

RESEARCH ARTICLE

Environmental drivers of spatial patterns of topsoil nitrogen and phosphorus under monsoon conditions in a complex terrain of South Korea

Gwanyong Jeong^{1†*}, Kwanghun Choi², Marie Spohn³, Soo Jin Park⁴, Bernd Huwe¹, Mareike Ließ⁵

1 Department of Geosciences/ Soil Physics Division, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany, **2** Biogeographical Modelling, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany, **3** Department of Soil Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany, **4** Department of Geography, Seoul National University, Shillim-dong, Kwanak-gu, Seoul, South Korea, **5** Department of Soil Physics, Helmholtz Centre for Environmental Research–UFZ, Halle, Germany

✉ Current address: Department of Geography, Chonnam National University, Yongbong-ro, Buk-gu, Gwangju, South Korea

* gyjeong@jnu.ac.kr



OPEN ACCESS

Citation: Jeong G, Choi K, Spohn M, Park SJ, Huwe B, Ließ M (2017) Environmental drivers of spatial patterns of topsoil nitrogen and phosphorus under monsoon conditions in a complex terrain of South Korea. PLoS ONE 12(8): e0183205. <https://doi.org/10.1371/journal.pone.0183205>

Editor: Budiman Minasny, The University of Sydney, AUSTRALIA

Received: May 22, 2016

Accepted: August 1, 2017

Published: August 24, 2017

Copyright: © 2017 Jeong et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All soil and environmental data used in this study are available in the supporting information file (S1 and S2 File).

Funding: This study was carried out as part of the International Research Training Group TERRECO (GRK 1565/1) at the University of Bayreuth. It was funded by the German Research Foundation (DFG). This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding programme Open Access Publishing. The funders had no role in study

Abstract

Nitrogen (N) and phosphorus (P) in topsoils are critical for plant nutrition. Relatively little is known about the spatial patterns of N and P in the organic layer of mountainous landscapes. Therefore, the spatial distributions of N and P in both the organic layer and the A horizon were analyzed using a light detection and ranging (LiDAR) digital elevation model and vegetation metrics. The objective of the study was to analyze the effect of vegetation and topography on the spatial patterns of N and P in a small watershed covered by forest in South Korea. Soil samples were collected using the conditioned latin hypercube method. LiDAR vegetation metrics, the normalized difference vegetation index (NDVI), and terrain parameters were derived as predictors. Spatial explicit predictions of N/P ratios were obtained using a random forest with uncertainty analysis. We tested different strategies of model validation (repeated 2-fold to 20-fold and leave-one-out cross validation). Repeated 10-fold cross validation was selected for model validation due to the comparatively high accuracy and low variance of prediction. Surface curvature was the best predictor of P contents in the organic layer and in the A horizon, while LiDAR vegetation metrics and NDVI were important predictors of N in the organic layer. N/P ratios increased with surface curvature and were higher on the convex upper slope than on the concave lower slope. This was due to P enrichment of the soil on the lower slope and a more even spatial distribution of N. Our digital soil maps showed that the topsoils on the upper slopes contained relatively little P. These findings are critical for understanding N and P dynamics in mountainous ecosystems.

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Nitrogen (N) and phosphorus (P) are the most important nutrients for primary productivity in terrestrial ecosystems [1,2]. Soil nutrient content varies during long-term soil development, such that N increases 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 minerals. As a result, primary productivity is initially N-limited in lightly weathered soils but becomes increasingly P-limited in highly weathered soils over millions of years [3].

P limitation is enhanced by atmospheric N deposition [2,4]. In East Asia, where the population and economy are growing rapidly, atmospheric N deposition is currently very high [5]. In South Korea, atmospheric N inputs have rapidly increased due to large industrial operations and agricultural intensification [6–8]. The annual average wet input of N ranged from 12.9 to 24.9 kg ha⁻¹ year⁻¹ from 2005 to 2010 [6], and is markedly higher than that during pre-industrial times. This might have effects on the productivity, biodiversity, and community composition of plants [9].

An understanding of nutrient contents in the organic layer is critical for mountainous ecosystem management. Organic layers are made up of freshly fallen organic matter, including whole leaves, twigs, and fruits. Following mineralization of organic matter, the organic layer slowly supplies nutrients, which are absorbed by plant roots [10]. Therefore, nutrients that are returned to soil by litterfall are important for plant nutrition [11]. In particular, the N/P ratio in topsoil is used as an indicator of potential growth limitation [12], and the spatial patterns of nutrients in the organic layer and in the A horizon can provide insight into soil-vegetation relationships.

Many studies have assessed spatial patterns of soil N [13–15] and P [16–18]. Previous studies on mountain ecosystems have found environmental correlations between the N contents in the organic layer and topographic parameters in a temperate forested watershed [19] and in boreal forests [20]. Wilcke et al. [21] reported an elevation gradient of decreasing N and P content in organic layers, and Soethe et al. [22] found that the N stocks of the organic layer differ significantly between different elevations in tropical mountain forests. However, our understanding of quantitative relationships between the content of nutrients (especially P) in the organic layer, topography, and vegetation is limited. In this regard, recent advances in digital soil mapping (DSM) have allowed us to improve our knowledge on spatial patterns of N and P and their environmental controls.

DSM often uses topographical predictors derived from digital elevation models (DEM), such as elevation, slope angle, curvature, and wetness index [23,24]. According to Ballabio [25], maps of soil properties can be produced with good accuracy using only terrain parameters as predictors in mountainous areas. In addition, vegetation data might improve DSM results, especially for the organic layer since it strongly depends on the vegetation [26]. Various vegetation parameters derived from satellite images have helped to explain the spatial variability of soil nutrients when used as DSM predictors [27,28]. However, to our knowledge, no attempt has been made to use Light detection and ranging (LiDAR) derived vegetation metrics for the spatial predictions of soil properties.

LiDAR-derived vegetation metrics could extend our understanding of spatial soil data by providing insight into the relationship between soils and vegetation as they are related to the vegetation's vertical variability, which reflects forest structure metrics [29]. Canopy cover percentage and maximum height can indicate the above ground biomass and forest productivity [30]. LiDAR predictors may also act as ecological indicators, such as light condition on the forest floor [30]. LiDAR intensity varies with land cover and forest types [31]. Additionally, LiDAR predictors are high-resolution data, which provide more detailed spatial information

than can be obtained from other types of remote sensing data (e.g. Aster [15 m] or Landsat [30 m] images). The normalized difference vegetation index (NDVI) and LiDAR data are expected to be important for N predictions related to forest biomass, but most probably not for P since it is assumed to mainly originate from bedrock.

LiDAR DEM could also be useful for predicting the spatial distributions of soil nutrients, especially P. P in soils tends to be fixed into stable forms as iron, aluminium, and calcium combinations [32]. Most P in soils is lost by soil erosion and is moved along surface configuration [33]. The LiDAR DEM can provide high resolution information on topography which might benefit the investigation of spatial P patterns.

To better understand the spatial patterns of N and P in the organic layer and mineral topsoil, the aim of this study was to use high-resolution LiDAR data and the derived DEM and vegetation metrics to predict topsoil N and P content by a DSM regression approach. The specific objectives of our research were: (1) to test the importance of LiDAR-derived vegetation and topographical parameters to understand the spatial patterns of N and P; (2) to identify subareas with critical P contents; and (3) to test different validation strategies for N and P.

Materials and methods

Research area

The study area has a size of 9.84 km² and is located in the downstream area of the Soyang lake watershed, Gangwon province, South Korea (Fig 1). The mean annual air temperature of the study area is 11.1°C and it receives a mean annual rainfall of 1,347 mm [34] with about 70% of the annual rain (824.4 mm) falling in the summer monsoon season (June, July, and August) [34]. The area's bedrock is part of the Gyeonggi gneiss complex, which consists of granitic gneiss and banded gneiss [35] formed in the Paleoproterozoic and belonging to the oldest basement rocks in the Korean Peninsula [36]. The elevation ranges between 320 and 868 m above sea level and the area consists of various steep slopes (over 45°) caused by a tectonic uplift that occurred during the Quaternary Period [37]. The area is a headwater catchment with narrow depositional areas and valleys, and plays an important role in the biogeochemical cycle of the downstream hydrological system as a key source of nutrients [38]. Its soils are

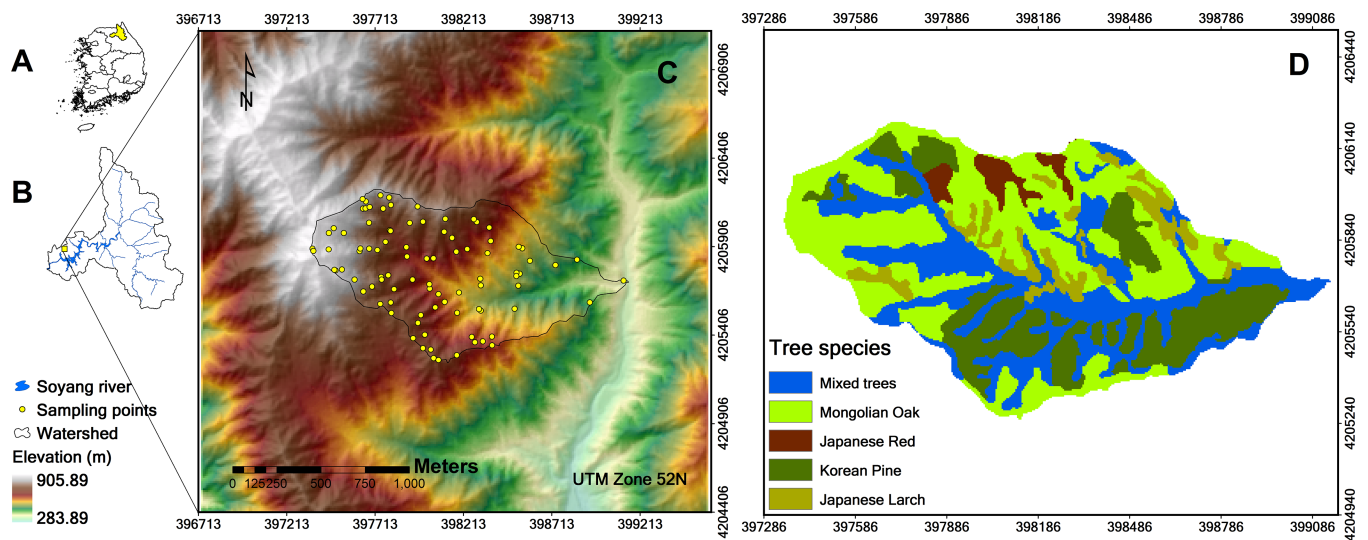


Fig 1. Research area. (A) The Soyang watershed within South Korea. (B) The research area within the Soyang watershed. (C) The research area with the sampling points. (D) The tree species map (fgis.forest.go.kr).

<https://doi.org/10.1371/journal.pone.0183205.g001>

mainly composed of fine gravelly sandy loam soils, fine sandy loam, and gravelly loam soils [39]. The area is part of a national forest and the 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*) (Fig 1).

Soil sampling and chemical analyses

Soil samples were collected from the organic layer and the A horizon at 91 sampling sites in 2014. Spatial position information of sampling points was recorded with a Qmini H3 global navigation satellite system (GNSS) GPS (accuracy within 5 m). Field studies were carried out under research permission from the Korea Forest Service of Chuncheon. We confirm that the field studies did not involve endangered or protected species. Conditioned Latin Hypercube Sampling (cLHS) was applied to optimize the density functions of the n-dimensional covariate space for the regression models [40]. This is a stratified random sampling approach that divides the empirical density functions of the predictor space into quantiles based on the number of samples. In order to obtain a Latin hypercube of exactly one sample per quantile for each of the predictors, an optimization approach is used. In the R package "clhs" [41], this is achieved by simulated annealing.

The organic layer had an average depth of 5 cm and was sampled using a metal frame of 0.3 x 0.3 m. The A horizon of the mineral soil was sampled using a shovel according to the depth of the A horizon, which differed between 10 and 30 cm. Mineral soil samples were air-dried and sieved (< 2 mm). The organic layer samples were oven-dried. Total P was extracted with HNO₃ and HF and measured according to DIN EN ISO 11885 / 22036 [42] by ICP-OES (Perkin Elmer, 2100 ZL, USA). After grinding to a fine powder, total N was measured by an elemental analyzer NA 1108 (CE Instruments, Milano, Italy). N/P ratios were calculated based on mass.

Environmental predictors

LiDAR is a remote sensing technology, which provides structural information on the illuminated surface, including the 3D terrain, vegetation canopy information, and object heights [43]. Point data, including x, y, and z coordinates, can be converted to a digital terrain model and a digital surface model [44]. The laser emits short pulses of light and the sensor records several returns from leaves, branches, and the underlying ground surface [29]. Vegetation heights can be derived from the difference between the ground and the non-ground returns [29]. LiDAR also generates intensity data, reflecting characteristics of objects, which can provide useful information on forest types and tree species [31]. Detailed overviews are provided by Asner et al. [45] and Hyypä et al. [46].

We used LiDAR point data which has a vertical accuracy of below 10 cm and an average of 4.08 points/m², surveyed by the National Geographic Information Institute (NGII) in South Korea [47]. The point data were pre-processed to identify ground returns, classify all returns, and calculate the normalized vegetation heights. Furthermore, we calculated a set of forest structural predictors using the LAStools software which provides a wide variety of methods to process LiDAR data [48] (Table 1). First, the ground and non-ground points were classified using the lasground module of LAStools. Then, the ground points were used to produce a digital elevation model with the las2dem module, and heights of non-ground points were calculated using the lasheight module. Finally, LiDAR vegetation metrics were derived using the lascanopy module. The maximum height (Hmax) was computed from the maximum point height within a grid cell. Variations of all vegetation point heights within a grid cell were converted to the standard deviation of heights (Hstd), which indicates the structural diversity of the forest. The canopy cover (Hccp) was calculated as the number of LiDAR first returns

Table 1. Environmental predictors for digital soil mapping.

	Predictor	Method	Reference
1	Elevation (ELEV)	Las2dem LAStools module	Isenburg [48]
2	Slope degree (SLO)	Slope, aspect, curvature SAGA module	Zevenbergen et al. [53]
3	Catchment area (CA)	Catchment area (Parallel) SAGA module (Multiple flow direction)	Freeman [54]
4	SAGA topographical wetness index (STWI)	SAGA wetness index SAGA module	Böhner et al. [55]
5	Surface curvature (CUR19)	CURV3 program	Park et al. [52]
6	Normalized difference vegetation index (NDVI)	(NIR-Red)/(NIR+Red)	Tucker and Sellers [56]
7	Maximum height (Hmax)	Lascanopy LAStools module	Isenburg [48]
8	Canopy cover percentage (Hccp)	Lascanopy LAStools module	Isenburg [48]
9	Standard deviation of heights (Hstd)	Lascanopy LAStools module	Isenburg [48]
10	Forest canopy and height (Hch)	Canopy cover percentage (Hccp) x maximum height (Hmax)	-
11	First return intensity average (Hfiravg)	Lasgrid LAStools module	Isenburg [48]

Note: NIR, near-infrared.

<https://doi.org/10.1371/journal.pone.0183205.t001>

greater than the cover cutoff (1.37 m by default) divided by the total number of first returns [48]. NDVI was derived from a 4-m Kompsat-2 image obtained on 11th October 2014 [49,50]. We selected the clear-sky image taken at the similar time as the field survey.

Most topographical predictors were calculated with the terrain analysis modules of the open source software SAGA based on the LiDAR DEM [51]. In addition, surface curvature, which reflects the degree of bending of the three-dimensional surface morphology, was calculated with the CURV3 program [52]. To consider the variability of surface configuration, surface curvature values were calculated with different search window sizes of 3 x 3 to 35 x 35 cells. The one with the highest Pearson’s correlation coefficient with the response variables N and P was finally selected as a predictor: 19 x 19 cells (CUR19). All predictors were converted to 10-m cell size via the nearest neighbor resampling method.

Random forest

Random forest (RF) is an ensemble learning method that operates by building a set of regression trees and averaging the results [57]. Each tree is built using bootstrap samples of the data and a subset of predictors. Providing the number of trees is large, the overall accuracy (out-of-bag error) of the RF converges [57]. Accordingly, the number of trees was set to 1000. The size of the predictor subset (mtry) was tuned by the R package “caret” [58]. The R package “randomForest” [57] was employed as a dependency.

RF is able to model complex nonlinear relationships between soil properties and environmental predictors. It is easier to apply than other supervised learning methods (e.g. neural networks and support vector regression) and does not require much tuning [58–60]. It also has a better interpretability due to the provision of a predictor importance measure. For this measure, the predictor values are permuted. The importance is then determined by the difference in mean square error before and after permutation [59]. Overall, RF has demonstrated good performance in DSM applications [16,61–64].

Predictor selection is reported to influence model performance [65–67]. Recursive feature elimination (RFE), a backward predictor selection method, begins with all predictors and iteratively eliminates the least important predictors one by one based on an initial measure of RF predictor importance until the best predictor remains [58]. At the end, the optimal number of predictors and the final list of selected predictors are returned. The package “caret” provides the functions for RFE [58].

To assess model performance, R^2 and root mean square error (RMSE) were calculated. For model validation, we used k-fold cross-validation (CV) where the dataset is randomly partitioned 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 k-fold CV, the choice of k determines the size of the test and training dataset. For example, in the case of 10-fold CV, 10% of the data are used for validation and the remaining 90% are used for calibration. The choice of k is usually 5 or 10; however there is no formal rule [58]. Although the subsets are generated randomly, the subdivision still affects model validation results. This can be acknowledged by repetitions of the k-fold CV. Still, the number of repetitions (n) might also affect the estimated model performance; for example, more repetitions lead to better results [68]. We explored 2-, 5-, 10-, 20-fold, and leave-one-out (LOO) CV in n repetitions to account for a total of 100 validation measures: $n \times k = 100$. Ultimately, 100 R-squares and RMSEs were returned for each soil property. Finally, the cell-wise standard deviation of the corresponding 100 predictions provides an estimate of spatial uncertainty.

Results

Descriptive statistics of soil nutrients

Summary statistics for the N and P data are shown in Table 2. The mean N value of the organic layer (N_o) was higher than that of the A horizon (N_a). N_o had the lowest coefficient of variation (CoV), while total P in the organic layer (P_o) showed a relatively higher variance based on the standard deviation and CoV. This indicates that the variability in the N/P ratios in the organic layer (N_o/P_o) was dependent on P_o content, and that there was major P input from the litter fall. The N/P ratio in the A horizon (N_a/P_a) showed a higher relative variability than did those in the organic layer, as indicated by the CoV. The mean N_o/P_o was 20.83 ± 4.82 and the mean N_a/P_a was 7.91 ± 2.42 .

Model validation

Fig 2 and S1 Fig show that with increasing k in repeated k-fold CV, mean R-square and RMSE values indicate a better model performance, while R-square and RMSE variance increases as well. Based on mean R-square, the LOO CV results were inferior to the repeated 10-fold and 20-fold, but superior to the repeated 2-fold results. Concerning repeated 5-fold CV, LOO CV was superior for the predictions of the organic layer nutrients, but inferior for the predictions of the mineral soil nutrients. Altogether, mean R-square values were higher for P_o and P_a compared to N_o and N_a respectively. The results for N_o/P_o and N_a/P_a were the worst, but showed

Table 2. Statistical summary of N and P content (mg kg⁻¹) and ratios.

	Mean	SD	MIN	Median	MAX	CoV (%)	Skew	Kurt
N_o	12245	1986	8000	12200	17800	16.22	0.35	2.92
P_o	624	190	310	610	1240	30.39	0.44	2.97
N_a	2990	1348	700	2600	7300	45.07	0.81	3.52
P_a	389	171	160	330	920	43.96	1.40	4.52
N_o/P_o	20.83	4.82	12.16	20.17	38.06	23.12	0.76	3.77
N_a/P_a	7.91	2.42	1.89	7.78	13.85	30.55	0.21	3.06

Notes: SD, standard deviation; MIN, minimum; MAX, maximum; CoV, coefficient of variation; Skew, skewness; Kurt, kurtosis; N, nitrogen; P, phosphorus; o , organic layer; and a , A horizon.

<https://doi.org/10.1371/journal.pone.0183205.t002>

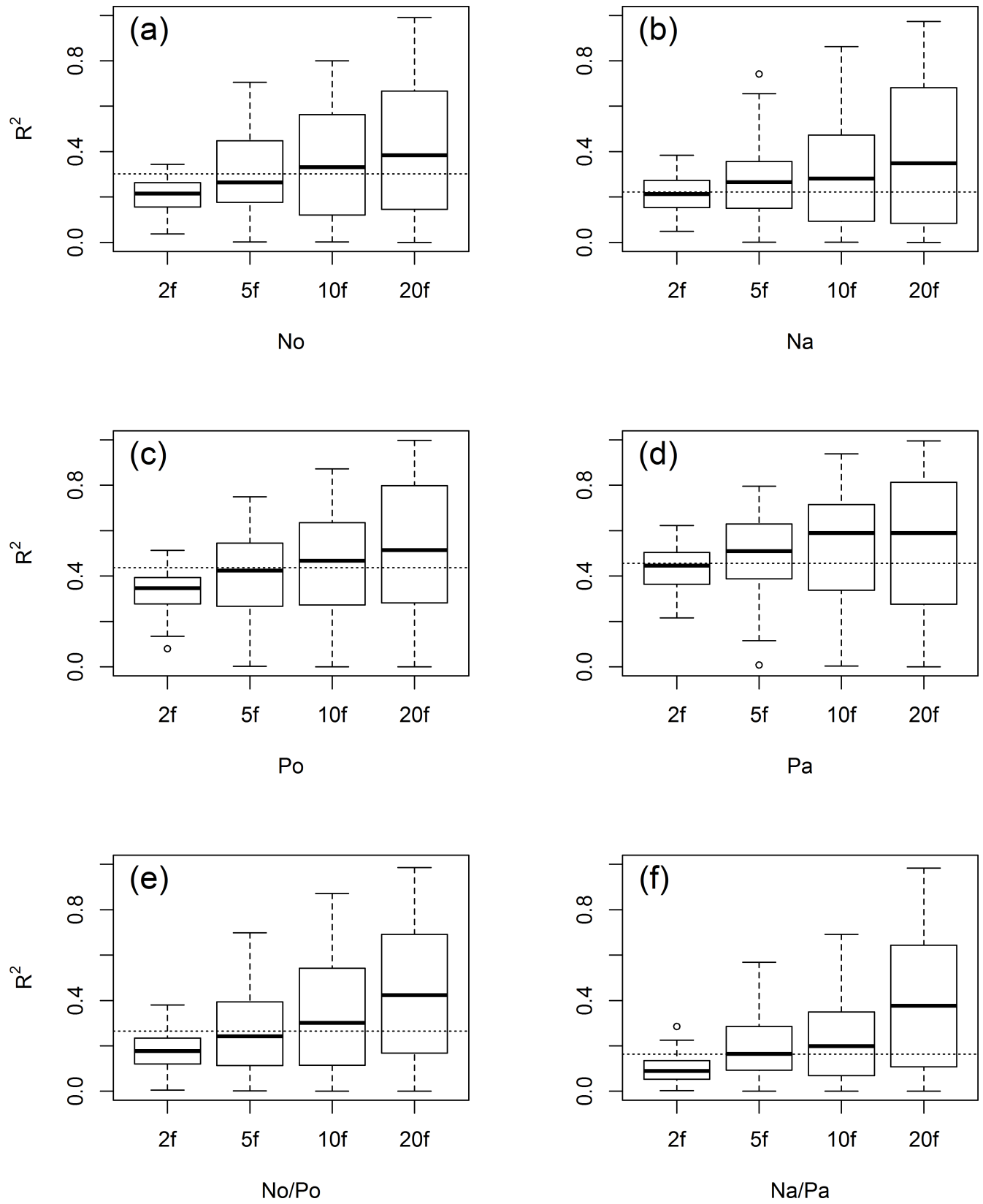


Fig 2. Model validation based on R-square with cross validation methods. The dotted lines indicate the leave-one-out cross-validated result. 2f, 2-fold 50 repetitions; 5f, 5-fold 20 repetitions; 10f, 10-fold 10 repetitions; 20f, 20-fold 5 repetitions; N, nitrogen; P, phosphorus; o, organic layer; and a, A horizon.

<https://doi.org/10.1371/journal.pone.0183205.g002>

the highest increase in model performance (mean R-square) with increasing k . Fig 3 shows the standard deviations of all raster cells according to the 100 spatial predictions resulting from the 100 models from the various CV schemes. The mean standard deviation and the variance of the standard deviations decrease with increasing k for all models.

As an example, spatial prediction patterns of P_o including mean values and the standard deviations from the 100 predictions according to the various CV schemes are displayed in Fig 4. In particular, spatial patterns of mean P_o of the repeated 5-, 10-, and 20-fold CV are optically very similar (Fig 4C, 4E and 4G). Only the results from repeated 2-fold CV (Fig 4A) show a comparatively smaller range of mean P_o values with lower values in the valleys and higher values along ridges. Furthermore, the increase of mean P_o values with elevation, which was particularly observable in the concave valley for repeated 5-, 10- and 20-fold CV, is less pronounced for repeated 2-fold CV. As already indicated by Fig 3, standard deviation values decrease with increasing k and a correspondingly bigger calibration dataset. The spatial patterns of the standard deviations show an abrupt increase in the concave valley in the lower part of the study area (Fig 4B, 4D, 4F and 4H).

Environmental drivers of spatial nutrient patterns

To analyze the influence of topography and vegetation on soil nutrients, the results from repeated 10-fold CV are displayed. These correspond to a comparatively good performance for all soil nutrients based on mean R-square, while R-square variance is not as high as for repeated 20-fold CV (Fig 2). The predictors selected with RFE are shown in Table 3. Surface curvature and elevation were selected for all soil nutrients. For P_o and P_a , they were the only selected predictors. NDVI and LiDAR vegetation predictors (Hfiravg, Hstd, and Hmax) were additionally selected for N_o . For the N/P ratios parameters corresponding to water flow were additionally selected. While the models for N_o/P_o in correspondence to N_o also included vegetation metrics as predictors (Hst, Hmax, and Hch), the model for N_a/P_a included the NDVI instead. We expected that the tree species influenced the spatial pattern of N/P ratios (Fig 1). Tree species were initially also tested as predictors; however, these were not considered important predictors based on previous results. Accordingly, they were excluded due to the simplicity of the model.

Our RF model revealed good performance for all soil nutrients based on R^2 (Fig 2). Mean R-square values ranged from 0.23 to 0.52. P_a showed the best result of the validation, while that of the R-square for N_a/P_a was lowest. Models for P showed better results than did models for N.

Fig 5 shows the mean relative predictor importance of the RF models created by repeated 10-fold CV. Terrain predictors exhibited 5.37–53.07% of the reduction in the mean square error (MSE). Surface curvature was the best or second best predictor for all soil nutrients, with the exception of N_o (Fig 5); contributed 6.50–53.07% of the MSE. Elevation exhibited a similarly high predictor importance: 9.55–39.22%. NDVI and LiDAR derived vegetation metrics (Hstd, Hmax, Hpdy, and Hfiravg) were also important predictors for the nutrients. The results showing the RF predictor importance were not consistent with the RFE results; however, the two results were similar and there was no difference in the most important predictors (Table 3).

The map of each nutrient displays the mean of the 100 predictions from repeated 10-fold CV (Fig 6). N_o and N_a content increased with elevation. We found that P content differed

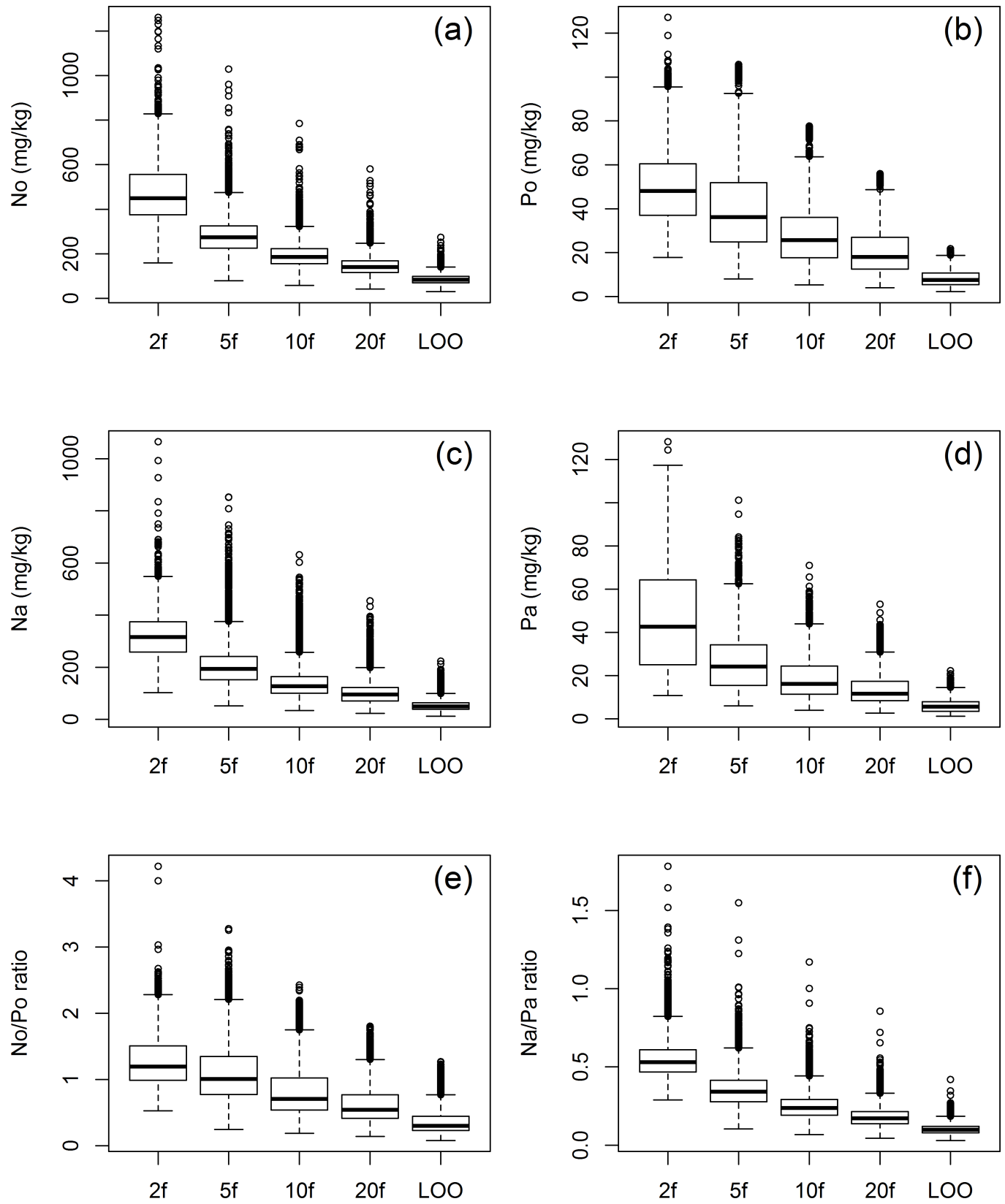


Fig 3. Boxplots showing standard deviations of 100 predicted values for each raster cell with cross validation methods. 2f, 2-fold 50 repetitions; 5f, 5-fold 20 repetitions; 10f, 10-fold 10 repetitions; 20f, 20-fold 5 repetitions; LOO, leave-one-out; N, nitrogen; P, phosphorus; o, organic layer; and a, A horizon.

<https://doi.org/10.1371/journal.pone.0183205.g003>

markedly between the upper and lower slopes. N_o/P_o and N_a/P_a were higher on the convex upper slope.

Higher standard deviations of P_o and N_o/P_o were found at lower elevations and on the valley floor (S2 Fig). The spatial uncertainties of P_a were higher at the upper part of the catchment. Uncertainties of N_o (S2 Fig) were similarly complex like the spatial pattern of the mean values (Fig 6A).

Discussion

Predictors of soil N and P

In this study, N_o ($r = 0.58$, $p < 0.001$) and N_a ($r = 0.49$, $p < 0.001$) were correlated with elevation. Likewise, Bedison and Johnson [69] also found a strong relationship between N_o and elevation ($R^2 = 0.41$, $P < 0.001$) in mountainous forested areas in the USA. Additionally, positive relationships between N_a and elevation were reported by Kunkel et al. [15], Wang et al. [70] and Peng et al. [13]. The catchment area (CA) and topographical wetness index (TWI) were important predictors of N_o in other studies [19,20]. In our study, CA and TWI were not significant for N_o , whereas N_a was correlated with TWI ($r = 0.26$, $p < 0.05$). According to Aandahl [71], higher nitrogen content is found on the lower slope. Higher N_a was found in areas with high elevation and on the lower slope (Fig 6C), which might have higher productivity (plants and microbes) and therefore, higher nitrogen fixation.

Vegetation can determine the spatial distribution of N in forest ecosystems [69,72]. For N_o , NDVI ranked as the second most important predictor and the LiDAR intensity of first returns (Hfiravg), which is often used as an indicator of forest type [31], was also an important predictor. Although NDVI and LiDAR predictors were not selected as predictors of the N_a model, N_a was weakly correlated with maximum height ($r = 0.24$, $p < 0.05$) and standard deviations of heights ($r = 0.23$, $p < 0.05$). Other studies have found significant relationships between N_a and NDVI which can measure vegetation density and aboveground biomass [15,16,73]. This implies that the density of forest cover and forest types affects the N_o content and N_o/P_o ratios. Vesterdal et al. [74] reported significant differences for N_o but not for N_a based on tree species and forest types. However, no relationship was found between P and LiDAR predictors.

As noted, LiDAR-derived predictors are promising for spatial soil predictions. In future studies, vegetation predictors should be applied to forest areas where there is difference in the variation of forest cover. Forest structure (LiDAR metrics) can have an effect on erosion and deposition of materials, which in turn, might alter the soil nutrient content. Hahm et al. [75] confirmed that differences in erosion rates are affected by tree canopy cover. However, to our knowledge, no studies have investigated the relationship between soil erosion, forest structures, and nutrient status using LiDAR data so far.

Spatial patterns of N/P ratios

We found that N/P ratios increased with surface curvature and were higher on the upper slope compared to the lower slope. This was due to P enrichment of the soil on the lower slope and a more even distribution of N (Fig 6). N_o/P_o and N_a/P_a were strongly related to surface curvature (Fig 6), which implies that P dynamics are affected strongly by topography. This is likely because P was carried from the upper slope by surface and subsurface flows and accumulated

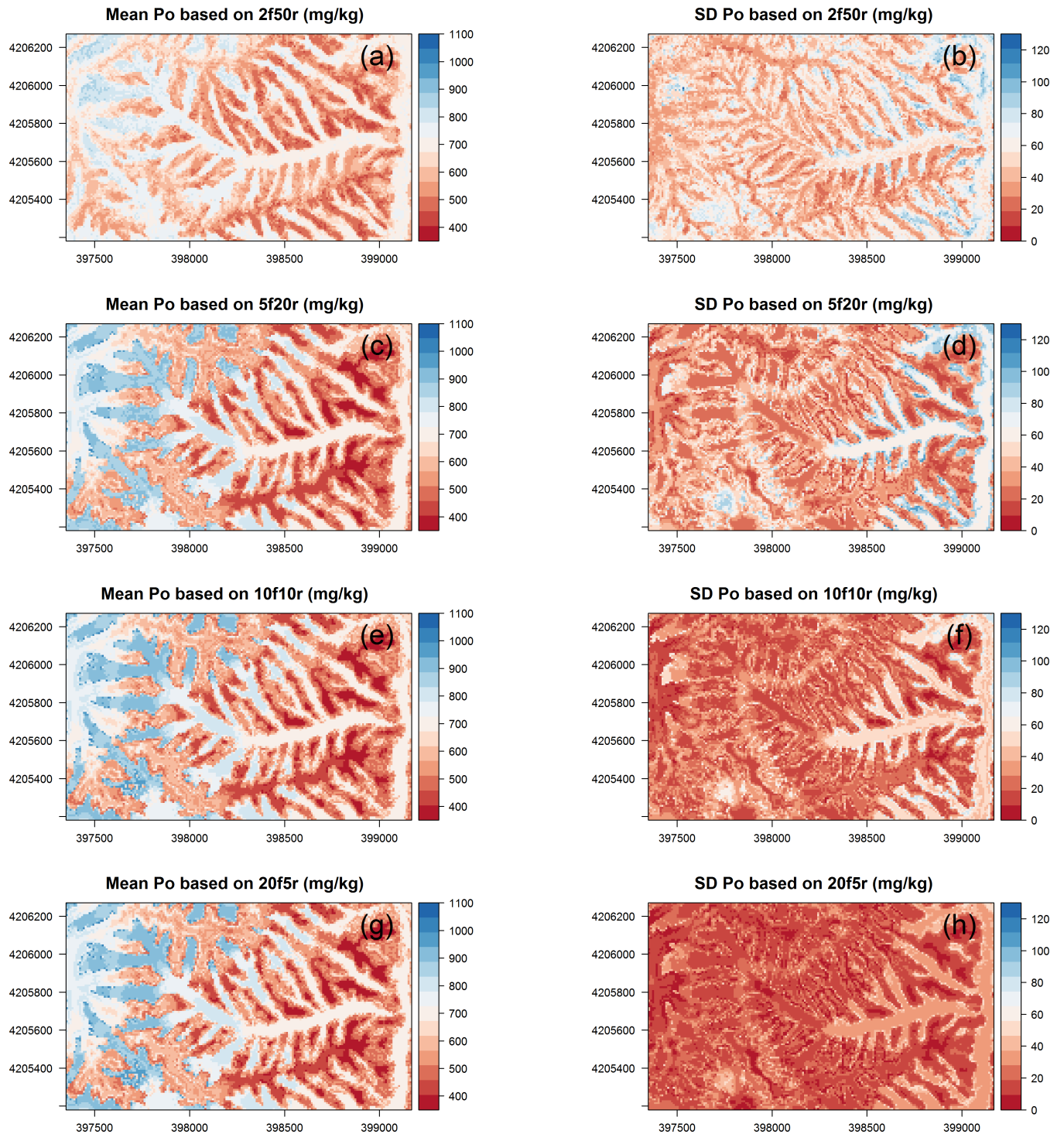


Fig 4. Maps of mean and coefficient of variation (CoV) of 100 models of phosphorus in the organic layer (P_o) with cross validation methods. 2f50r, 2-fold 50 repetitions; 5f20r, 5-fold 20 repetitions; 10f10r, 10-fold 10 repetitions; 20f5r, 20-fold 5 repetitions.

<https://doi.org/10.1371/journal.pone.0183205.g004>

on the lower slope, as observed previously in other areas [33]. Soil erosion in the watershed under study is strong due to storm events and steep slopes [76,77]. Consequently, higher soil P content on the lower slope than on the upper slope can lead to higher plant P uptake and

Table 3. Selected predictors using recursive feature elimination (RFE) based on repeated 10-fold cross validation.

Soil properties	Predictors
N _o	ELEV, NDVI, Hfiravg, CUR19, STWI, Hstd, Hmax
P _o	CUR19, ELEV
N _a	ELEV, CUR19
P _a	CUR19, ELEV
N _o /P _o	CUR19, CA, Hstd, ELEV, Hmax, Hch
N _a /P _a	CUR19, CA, NDVI, ELEV, STWI

Notes: ELEV, elevation; CUR19, surface curvature (19 x 19 local window); STWI, SAGA topographical wetness index; CA, Catchment area; SLO, slope degree; NDVI, normalized difference vegetation index; Hfiravg, first return intensity average; Hstd, standard deviations of heights; Hmax, maximum height; Hccp, canopy cover percentage; Hch, forest canopy and height (Hmax X Hccp); N, nitrogen; P, phosphorus; _o, organic layer; _a, A horizon.

<https://doi.org/10.1371/journal.pone.0183205.t003>

higher plant litter P content, leading to a lower N_o/P_o. This implies that spatial patterns of N_o/P_o might be generated by the interconnected relationships between soil, topography, and vegetation. Similarly, Uriarte et al. [78] found that soil N/P was correlated with leaf litter N/P, and was determined by topography in a tropical mountainous forest with heavy rainfall and steep slopes.

Model performance based on different cross validation schemes

We observed the typical bias-variance tradeoff when comparing the various CV schemes as was discussed at length in Hastie et al. [79]. With a higher k, the mean test error decreases, while test error variance increases (Fig 2, S1 Fig). In general, the performance of the learning

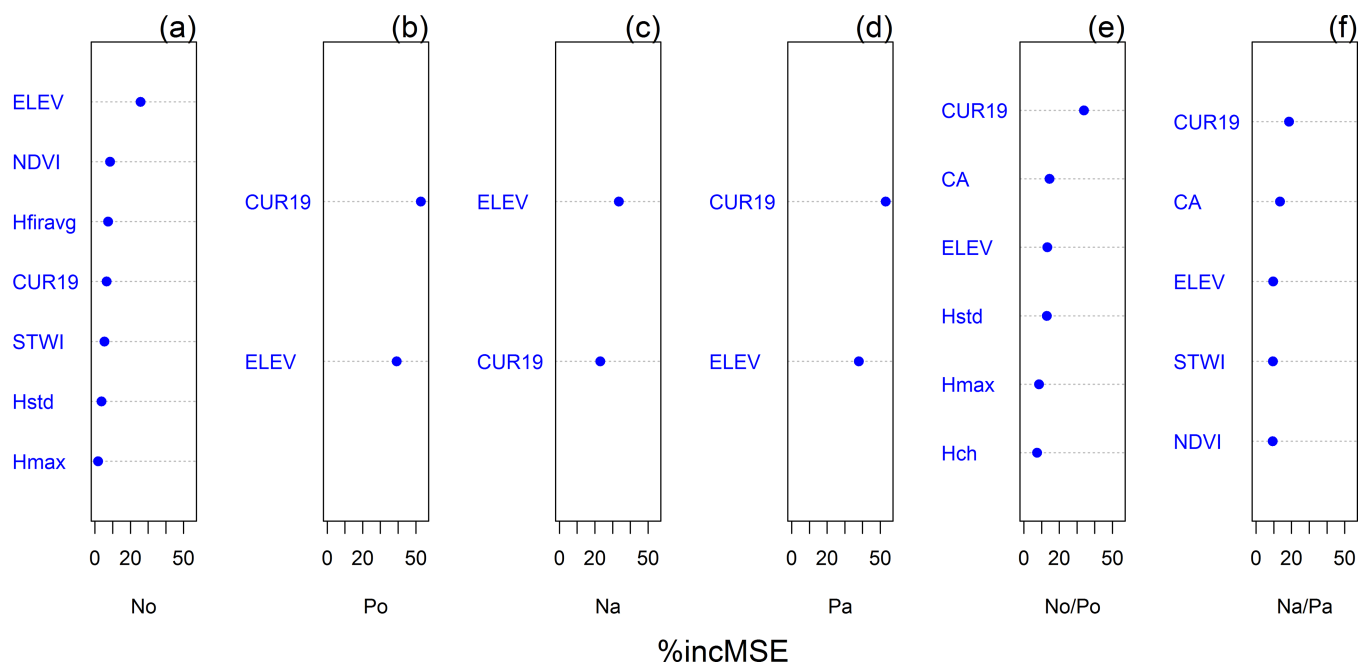


Fig 5. Mean relative importance of predictors for nitrogen and phosphorus based on the increased mean square error (%incMSE) from random forest. N, nitrogen; P, phosphorus; o, organic layer; and a, A horizon.

<https://doi.org/10.1371/journal.pone.0183205.g005>

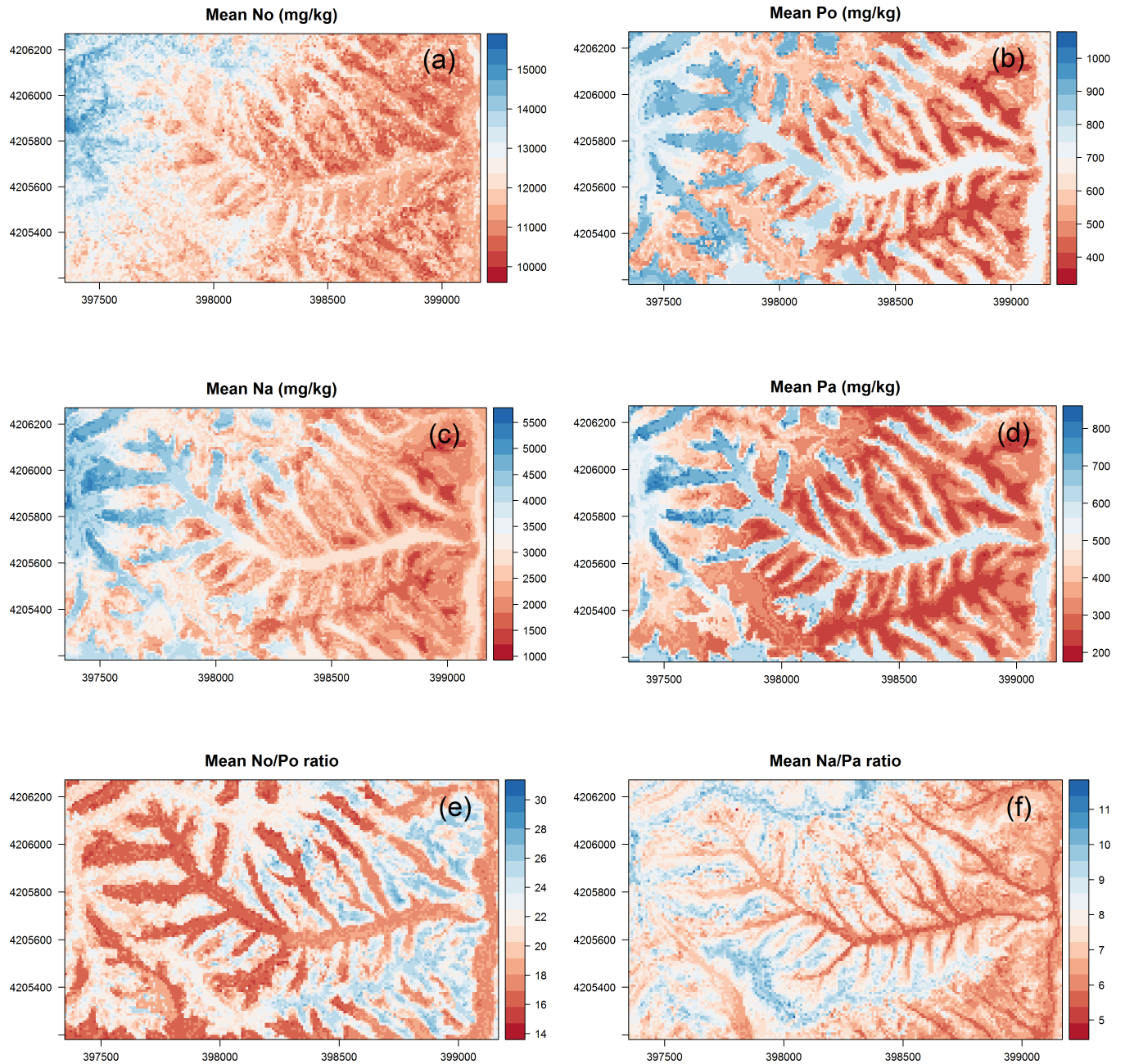


Fig 6. Predicted mean soil N and P content and ratios. N, nitrogen; P, phosphorus; o, organic layer; and a, A horizon.

<https://doi.org/10.1371/journal.pone.0183205.g006>

method varies with the size of the training set. A higher k results in a higher amount of training data, which can be crucial with small datasets. This pattern was consistent with the findings of previous studies. Park and Vlek [80] tested the change in prediction error with different numbers of training soil data sets, and confirmed that the prediction accuracy increases when increasing numbers of soil samples are used for the tuning dataset. A similar decrease in the prediction error was found using various methods for soil prediction according to Ballabio [25]. Generally, 10-fold CV is recommended in most studies [81–86]. Remesan and Mathew [81] noted that the use of very few datasets might result in poorly calibrated models, while

high amounts of data for calibration might lead to overfitting. For small sample sizes, model calibration requires all possible datasets to improve the model performance, while validation results can differ markedly depending on which samples are included in the validation [58]. Therefore, Kuhn and Johnson [58] suggested repeated 10-fold CV for small sample sizes because the bias and variance are somewhat balanced and the computational efficiency is good.

The size of the standard deviations of the spatial predictions corresponds to the applied CV scheme (Fig 3). Naturally, a low model bias goes along with low standard deviations. With a high amount of samples included in the training dataset, the training datasets and hence the 100 models are very similar to one another and will, therefore, make similar predictions. That this ensemble of RF models (e.g. from repeated 20-fold or LOO CV) comes along with a high error variance indicates that it is not a good choice, as the corresponding model might be overfitting the data and perform poorly on other data.

Conclusions

Here, we created the first digital soil maps, showing the spatial pattern of N/P ratios using LiDAR-derived vegetation and topographic predictors. These maps help to identify areas with low nutrient availability. In our study, repeated 10-fold CV was recommended for model validation with small sample sizes. While surface curvature and elevation were mostly sufficient to explain the overall spatial pattern, particularly N contents as well as nutrient ratios in the organic layer benefited from the inclusion of the LiDAR derived vegetation metrics. N/P ratios on the upper slope were higher than those on the lower slope and therefore, productivity on the upper slope might be limited by P in mountainous ecosystems under monsoon conditions. Finally, our analyses show that topographic and vegetation characteristics may help to predict the spatial distribution of nutrients and hence, nutrient limitation in mountainous regions.

Supporting information

S1 Fig. Model validation based on root mean square error (RMSE) with cross validation methods. The dotted 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 layer; and a, A horizon.
(TIFF)

S2 Fig. Predicted SD nitrogen and phosphorus content and ratios. SD, standard deviation; N, nitrogen; P, phosphorus; o, organic layer; and a, A horizon.
(TIFF)

S1 File. Soil nitrogen and phosphorus at soil sampling sites.
(XLS)

S2 File. Environmental predictors for digital soil mapping.
(ZIP)

Acknowledgments

The authors would like to thank Bomchul Kim, Youngsoon Choi, and Jaesung Eum from Kangwon National University, South Korea for providing a laboratory for the pretreatment of soil samples.

Author Contributions

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

Data curation: Gwanyong Jeong.

Formal analysis: Gwanyong Jeong, Kwanghun Choi.

Funding acquisition: Bernd Huwe.

Investigation: Gwanyong Jeong.

Methodology: Gwanyong Jeong, Kwanghun Choi, Bernd Huwe.

Project administration: Bernd Huwe.

Resources: Gwanyong Jeong.

Software: Gwanyong Jeong, Kwanghun Choi.

Supervision: Marie Spohn, Soo Jin Park, Bernd Huwe, Mareike Ließ.

Validation: Gwanyong Jeong.

Visualization: Gwanyong Jeong.

Writing – original draft: Gwanyong Jeong.

Writing – review & editing: Gwanyong Jeong, Marie Spohn, Bernd Huwe, Mareike Ließ.

References

1. Vitousek PM, Hättenschwiler S, Olander L, Allison S. Nitrogen and nature. *Ambio* 2002; 31: 97–101. <https://doi.org/10.1579/0044-7447-31.2.97>
2. Vitousek PM, Porder S, Houlton BZ, Chadwick O a, Applications SE, January N, et al. Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen—phosphorus interactions. *Ecol Appl*. 2010; 20: 5–15. <https://doi.org/10.1890/08-0127.1> PMID: 20349827
3. Laliberté E, Grace JB, Huston M a., Lambers H, Teste FP, Turner BL, et al. How does pedogenesis drive plant diversity? *Trends Ecol Evol*. 2013; 28: 331–340. <https://doi.org/10.1016/j.tree.2013.02.008> PMID: 23561322
4. Braun S, Thomas VFD, Quiring R, Flückiger W. Does nitrogen deposition increase forest production? The role of phosphorus. *Environ Pollut*. 2010; 158: 2043–2052. <https://doi.org/10.1016/j.envpol.2009.11.030> PMID: 20015583
5. Manning P. The impact of nitrogen enrichment on ecosystems and their services. In: Wall DH, editor. *Soil Ecology and Ecosystem Services*. Oxford University Press; 2012. pp. 256–269.
6. Jang S-K, Sung M-Y, Shin A-Y, Choi J-S, Son J-S, Ahn J-Y, et al. A Study for Long-term Trend of Acid Deposition in Korea. *J Korea Soc Environ Adm*. 2011; 17: 183–192.
7. Kim I, Lee K, Gruber N, Karl DM, Bullister JL, Yang S, et al. Increasing anthropogenic nitrogen in the North Pacific Ocean. *Science*. 2014; 346: 1102–1106. <https://doi.org/10.1126/science.1258396> PMID: 25430767
8. Kim T-W, Lee K, Najjar RG, Jeong H-D, Jeong HJ. Increasing N Abundance in the Northwestern Pacific Ocean Due to Atmospheric Nitrogen Deposition. *Science*. 2011; 334: 505–509. <https://doi.org/10.1126/science.1206583> PMID: 21940860
9. Turner BL. Resource partitioning for soil phosphorus: A hypothesis. *J Ecol*. 2008; 96: 698–702. <https://doi.org/10.1111/j.1365-2745.2008.01384.x>
10. Osman KT. *Soils: Principles, Properties and Management*. Dordrecht: Springer Science+ Business Media; 2012.
11. Huang W, Spohn M. Effects of long-term litter manipulation on soil carbon, nitrogen, and phosphorus in a temperate deciduous forest. *Soil Biol Biochem*. Elsevier Ltd; 2015; 83: 12–18. <https://doi.org/10.1016/j.soilbio.2015.01.011>
12. Cleveland CC, Liptzin D. C:N:P stoichiometry in soil: Is there a “Redfield ratio” for the microbial biomass? *Biogeochemistry*. 2007; 85: 235–252. <https://doi.org/10.1007/s10533-007-9132-0>

13. Peng G, Bing W, Guangpo G, Guangcan Z. 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*. 2013; 8: 1–9. <https://doi.org/10.1371/journal.pone.0083592> PMID: 24391791
14. Liu Z-P, Shao M-A, Wang Y-Q. Spatial patterns of soil total nitrogen and soil total phosphorus across the entire Loess Plateau region of China. *Geoderma*. 2013; 197–198: 67–78. <https://doi.org/10.1016/j.geoderma.2012.12.011>
15. Kunkel ML, Flores AN, Smith TJ, McNamara JP, Benner SG. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. *Geoderma*. Elsevier B.V.; 2011; 165: 1–11. <https://doi.org/10.1016/j.geoderma.2011.06.011>
16. Kim J, Grunwald S, Rivero RG. Soil phosphorus and nitrogen predictions across spatial escalating scales in an aquatic ecosystem using remote sensing images. *IEEE Trans Geosci Remote Sens*. 2014; 52: 6724–6737.
17. Roger A, Libohova Z, Rossier N, Joost S, Maltas A, Frossard E, et al. Spatial variability of soil phosphorus in the Fribourg canton, Switzerland. *Geoderma*. 2014; 217–218: 26–36. <https://doi.org/10.1016/j.geoderma.2013.11.001>
18. McKenzie NJ, Ryan PJ. Spatial prediction of soil properties using environmental correlation. *Geoderma*. 1999; 89: 67–94. [https://doi.org/10.1016/S0016-7061\(98\)00137-2](https://doi.org/10.1016/S0016-7061(98)00137-2)
19. Johnson CE, Ruiz-Mendez JJ, Lawrence GB. Forest Soil Chemistry and Terrain Attributes in a Catskills Watershed. *Soil Sci Soc Am J*. 2000; 64: 1804–1814.
20. Seibert J, Stendahl J, Sørensen R. Topographical influences on soil properties in boreal forests. *Geoderma*. 2007; 141: 139–148. <https://doi.org/10.1016/j.geoderma.2007.05.013>
21. Wilcke W, Oelmann Y, Schmitt A, Valarezo C, Zech W, Homeier J. Soil properties and tree growth along an altitudinal transect in Ecuadorian tropical montane forest. *J Plant Nutr Soil Sci*. 2008; 171: 220–230. <https://doi.org/10.1002/jpln.200625210>
22. Soethe N, Lehmann J, Engels C. Nutrient availability at different altitudes in a tropical montane forest in Ecuador. *J Trop Ecol*. 2008; 24: 397–406. <https://doi.org/10.1017/S026646740800504X>
23. McBratney AB, Mendonça Santos ML, Minasny B. On digital soil mapping. *Geoderma*. 2003; 117: 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
24. Grunwald S. *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics*. Boca Raton: Taylor & Francis, Boca Raton; 2006.
25. Ballabio C. Spatial prediction of soil properties in temperate mountain regions using support vector regression. *Geoderma*. 2009; 151: 338–350. <https://doi.org/10.1016/j.geoderma.2009.04.022>
26. Binkley D, Fisher R. *Ecology and Management of Forest Soils*. West Sussex: John Wiley & Sons; 2012.
27. Grunwald S, Vasques GM, Rivero RG. Fusion of soil and remote sensing data to model soil properties. *Adv Agron*. 2015; 131: 1–109. <https://doi.org/10.1016/bs.agron.2014.12.004>
28. Mulder VL, de Bruin S, Schaepman ME, Mayr TR. The use of remote sensing in soil and terrain mapping—A review. *Geoderma*. Elsevier B.V.; 2011; 162: 1–19. <https://doi.org/10.1016/j.geoderma.2010.12.018>
29. Jones HG, Vaughan RA. *Remote Sensing of Vegetation: Principles, Techniques, and Applications*. Oxford: Oxford University Press; 2010.
30. Zellweger F, Braunisch V, Morsdorf F, Baltensweiler A, Abegg M, Roth T, et al. Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests. *For Ecol Manage*. 2015; 349: 36–44. <https://doi.org/10.1016/j.foreco.2015.04.008>
31. Ørka HO, Næsset E, Bollandsås OM. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. *Remote Sens Environ*. 2009; 113: 1163–1174. <https://doi.org/10.1016/j.rse.2009.02.002>
32. Walker TW, Syers JK. The fate of phosphorus during pedogenesis. *Geoderma*. 1976; 15: 1–19. [https://doi.org/10.1016/0016-7061\(76\)90066-5](https://doi.org/10.1016/0016-7061(76)90066-5)
33. Smeck NE. Phosphorus dynamics in soils and landscapes. *Geoderma*. 1985; 36: 185–199.
34. Korea meteorological administration. Korea weather service. 2015. Available: <http://www.kma.go.kr/>.
35. Korea Institute of Geology Mining and Material. Explanatory note of the Gangreung Sokcho sheet 1:250,000. Daejeon: Korea Institute of Geology, Mining and Material; 2001.
36. Chough SK. *Geology and Sedimentology of the Korean Peninsula*. London: Elsevier; 2013.
37. Lee G. Characteristics of geomorphological surface and analysis of deposits in fluvial terraces at upper reach of Soyang river. *J Korean Geogr Soc*. 2004; 39: 27–44.
38. Wohl E. *Mountain Rivers*. Washington, DC: American Geophysical Union; 2010.

39. National Academy of Agricultural Science. Korean Soil Information System. 2013. Available: <http://soil.rda.go.kr/soil/index.jsp>.
40. Minasny B, McBratney AB. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput Geosci*. 2006; 32: 1378–1388. <https://doi.org/10.1016/j.cageo.2005.12.009>
41. Roudier P, Beaudette DE, Hewitt AE. A conditioned Latin hypercube sampling algorithm incorporating operational constraints. In: Minasny B, Malone B, McBratney A, editors. *Digital Soil Assessments and Beyond*. Boca Raton: CRC Press; 2012. pp. 227–231.
42. DEV. German Standard Methods for the Examination of Water, Wastewater and Sludge. Berlin: WILEY-VCH; 2002.
43. Franklin SE. *Remote Sensing for Biodiversity and Wildlife Management: Synthesis and Applications*. New York: McGrawHill; 2010.
44. Hyypä J, Hyypä H, Leckie D, Gougeon F, Yu X, Maltamo M. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. *Int J Remote Sens*. 2008; 29: 1339–1366. <https://doi.org/10.1080/01431160701736489>
45. Asner G, Ustin S, Townsend P, Martin R, Chadwick Kd. Forest biophysical and biochemical properties from hyperspectral and LiDAR remote sensing. In: Thenkabail PS, editor. *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*. Boca Raton: CRC Press; 2015. pp. 429–448. <https://doi.org/10.1201/b19322-22>
46. Hyypä J, Karjalainen M, Liang X, Jaakkola A, Yu X, Wulder M, et al. Remote sensing of forests from Lidar and Radar. In: Thenkabail PS, editor. *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*. Boca Raton: CRC Press; 2015. pp. 397–427. <https://doi.org/10.1201/b19322-21>
47. National Geographic Information Institute. National Spatial Information Clearinghouse. 2015. Available: <https://www.nsic.go.kr>.
48. Isenburg M. LAStools—efficient tools for LiDAR processing, version 2.1. 2014. Available: <http://lastools.org>.
49. Jensen JR. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Upper Saddle River, NJ, USA: Prentice Hall Press; 2015.
50. Thenkabail PS, Lyon JG, Huete A. Advances in hyperspectral remote sensing of vegetation and agricultural croplands. In: Thenkabail PS, Lyon JG, Huete A, editors. *Hyperspectral Remote Sensing of Vegetation*. Boca Raton: CRC Press; 2011. pp. 3–36. <https://doi.org/10.1201/b11222-3>
51. Conrad O, Bechtel B, Bock M, Dietrich H, Fischer E, Gerlitz L, et al. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci Model Dev*. 2015; 8: 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
52. Park SJ, McSweeney KK, Lowery BB. Identification of the spatial distribution of soils using a process-based terrain characterization. *Geoderma*. 2001; 103: 249–272. [https://doi.org/10.1016/S0016-7061\(01\)00042-8](https://doi.org/10.1016/S0016-7061(01)00042-8)
53. Zevenbergen LW, Zevenbergen LW, Thorne CR, Thorne CR. Quantitative analysis of land surface topography. *Earth Surf Process Landforms*. 1987; 12: 47–56.
54. Freeman TG. Calculating catchment area with divergent flow based on a regular grid. *Comput Geosci*. 1991; 17: 413–422. [https://doi.org/10.1016/0098-3004\(91\)90048-1](https://doi.org/10.1016/0098-3004(91)90048-1)
55. Böhner J, Köthe R, Conrad O, Gross J, Ringeler A, Selige T. Soil regionalisation by means of terrain analysis and process parameterisation. In: Micheli E, Nachtergaele F, Montanarella L, editors. *Soil Classification 2001*. Luxembourg: The European Soil Bureau, Joint Research Centre; 2002. pp. 213–222.
56. Tucker CJ, Sellers PJ. Satellite remote sensing of primary production. *Int J Remote Sens*. 1986; 7: 1395–1416. <https://doi.org/10.1080/01431168608948944>
57. Breiman L. Random forests. *Mach Learn*. 2001; 45: 5–32.
58. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York: Springer; 2013.
59. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Methods*. 2009; 14: 323–348. <https://doi.org/10.1037/a0016973> PMID: 19968396
60. Kampichler C, Wieland R, Calmé S, Weissenberger H, Arriaga-Weiss S. Classification in conservation biology: A comparison of five machine-learning methods. *Ecol Inform*. 2010; 5: 441–450. <https://doi.org/10.1016/j.ecoinf.2010.06.003>
61. Grimm R, Behrens T, Märker M, Elsenbeer H. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma*. 2008; 146: 102–113. <https://doi.org/10.1016/j.geoderma.2008.05.008>

62. Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil*. 2011; 340: 7–24. <https://doi.org/10.1007/s11104-010-0425-z>
63. Tesfa TK, Tarboton DG, Chandler DG, McNamara JP. Modeling soil depth from topographic and land cover attributes. *Water Resour Res*. 2009; 45: 1–16. <https://doi.org/10.1029/2008WR007474>
64. Ließ M, Hitziger M, Huwe B. The sloping mire soil-landscape of Southern Ecuador: Influence of predictor resolution and model tuning on random forest predictions. *Appl Environ Soil Sci*. 2014; 2014: 1–10. <https://doi.org/10.1155/2014/603132>
65. Brungard CW, Boettinger JL, Duniway MC, Wills SA, Edwards TC. Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*. 2015; 239–240: 68–83. <https://doi.org/10.1016/j.geoderma.2014.09.019>
66. Miller BA, Koszinski S, Wehrhan M, Sommer M. Impact of multi-scale predictor selection for modeling soil properties. *Geoderma*. 2015; 239–240: 97–106. <https://doi.org/10.1016/j.geoderma.2014.09.018>
67. Poggio L, Gimona A, Brewer MJ. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. *Geoderma*. 2013; 209–210: 1–14. <https://doi.org/10.1016/j.geoderma.2013.05.029>
68. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: A comparison of resampling methods. *Bioinformatics*. 2005; 21: 3301–3307. <https://doi.org/10.1093/bioinformatics/bti499> PMID: 15905277
69. Bedison JE, Johnson AH. Controls on the spatial patterns of carbon and nitrogen in Adirondack forest soils along a gradient of nitrogen deposition. *Soil Sci Soc Am J*. 2009; 73: 2105–2117. <https://doi.org/10.2136/sssaj2008.0336>
70. Wang K, Zhang C, Li W. Predictive mapping of soil total nitrogen at a regional scale: A comparison between geographically weighted regression and cokriging. *Appl Geogr*. Elsevier Ltd; 2013; 42: 73–85. <https://doi.org/10.1016/j.apgeog.2013.04.002>
71. Aandahl A. 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*. 1948; 13: 449–454.
72. Zhang ZM, Yu XX, Qian S, Li JW. Spatial variability of soil nitrogen and phosphorus of a mixed forest ecosystem in Beijing, China. *Environ Earth Sci*. 2010; 60: 1783–1792. <https://doi.org/10.1007/s12665-009-0314-z>
73. Sumfleth K, Duttman R. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecol Indic*. 2008; 8: 485–501. <https://doi.org/10.1016/j.ecolind.2007.05.005>
74. Vesterdal L, Schmidt IK, Callesen I, Nilsson LO, Gundersen P. Carbon and nitrogen in forest floor and mineral soil under six common European tree species. *For Ecol Manage*. 2008; 255: 35–48. <https://doi.org/10.1016/j.foreco.2007.08.015>
75. Hahm WJ, Riebe CS, Lukens CE, Araki S. Bedrock composition regulates mountain ecosystems and landscape evolution. *Proc Natl Acad Sci U S A*. 2014; 111: 3338–3343. <https://doi.org/10.1073/pnas.1315667111> PMID: 24516144
76. Jeong JJ, Bartsch S, Fleckenstein JH, Matzner E, Tenhunen JD, Lee SD, et al. 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*. 2012; 117: 1–13. <https://doi.org/10.1029/2012JG001999>
77. Jung BJ, Lee HJ, Jeong JJ, Owen J, Kim B, Meusburger K, et al. Storm pulses and varying sources of hydrologic carbon export from a mountainous watershed. *J Hydrol*. 2012; 440–441: 90–101. <https://doi.org/10.1016/j.jhydrol.2012.03.030>
78. Uriarte M, Turner BL, Thompson J, Zimmerman JK. Linking Spatial Patterns of Leaf Litterfall and Soil Nutrients in a Tropical Forest: a Neighborhood Approach. *Ecol Appl*. 2015; 25: 150313143409001. <https://doi.org/10.1890/15-0112.1>
79. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer Science+Business Media; 2009.
80. Park SJ, Vlek PLG. Environmental correlation of three-dimensional soil spatial variability: A comparison of three adaptive techniques. *Geoderma*. 2002; 109: 117–140. [https://doi.org/10.1016/S0016-7061\(02\)00146-5](https://doi.org/10.1016/S0016-7061(02)00146-5)
81. Remesan R, Mathew J. *Hydrological Data Driven Modelling: A Case Study Approach*. Springer International Publishing Switzerland; 2015.
82. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning with Applications in R*. New York: Springer; 2013.
83. Cichosz P. *Data Mining Algorithms: Explained Using R*. West Sussex: John Wiley & Sons; 2015.

84. Feigelson ED, Babu GJ. *Modern Statistical Methods for Astronomy: With R Applications*. New York: Cambridge University Press; 2012.
85. Malley JD, Malley KG, Pajevic S. *Statistical Learning for Biomedical Data*. New York: Cambridge University Press; 2011.
86. Ambrose C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci U S A*. 2002; 99: 6562–6566. <https://doi.org/10.1073/pnas.102102699> PMID: [11983868](https://pubmed.ncbi.nlm.nih.gov/11983868/)