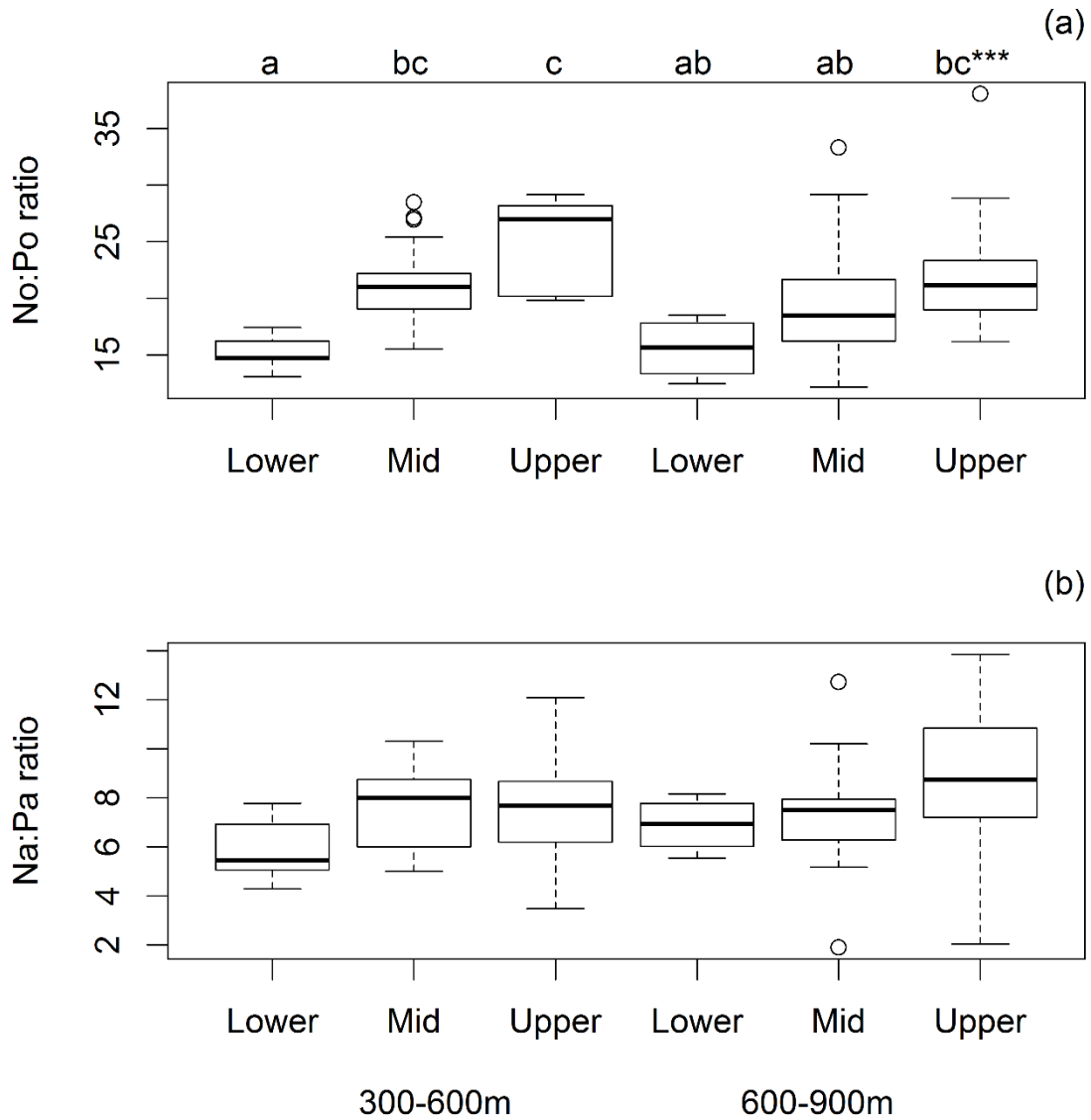
Appendix 4.3 Nitrogen and phosphorus ratios in the organic horizon (a) and in the mineral soil A horizon (b) based on two altitudes of 300-600 m and 600-900 m and three topographic positions (lower slope, midslope, and upper slope). Different letters indicate significant differences significantly different at $p<0.05$ with Kruskal-Wallis test. *** $p<0.001$.
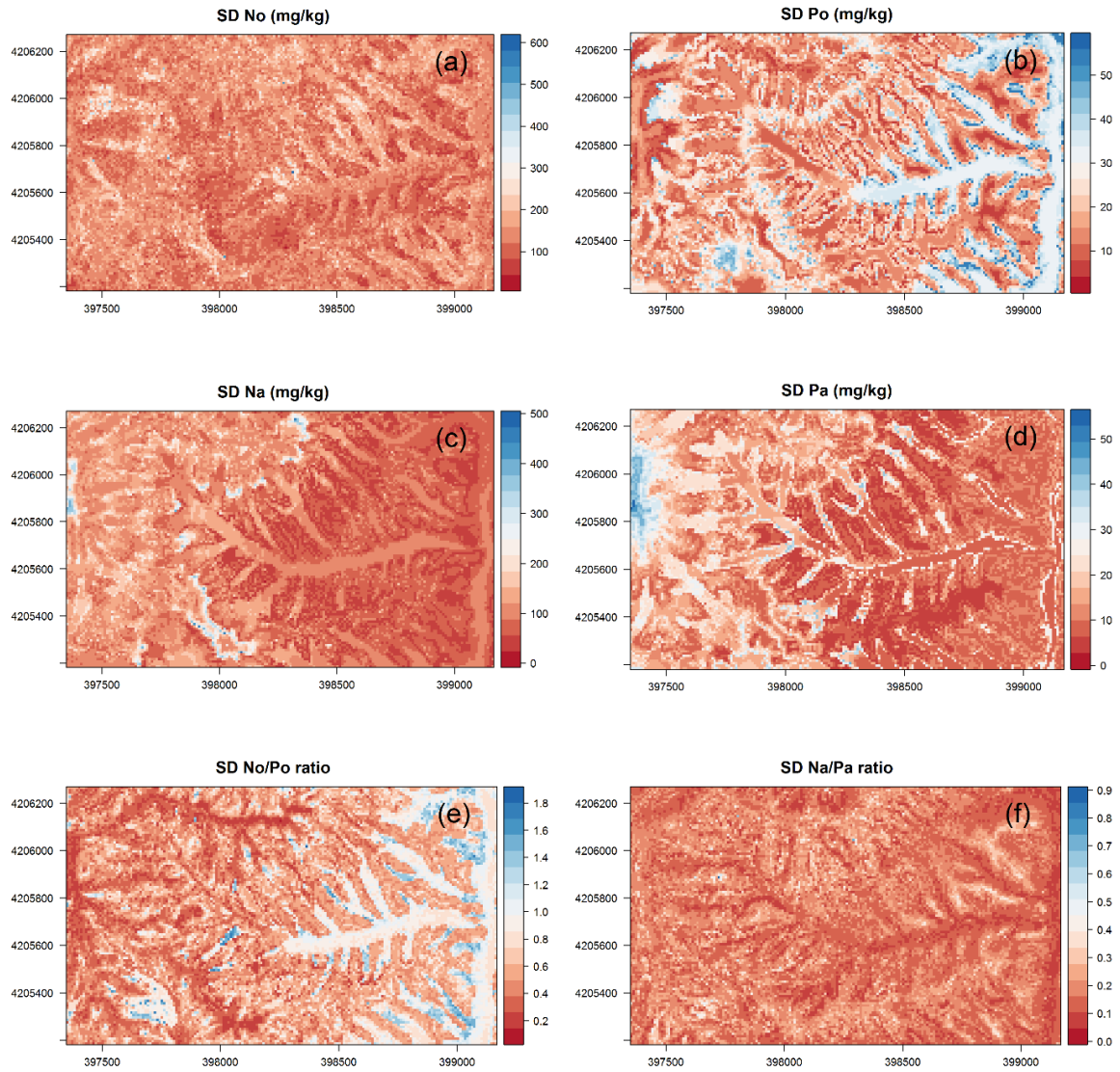
Appendix 4.4 Predicted SD soil N and P contents and ratios. SD: standard deviation. N: nitrogen, P: phosphorus, o: organic horizon, a: mineral A horizon.

(Eidesstattliche) Versicherungen und Erklärungen

(§ 8 S. 2 Nr. 6 PromO)

Hiermit erkläre ich mich damit einverstanden, dass die elektronische Fassung meiner Dissertation unter Wahrung meiner Urheberrechte und des Datenschutzes einer gesonderten Überprüfung hinsichtlich der eigenständigen Anfertigung der Dissertation unterzogen werden kann.

(§ 8 S. 2 Nr. 8 PromO)

Hiermit erkläre ich eidesstattlich, dass ich die Dissertation selbständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

(§ 8 S. 2 Nr. 9 PromO)

Ich habe die Dissertation nicht bereits zur Erlangung eines akademischen Grades anderweitig eingereicht und habe auch nicht bereits diese oder eine gleichartige Doktorprüfung endgültig nicht bestanden.

(§ 8 S. 2 Nr. 10 PromO)

Hiermit erkläre ich, dass ich keine Hilfe von gewerbliche Promotionsberatern bzw. -vermittlern in Anspruch genommen habe und auch künftig nicht nehmen werde.

. . . .

Ort, Datum, Unterschrift