Research Article

The Sloping Mire Soil-Landscape of Southern Ecuador: Influence of Predictor Resolution and Model Tuning on Random Forest Predictions

Mareike Ließ,[1] Martin Hitziger,[2] and Bernd Huwe[1]

[1] *Department of Geosciences/Soil Physics Division, University of Bayreuth, Universitaetsstraße 30, 95447 Bayreuth, Germany*
[2] *ETH Zürich, Environmental Natural and Social Sciences, Universitaetsstraße 22, 8092 Zürich, Switzerland*

Correspondence should be addressed to Mareike Ließ; mareike.liess@uni-bayreuth.de

The sloping mire landscape of the investigation area, in the southern Andes of Ecuador, is dominated by stagnic soils with thick organic layers. The recursive partitioning algorithm Random Forest was used to predict the spatial water stagnation pattern and the thickness of the organic layer from terrain attributes. Terrain smoothing from 10 to 30 m raster resolution was applied in order to obtain the best possible model. For the same purpose, several model tuning parameters were tested and a prepredictor selection with the R-package Boruta was applied. Model versions were evaluated and compared by 100 repetitions of the calculation of the residual mean square error of a five-fold cross-validation. Position specific density functions of the predicted soil parameters were then used to display prediction uncertainty. Prepredictor selection and tuning of the Random Forest algorithm in some cases resulted in an improved model performance. We therefore recommend testing prepredictor selection and tuning to make sure that the best possible model is chosen. This needs particular emphasis in complex tropical mountain soil-landscapes which provide a real challenge to any soil mapping approach but where Random Forest has proven to be successful due to the testing of model tuning and prepredictor selection.

1. Introduction

Tropical forests store significant amounts of organic carbon not only in their aboveground biomass but particularly in their soils. In tropical mountain forests which receive a high amount of precipitation, soil wetness is usually assumed to even increase organic layer thickness and, therefore, soil carbon stocks due to a lower nutrient availability and turnover rate [1, 2]. Furthermore, with the altitude decreasing temperatures were often reported to cause an additional increase of the thickness of the organic layer. Accordingly, Dieleman et al. [3] claim that montane tropical forests consistently contain larger amounts of soil organic carbon compared to tropical lowland forests. We will show in this study that the relation between climate and paludification (the formation of peatlands) in tropical mountain landscapes is still not completely understood and needs further investigation.

The soil landscape of the tropical mountain forest area in San Francisco, Ecuador, is dominated by soils with hydromorphic properties and thick organic layers [4, 5]. The soils are influenced by slope processes such as shallow slope parallel subsurface flow within the organic layer and the stagnic soil horizon [6, 7]. Rainfall and the geomorphology of the landscape have a strong influence on the genesis of stagnic soil properties [7] in determining how much water is accumulated. The more water reaches a particular soil compartment due to rainfall and shallow subsurface flow and the lesser the discharge capacities of the soil are due to its saturated hydraulic conductivity and its position in the landscape, the more likely it is that stagnic properties will develop. The term catena [8] refers to the relief determined pattern of soils on hillslopes. It is defined as a sequence of soils of about the same age derived from similar parent material and occurring under similar climatic conditions but having

different characteristics due to variation in relief and drainage [9].

Being an important control factor of soil formation and water logging in particular, relief parameters calculated from a digital terrain model (DTM) are often used to regionalise soil properties. Walker et al. [10], Thompson et al. [11], and Chaplot et al. [12] predicted hydromorphic soil properties by landform parameters. Furthermore, the relation between terrain parameters and soil drainage classes was analysed by Troeh [13] and Peng et al. [14]. Soil horizon thickness was regionalised by Moore et al. [15]. Moreover, Park et al. [16] classified the soil morphological pattern based on soil landscape units. For a detailed overview regarding digital soil terrain modelling please refer to [17]. However, so far only a few attempts have been made to apply digital soil mapping techniques in tropical mountain landscapes, due to their often high heterogeneity, difficult terrain, and low accessibility. Moreover, rainfall induced soil slides are an important factor within the development of this particular soil landscape. Accordingly, first attempts to predict the histic and stagnic soil layer [7] by classification and regression trees (CART) [5] resulted in very high prediction uncertainty.

The recursive partitioning algorithm Random Forest is known for its strong predictive force due to the prediction by multiple decision trees and the implemented algorithms to guarantee decision tree variability. However, for the random selection of a subset of predictors to develop each tree within Random Forest, each single tree is also more prone to lose its predictive force while weak predictors are included. This is a well-known problem concerning machine learning algorithms [18]. Kursa and Rudnicki [19] provided the R-package Boruta to select the minimal predictor set and exclude predictors without predictive force to solve this so-called minimal-optimal problem. Hence, different predictor resolutions, Boruta pre selection of predictors, and two tuning parameters to optimize the Random Forest algorithm shall be tested in order to obtain the best possible spatial prediction of three soil parameters: (1) the occurrence probability of hydromorphic/stagnic properties in the soil, (2) the vertical extent of these properties, that is, the thickness of the stagnic soil layer, and (3) the thickness of the organic layer. In order to refer to the former two, the term water stagnation pattern will be used. Furthermore, we use the term stagnic layer and not stagnic horizon to differentiate from its usage within the World Reference Base of Soil resources [7], where a stagnic soil horizon needs a clearly defined minimum thickness.

2. Material and Methods

2.1. Dataset.

The investigated soil-landscape comprises an area of c. 26 km^2 around the research station San Francisco (Figure 1). Soils were assessed by 56 soil profiles and 315 auger sampling points. To guarantee a representative dataset, sampling sites were selected according to a 24-terrain classes comprising sampling design [4]. While soil profiles, due to limited time, had to be positioned close to the existing footpath network, auger sampling took place along transects
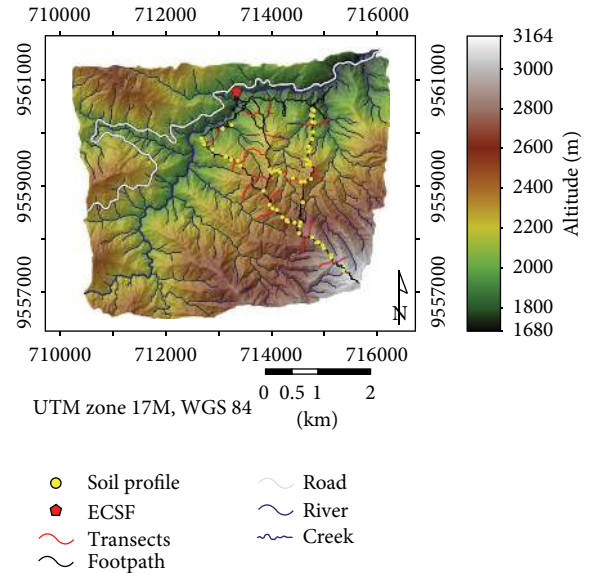

Figure 1: Sampling positions (adapted from [20]; overlaid hill shading with light source from north).

laid from hilltops to valley bottoms (Figure 1). Transect positions also had to be selected due to accessibility of the often very steep terrain.

2.2. Prediction Parameters.

In earlier studies [4, 5] the terrain parameters altitude, aspect, slope, terrain curvature, distance to the channel network (OFD), and specific upslope contributing catchment area, calculated by the kinematic routing algorithm (KRAA), were used as prediction parameters for organic layer and stagnic layer thickness [7].

According to Ließ et al. [4, 21], Stagnosol [7] probability increases above an altitude of about 2150 m a.s.l. on slope angles <40°. Schrumpf et al. [22] also report an increase in hydromorphic properties with altitude. The increase with altitude can be attributed to the increasing rainfall [23]. Lesser steep slope angles account for a slower discharge. Against all expectations, rather small specific upslope contributing catchment areas showed the highest Histosol [7] probability [21].

Apart from parameters being related to rainfall intensity like altitude or parameters describing relative water accumulation such as slope, OFD, and KRAA more predictors describing water accumulation are added in this study, that is, convergence index and Saga wetness index (SWI). We further assume that the soil's relative position on the slope is of high importance not only due to received rainfall and water accumulation potential but also regarding wind and incoming solar radiation, which are expected to have a stronger drying effect along the exposed mountain ridges and on the eastern slopes that are exposed to the main wind direction. To best describe these processes, the terrain parameters valley depth, normalised height, wind effect, terrain ruggedness index (TRI), and potential incoming solar radiation (PISR) were included as prediction parameters. All considered terrain parameters were

Table 1: SAGA modules to calculate terrain parameters.

Terrain parameter	Module library	Module	Reference	Saga author/year
Altitude	Terrain analysis—preprocessing	Fill sinks (Planchon/Darboux, 2001)	[24]	Wichmann/2003
Slope Aspect Profile curvature Plan curvature	Terrain analysis—morphometry	Slope, aspect, curvature	[25]	Conrad/2001
Convergence index	Terrain analysis—morphometry	Convergence index (search radius)	[26]	Conrad/2003
Normalised height Valley depth	Terrain analysis—morphometry	Relative heights and Slope positions	—	Böhner and Conrad/2008
TRI	Terrain analysis—morphometry	Terrain ruggedness index	[27]	Conrad/2010
Wind effect	Terrain analysis—morphometry	Wind effect	—	Böhner and Ringeler/2008, Conrad/2011
KRAA KRAS	Terrain analysis—hydrology	Catchment area (flow tracing)	[28]	Conrad/2001
SWI	Terrain analysis—hydrology	Saga wetness index	[29]	Böhner and Conrad/2001
PISR	Terrain analysis—lighting, visibility	Potential incoming solar radiation—direct insolation	[30–32]	Conrad/2010

calculated by SAGA modules according to Table 1 [33–35].

The predictor altitude, which corresponds to air temperature [36] as well as a different forest composition and structure [37], is assumed to strongly influence the organic layer thickness due to a lower nutrient availability and turnover rate [1, 2]. The relative position on the slope, represented by valley depth and normalised height, is the second important factor determining forest structure and results in a denser more light consuming vegetation with taller trees in the deep side valleys compared to the ridge structures [37]. Furthermore, Oesker et al. [38] report a different soil nutrient status for ridges and gorges within this area.

We, therefore, expect the complex interactions of (1) climatic factors such as rain and temperature represented by altitude, (2) factors representing water accumulation and discharge such as slope, KRAA, KRAS, SWI, and TRI, and (3) factors influencing the drying mechanisms such as wind effect, PISR, and relative slope position, represented by normalised height and valley depth to determine the spatial water stagnation pattern. The thickness of the organic layer is assumed to be determined by (1) the forest structure, (2) the soil nutrient status, (3) air temperature represented by altitude, valley depth, and normalised height, and (4) the soil water stagnation pattern.

The influence of the predictor resolution, particularly of the DTM, on digital soil mapping has been widely analysed and has been summarized by Behrens et al. [39]. Ließ [40] also assumed an effect of the DTM resolution on model uncertainty for this particular research area. Therefore, this study aims at finding the best possible Random Forest model to predict the spatial water stagnation pattern and the organic layer thickness within this tropical mountain landscape by means of a range of predictor resolutions and various model tuning parameters which include predictor selection and the size of the predictor set for each Random Forest tree.

2.3. Random Forest Algorithm. Methods from the field of statistical learning theory are often applied to understand the influence of terrain parameters on soil properties and use this relation to develop digital soil maps. Recursive partitioning methods, that is, CART, boosted classification trees, and Random Forest (RF) are used to predict soil units, clay content, or soil drainage classes from terrain parameters (e.g., [41–43]). These tree methods are amongst the most popular and widely used techniques for nonparametric regression and classification [44].

Recursive partitioning [45] is a procedure by which a data set, comprising one dependent (e.g., soil property) and many predictor (e.g., terrain parameters) variables, is progressively split into a dichotomously branching tree that optimizes the homogeneity of samples within subsets based on the dependent variable. At each splitting location within a tree, it is possible to determine the value of the predictor variable that best predicts the split in the data. For a regression problem (e.g., organic layer thickness), the optimal split is found by minimizing the mean square error. For a classification problem (e.g., occurrence of stagnic properties), the optimization criterion is the Gini Index [45].

RF, developed by Breiman [46], is an ensemble method which grows a number of classification or regression trees. Model stability is achieved through tree diversity by (1) choosing at random a subset of predictor variables (mtry) to grow each tree and (2) sampling with replacement (bootstrapping) and thereby varying the input dataset. The size of mtry has to be selected by the user. It is a sensitive parameter determining model strength for it defines the strength of each individual tree and the correlation between any two trees in the forest. Tree strength improves model performance, whereas correlation among trees weakens it.

mtry can be optimized by the R-function tuneRF. The tuning parameter sampsize determines the size of the data subset used for model development. It is set to 2/3 by default but can be varied. Hence, Random Forest contains several tuning parameters which control internal random processes.

RF was performed within the open-source data analysis environment R (version 2.13.2; R Development Core Team, 2011). It is implemented with the package randomForest which is based on Breiman and Cutler's FORTRAN code. The R-package Boruta for predictor selection is described by Rudnicki and Kursa [47]. Several tuning strategies were applied and compared to the Random Forest model with default parameters resulting in four model versions.

(1) No predictor selection and no tuning (R default values for *mtry* and *sampsize*).

(2) Additional predictor selection with the R package Boruta [19]. All confirmed and tentative predictors were used. In case predictor selection with Boruta led to better models compared to (1) it was also included into (3) and (4).

(3) Additional tuning of sampsize. Sampsize was tuned by fitting 12 models with different sampsizes (5, 10, 15, 20, . . . , 60).

(4) Additional tuning of mtry instead of sampsize.

As analysed [5], not only CART methodology, but also RF shows a strong dependence on the used dataset, especially while the dataset is small. In order to compare model performance and to estimate modelling uncertainty of the 12 adapted models—three spatial predictor resolutions times four model tuning strategies—a 5-fold cross-validation scheme was computed and conducted in 100-fold repetition to account for the effect of internal and external random effects. Internal effects refer to the bootstrapping and predictor selection procedure (mtry) implemented within Random Forest; external effects refer to the sample attribution to cross-validation groups. The so obtained RMSE distributions of the 12 models are then also compared to the RMSE distribution of the mean of the data which was calculated by the same scheme.

3. Results and Discussion

3.1. Preliminary Data Mining. Thick organic layers in the tropics are assumed to develop above poorly drained basins and depressions or in highland areas with a high precipitation/evapotranspiration ratio [7]. However, Ließ et al. [48] state that the wettest points in highland landscapes might be too wet to carry thick organic layers. It is often assumed that soil water logging limits organic matter turnover [49, 50] which results in the expectation of a positive correlation between the occurrence of stagnic properties and organic layer thickness. However, Figures 2(a) and 2(b) clearly show that there is no correlation between the occurrence of stagnic properties and the organic layer thickness for the investigated area as was already assumed by Ließ et al. [48].
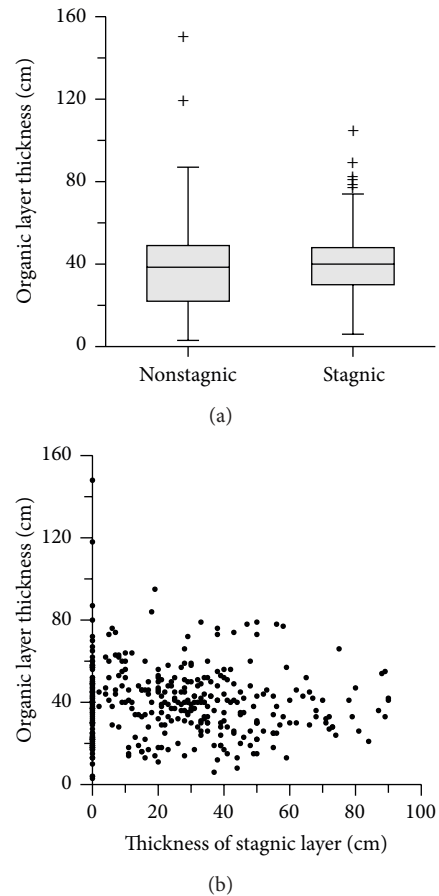

Figure 2: (a) Boxplots of organic layer thickness for soils with and without stagnic properties based on collected data set. (b) Scatter plot relating stagnic and organic layer thickness based on collected data set.

3.2. Model Performance. Predictor selection with Boruta (model version 2) resulted in no model improvement concerning the prediction of the occurrence probability of stagnic properties (Figures 3(a)–3(c)) and organic layer thickness (Figures 5(a)–5(c)) as indicated by the median RMSE. It rather impaired model performance. For the prediction of the thickness of the stagnic soil horizon it did, however, have a positive impact for the models using 10 and 20 m terrain resolution (Figures 4(a) and 4(b)).

The tuning of sampsize also had an ambivalent effect on model performance (model version 3), for it improved model performance for the models predicting the occurrence probability of stagnic properties, using 20 or 30 m terrain resolution, and even resulted in the best model (Figures 3(b) and 3(c)). The same is true for the prediction of the stagnic layer thickness, where it improved the models of all three terrain resolutions (Figures 4(a)–4(c)) and resulted in the overall best model for one of them (Figure 4(b)). In the latter case of 20 m terrain resolution being used for the prediction of the stagnic layer thickness, all other models (versions 1, 2, and 4) resulted in models worse than the data mean. In predicting organic layer thickness, the tuning of sampsize impaired model performance for 20
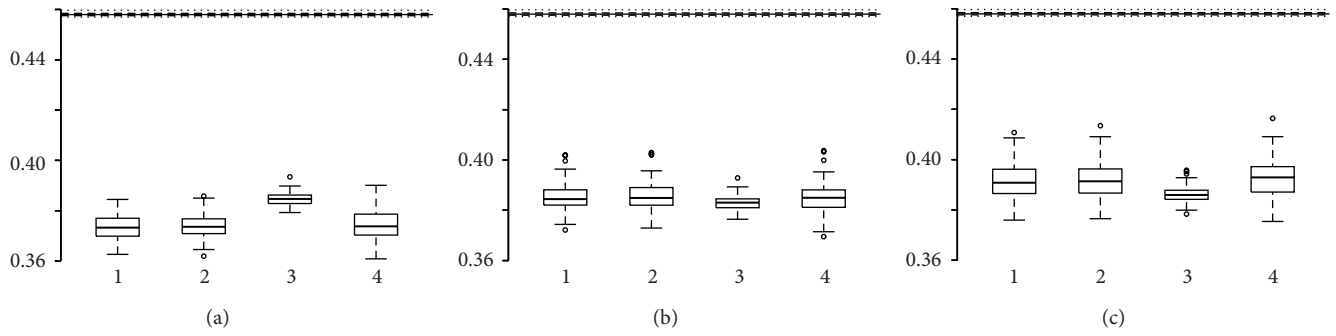
FIGURE 3: Boxplots of the 100 RMSE of the models predicting the occurrence probability of stagnic properties. (a) 10 m terrain resolution, (b) 20 m terrain resolution, and (c) 30 m terrain resolution. 1 = default Random Forest parameters, 2 = additional predictor selection, 3 = 1, 2+ tuning of sampsize, and 4 = 1, 2+ tuning of mtry. The lines refer to the boxplot parameters of the RMSE distribution of the mean of the data as prediction model. Solid line: median, dashed line: upper/lower hinges, and dotted line: upper/lower whiskers.

FIGURE 4: Boxplots of the 100 RMSE of the models predicting stagnic layer thickness. (a) 10 m terrain resolution, (b) 20 m terrain resolution, and (c) 30 m terrain resolution. 1 = default Random Forest parameters, 2 = additional predictor selection, 3 = 1, 2+ tuning of sampsize, and 4 = 1, 2+ tuning of mtry. The lines refer to the boxplot parameters of the RMSE distribution of the mean of the data as prediction model. Solid line: median, dashed line: upper/lower hinges, and dotted line: upper/lower whiskers.
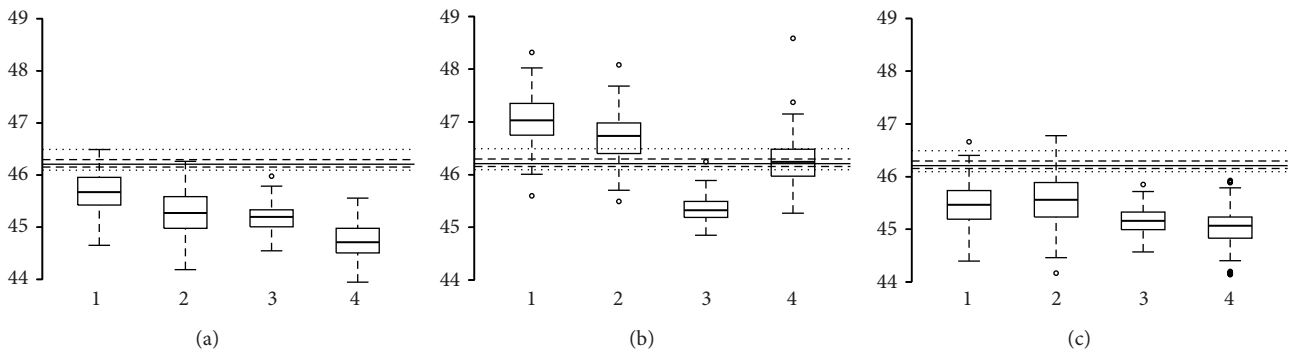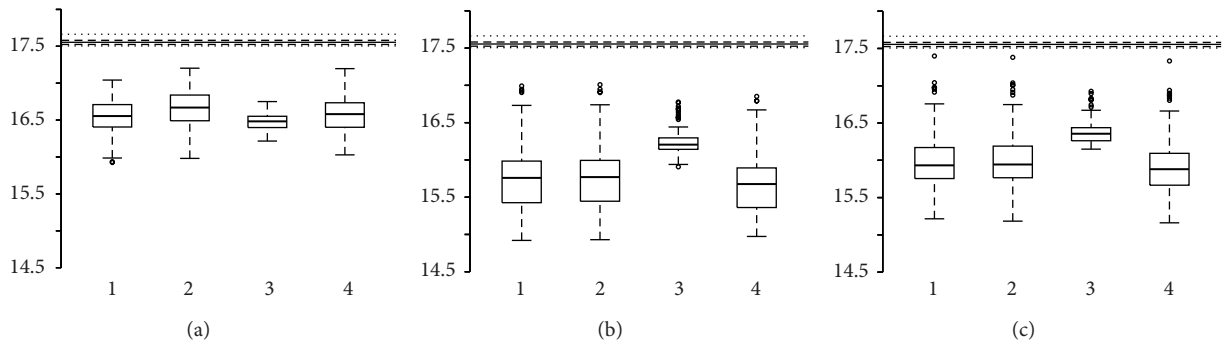

FIGURE 5: Boxplots of the 100 RMSE of the models predicting organic layer thickness. (a) 10 m terrain resolution, (b) 20 m terrain resolution, and (c) 30 m terrain resolution. 1 = default Random Forest parameters, 2 = additional predictor selection, 3 = 1, 2+ tuning of sampsize, and 4 = 1, 2+ tuning of mtry. The lines refer to the boxplot parameters of the RMSE distribution of the mean of the data as prediction model. Solid line: median, dashed line: upper/lower hinges, and dotted line: upper/lower whiskers.

and 30 m terrain resolution, resulting in the models with the highest RMSE median (Figures 5(b) and 5(c)) but for 10 m terrain resolution it resulted in the overall best model (Figure 5(a)). The tuning of sampsize always resulted in the lowest interquartile range of the RMSE distribution.

The tuning of mtry (model version 4) had only little or no negative effect in model uncertainty concerning the prediction of the occurrence of stagnic properties and the organic layer thickness. It did, however, always improve model performance regarding the prediction of the thickness of the stagnic soil horizon (Figures 4(a)–4(c)).

Because of their lowest median of the RMSE distribution the following models are considered best and will be used for prediction and map generation.
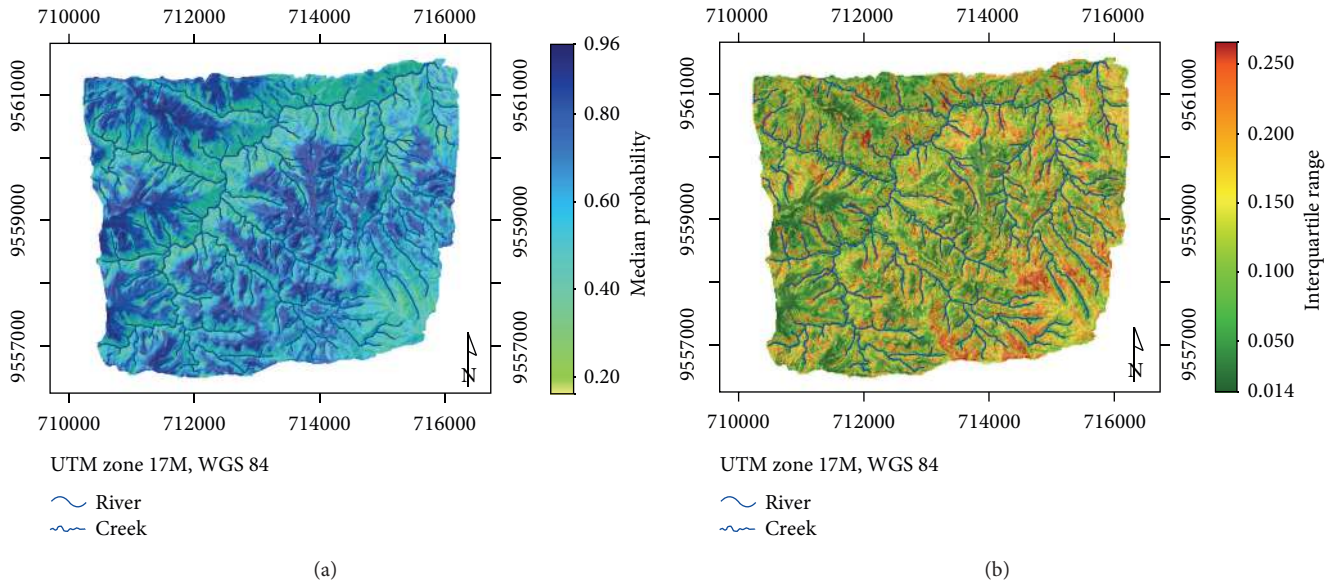
(a)

(b)

FIGURE 6: Median occurrence probability of hydromorphic properties (a) and interquartile range (b) (overlaid hillshading from north).

(i) Occurrence probability of stagnic properties: 10 m DTM, model 1 (no predictor selection, no tuning).

(ii) Stagnic layer thickness: 10 m DTM, model 4 (predictor selection, tuning of mtry).

(iii) Organic layer thickness: 20 m DTM, model 4 (no predictor selection, tuning of mtry).

Of these three models, a combination of tuning of mtry and sampsize might have resulted in an even better model only for the model to predict stagnic layer thickness. The model to predict the occurrence probability of hydromorphic properties reduced the Median RMSE by 18%, the model to predict stagnic layer thickness reduced it by 3%, and the model to predict organic layer thickness reduced it by 11% compared to using the data mean for prediction. That higher predictor resolution resulted in better models to predict water stagnation was also reported by Chaplot et al. [12]. According to Campling et al. [51], vegetation indices and terrain parameters have a complementary role in predicting soil drainage classes. Hence, classified satellite image information could improve model performance and will, therefore, be included in future modelling approaches.

3.3. Digital Soil Maps. The digital soil maps display the soil parameters' distribution function for every point in the landscape. While the median (Figures 6(a), 7(a), and 8(a)) displays the spatial prediction estimate, the interquartile range (Figures 6(b), 7(b), and 8(b)) provides a spatial uncertainty estimate due to the data.

The digital soil map of the median occurrence probability of stagnic properties and its interquartile range is shown in Figures 6(a) and 6(b). The best of the 12 models predicted by terrain parameters of 10 m resolution with no tuning and no predictor selection was selected. For the development of the digital soil map of the occurrence probability of

stagnic properties, all terrain parameters were included. This indicates that it is the complex pattern of climate (altitude, PISR), water accumulation (curvature, convergence, KRAA), water discharge (slope, KRAS), the insulating effect of the heterogeneous geomorphology with the ridge-side valley structure in particular (TRI, normalised height, valley depth), and the wind effect (wind effect, aspect) which lead to the distribution pattern of stagnic soil properties within the investigation area.

The spatial pattern of stagnic properties occurrence probability in Figure 6(a) follows that described by Ließ et al. [48] with a minimum median probability of 0.2 throughout the area and a particularly high probability between 2100 and 2500 m a.s.l. The lower probability below 2100 m a.s.l. must be attributed to the higher inclination that supports a higher discharge of surface and subsurface flow and the higher bulk soil density [52]. In contrast, particularly the flat platform-like areas above 2100 m show a much higher probability. The lower probability above 2500 m a.s.l according to Ließ et al. [48] can be attributed to a higher soil hydraulic conductivity due to a sandier soil texture [5] and therefore less chance for the development of stagnic soil properties. A low interquartile range, <0.1 for 99% of the area (Figure 9(a)), shows that the dataset is well suited to model the spatial pattern of hydromorphic properties within this area. However, particularly some parts along the upper mountain ridges (Figure 6(b) south-eastern part) display a rather high interquartile range. This shows that the model is better suited for certain parts of the landscape. However, another possible explanation is that the degree of soilscape complexity for certain geomorphological positions might be higher than for others.

The model to regionalise stagnic layer thickness is less stable than the model to predict the horizon's occurrence probability. This is indicated by the rather high interquartile ranges in Figure 7(b). Still 70% of the area displays a range
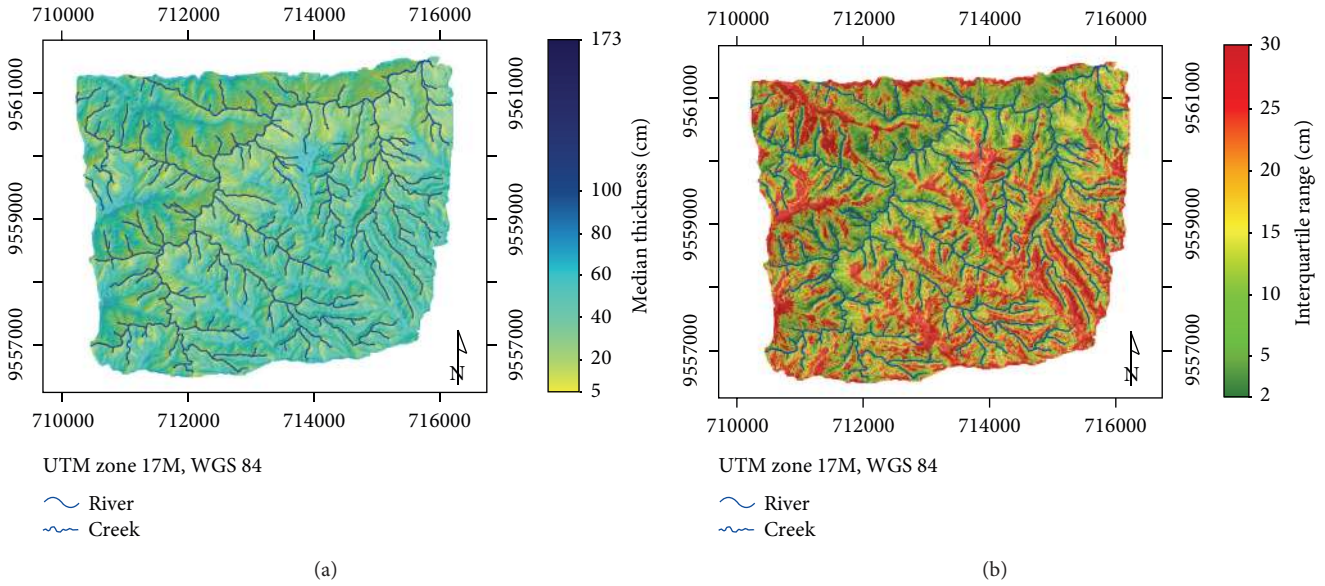
FIGURE 7: Median thickness of stagnic soil horizon (a) and interquartile range (b) (overlaid hillshading from north).
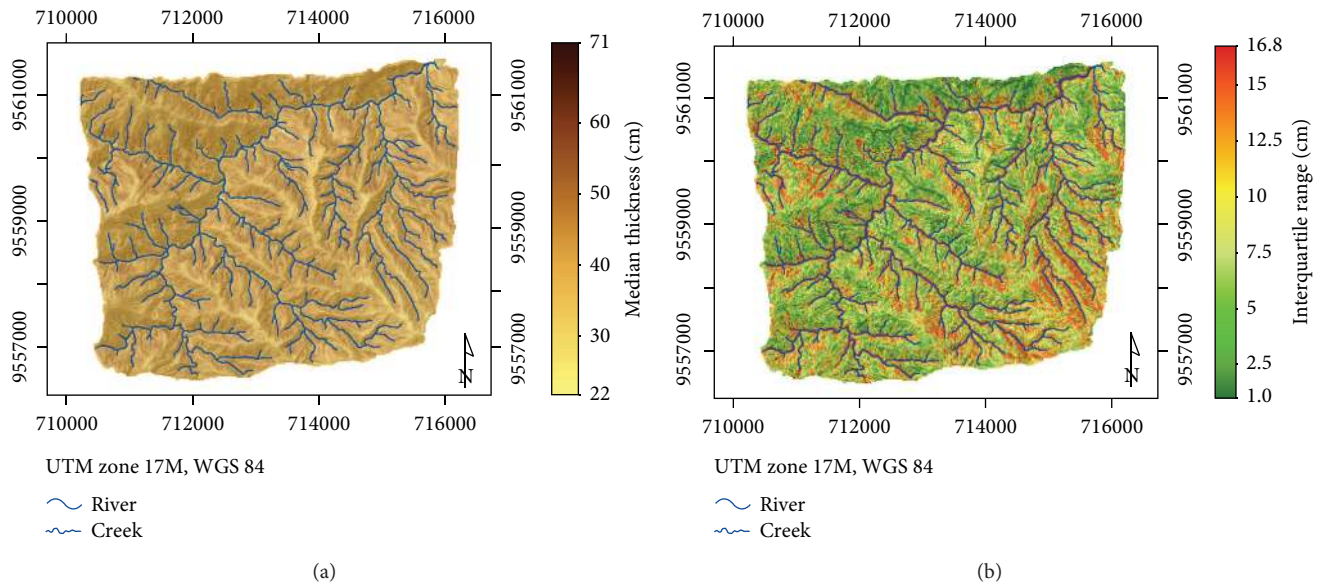

FIGURE 8: Median organic layer thickness (a) and interquartile range (b) (overlaid hillshading from north).

<6 cm (Figure 9(b)). The vertical development of the soil profile is less influenced by surface processes. According to Park and Vlek [53], soil attributes whose vertical distribution is strongly determined by vertical pedogenesis or unknown factors are poorly modelled by environmental variables. Therefore, terrain attributes can only explain horizon thickness to some extent. Bauer [6] limited slope-parallel subsurface flow within the research area to the topsoil (stagnic layer). The frequent change of parent material within one soil profile [5] might be the reason why stagnic layer thickness cannot be explained by geomorphology alone.

The chosen model version includes prepredictor selection and the tuning of mtry so that we assume that at least some of the predictors are not or are less suitable for predicting the stagnic layer thickness. The thickest stagnic layers >40 or even >60 cm are found along the mountain ridges, with decreasing thickness while proceeding down slope towards the side valley creeks. In contrast to the occurrence probability of stagnic properties, this pattern seems to apply throughout the area.

For the development of the digital soil map of the organic layer thickness, all terrain parameters were included. This means that the assumptions about the predictors were reasonable. However, from the final model it is not clear in which way the predictors influence the spatial pattern of the organic layer thickness. The digital soil map is shown in Figure 8. Its uncertainty expressed as the interquartile range (Figure 8(b)) according to the dataset is very low;
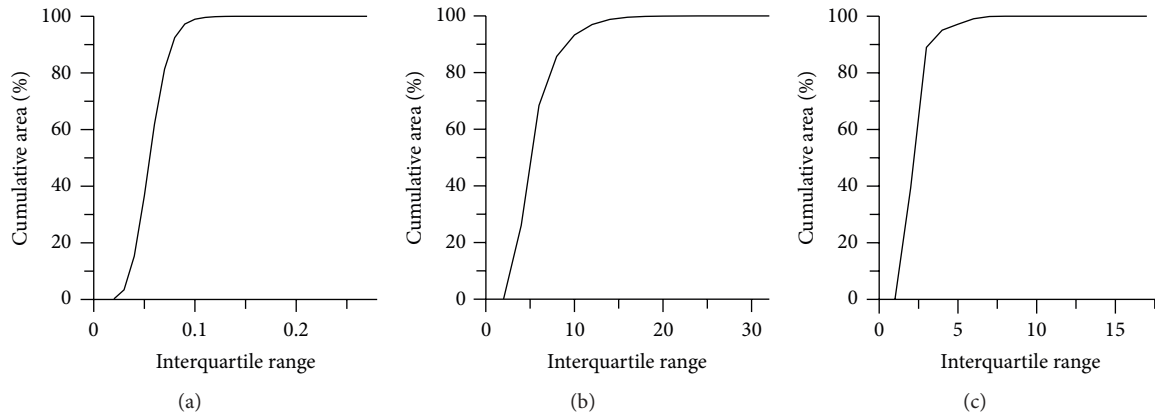
FIGURE 9: Covered cumulative area percentage in dependence of the interquartile range. (a) Occurrence probability of stagnic properties, (b) stagnic layer thickness, and (c) organic layer thickness.

that is, different data subsets predict a similar organic layer thickness. 97% of the area displays an interquartile range <5 cm (Figure 9(c)). This indicates a stable model.

The thickest organic layers are not found along the mountain ridges which are supposed to be the wettest due to their exposedness and low slope angle, resulting in the highest stagnic layer thickness (and probability). The organic layer thickness rather seems to be the highest on mid slope positions, decreasing towards the creeks and towards the crests. Water logging surely limits organic layer decomposition, but is not the only cause. A higher direct solar radiation and wind exposure might favour decomposition rates along the exposed and mostly flat mountain ridges. In addition, wind might also be responsible for a lower litter fall rate on these sites. Accordingly, apart from prediction parameters indicating rainfall (altitude) soil water accumulation (plan curvature, KRAA, SWI, TRI) and water discharge (slope), solar radiation (PISR) and wind effect are also important. Last but not least, the relative slope position (normalised height, valley depth) is a good indicator for organic layer thickness.

With the overall median organic layer thickness >21 cm and 46 area % even ≥40 cm, the area is influenced by paludification, controlled by factors such as climate, geomorphology, and soil water stagnation. International mire classification [54] acknowledges soligenous surface flow mires, the so-called sloopy fens. German Bavarian classification [55] is more precise in acknowledging the fact that sloping mires are influenced by rainwater and shallow subsurface flow at the same time and refers to them as soliombrogen sloping mires. However, information on tropical mountain mires is still scarce. Chimner and Karlberg [56] state that tropical mountain peatlands unlike lowland peatlands are covered by cushion plants, bryophytes, and herbaceous. We can now report that the mires within the tropical Andes of southern Ecuador are also found within the tropical mountain forest zone at 1800–2800 m a.s.l. and under páramo vegetation starting above c. 2800 m a.s.l.

Furthermore, organic layer thickness does not increase with altitude as was concluded by studies based on a dataset with a smaller spatial coverage [13, 47] and is usually explained by a decrease in organic matter turnover and limitation in nutrient supply due to decreasing mean temperatures and increasing rainfall [1, 20, 57].

The landslides occurring with high frequency within the investigation area might give one possible explanation. However, old landslide positions, covered again by dense forest, were sampled by chance and make up less than 5% of the data. Furthermore, open landslide scars, which are visible on aerial photographs, are mainly found at upper slope positions and on steep slopes and therefore cannot explain the low organic layer thickness along the ridges which are left unaffected.

4. Conclusions

Soliombrogen sloping mires do not only occur in tropical páramo landscapes, but also under tropical mountain forest vegetation. However, there is no simple relation between the water stagnation pattern and organic layer thickness, but a complex interaction of various parameters has to be considered. Furthermore, soil organic layer thickness within this tropical mountain landscape does not increase along the altitudinal gradient as was assumed by authors analysing datasets of a smaller spatial coverage.

The Random Forest algorithm was successfully applied to predict the spatial distribution of the occurrence probability of stagnic properties and of the organic layer thickness in this complex tropical landscape that is influenced by landslides. The RMSE as compared to the mean was reduced by 18% by the model to predict the former and 11% by the model to predict the latter. Still stagnic layer thickness was the most difficult parameter to be predicted, as indicated by the low improvement in RMSE by only 3% as was also described by Ließ et al. [48].

Our modeling exercise has shown that position specific density functions of soil properties, characterized by median and interquartile range (Figures 6, 7, and 8), may be an appropriate way of mapping prediction uncertainty. They show that a particular model is better suited for certain parts of the landscape. Furthermore, they might indicate that the degree of soilscape complexity for certain geomorphological

positions might be higher than for others, which needs to be further investigated in the near future.

All considered terrain parameters proved to be important predictors for the spatial water stagnation pattern and the paludification process in this particular landscape and should, therefore, be considered while investigating other tropical mountain landscapes, and—in case they are available—amended by others. As expected, DTM resolution also showed an impact on model performance. Higher resolution was favoured for the prediction of hydromorphic properties, while using a smoothened DTM lead to better results concerning the organic layer thickness models. Tropical soils may vary in their hydromorphic properties within a few meters. Nevertheless, it would have been interesting to extend the range of the GIS raster resolutions beyond 30 m. However, with the available dataset this was not possible. As the spatial distance between many data points is not more than 25 meters, a further terrain smoothing would have only yielded in a higher noise and model uncertainty.

However, no general decision should be made in regard to whether Random Forest tuning is necessary or not. Prepredictor selection with Boruta improved model performance in one of the three predicted soil parameters. The tuning of mtry resulted twice in the overall best model. Sampsize reduced model variation as indicated by the RMSE range and improved model performance in some cases. We, therefore, recommend always testing prepredictor selection and model tuning in order to make sure that the best possible model is chosen. This needs particular emphasis in complex tropical mountain soil landscapes which provide a real challenge to any soil mapping approach but where a supervised learning technique has proven to be successful due to the testing of model tuning and prepredictor selection.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] A. R. Townsend, P. M. Vitousek, and S. E. Trumbore, "Soil organic matter dynamics along gradients in temperature and land use on the island of Hawaii," *Ecology*, vol. 76, no. 3, pp. 721–733, 1995.

[2] W. Wilcke, S. Yasin, A. Schmitt, C. Valarezo, and W. Zech, "Soils along the altitudinal transect and in catchments," in *Gradients in a Tropical Mountain Ecosystem of Ecuador*, E. Beck, J. Bendix, I. Kottke, F. Makeschin, and R. Mosandl, Eds., Chapter 9, pp. 75–85, Springer, Berlin, Germany, 2008.

[3] W. I. J. Dieleman, M. Venter, A. Ramachandra, A. K. Krockenberger, and M. I. Bird, "Soil carbon stocks vary predictably with altitude in tropical forests: implications for soil carbon storage," *Geoderma*, vol. 204-205, pp. 59–67, 2013.

[4] M. Liess, B. Glaser, and B. Huwe, "Digital soil mapping in southern Ecuador," *Erdkunde*, vol. 63, no. 4, pp. 309–319, 2009.

[5] M. Ließ, B. Glaser, and B. Huwe, "Uncertainty in the spatial prediction of soil texture: comparison of regression tree and Random Forest models," *Geoderma*, vol. 170, pp. 70–79, 2012.

[6] F. C. Bauer, *Water flow paths of an undisturbed and landslide affected mature montane rainforest in south Ecuador [Ph.D. thesis]*, University of Bayreuth, Bayreuth, Germany, 2010, http://opus.ub.uni-bayreuth.de/volltexte/2011/761/.

[7] FAO and IUSS Working Group WRB, "World Reference Base for Soil Resources 2006, first update 2007," World Soil Resources Reports 103, 2007.

[8] G. Milne, "Normal erosion as a factor in soil profile development," *Nature*, vol. 138, no. 3491, pp. 548–549, 1936.

[9] D. A. Wysocki, P. J. Schoeneberger, and H. E. LaGarry, "Geomorphology of soil landscapes," in *Handbook of Soil Science*, M. Sumner, Ed., pp. E1–E39, CRC Press, Boca Raton, Fla, USA, 2000.

[10] P. H. Walker, G. F. Hall, and R. Protz, "Relation between landform parameters and soil properties," *Soil Science Society of America Journal*, vol. 32, pp. 101–104, 1968.

[11] J. A. Thompson, J. C. Bell, and C. A. Butler, "Quantitative soil-landscape modeling for estimating the areal extent of hydromorphic soils," *Soil Science Society of America Journal*, vol. 61, no. 3, pp. 971–980, 1997.

[12] V. Chaplot, C. Walter, and P. Curmi, "Improving soil hydromorphy prediction according to DEM resolution and available pedological data," *Geoderma*, vol. 97, no. 3-4, pp. 405–422, 2000.

[13] F. R. Troeh, "Landform parameters correlated to soil drainage," *Soil Science Society of America Journal*, vol. 28, pp. 808–812, 1964.

[14] W. Peng, D. B. Wheeler, J. C. Bell, and M. G. Krusemark, "Delineating patterns of soil drainage class on bare soils using remote sensing analyses," *Geoderma*, vol. 115, no. 3-4, pp. 261–279, 2003.

[15] I. D. Moore, P. E. Gessler, G. A. Nielsen, and G. A. Peterson, "Soil attribute prediction using terrain analysis," *Soil Science Society of America Journal*, vol. 57, no. 2, pp. 443–452, 1993.

[16] S. J. Park, K. K. McSweeney, and B. B. Lowery, "Identification of the spatial distribution of soils using a process-based terrain characterization," *Geoderma*, vol. 103, no. 3-4, pp. 249–272, 2001.

[17] T. F. A. Bishop and B. Minasny, "Digital soil-terrain modeling: the predictive potential and uncertainty," in *Environmental Soil-Landscape Modeling*, S. Grunwald, Ed., pp. 185–213, CRC Press, Boca Raton, Fla, USA, 2006.

[18] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[19] M. B. Kursa and W. R. Rudnicki, "Feature selection with the boruta package," *Journal of Statistical Software*, vol. 36, no. 11, pp. 1–13, 2010.

[20] C. Leuschner, G. Moser, C. Bertsch, M. Röderstein, and D. Hertel, "Large altitudinal increase in tree root/shoot ratio in tropical mountain forests of Ecuador," *Basic and Applied Ecology*, vol. 8, no. 3, pp. 219–230, 2007.

[21] Ließ, M, B. Glaser, and B. Huwe, "Soil-landscape modelling—reference soil group probability prediction in southern Ecuador," in *Principles, Application and Assessment in Soil Science*, E. B. Özkaraova Güngör, Ed., pp. 241–256, InTech Open Access, 2011.

[22] M. Schrumpf, G. Guggenberger, C. Valarezo, and W. Zech, "Tropical montane rain forest soils. Development and nutrient status along an altitudinal gradient in the South Ecuadorian Andes," *Erde*, vol. 132, no. 1, pp. 43–59, 2001.

[23] R. Rollenbeck, "Variability of precipitation in the Reserva Biológica San Francisco/Southern Ecuador," *Lyonia*, vol. 9, no. 1, pp. 43–51, 2006.

[24] O. Planchon and F. Darboux, "A fast, simple and versatile algorithm to fill the depressions of digital elevation models," *Catena*, vol. 46, no. 2-3, pp. 159–176, 2002.

[25] L. W. Zevenbergen and C. R. Thorne, "Quantitative analysis of land surface topography," *Earth Surface Processes & Landforms*, vol. 12, no. 1, pp. 47–56, 1987.

[26] R. Köthe and F. Lehmeier, *SARA-System Zur Automatischen Relief-Analyse*, User Manual, Deptartement of Geography, University of Goettingen, 2nd edition, 1996.

[27] S. J. Riley, S. D. De Gloria, and R. Elliot, "A terrain ruggedness that quantifies topographic -heterogeneity," *Intermountain Journal of Science*, vol. 5, no. 1-4, pp. 23–27, 1999.

[28] N. L. Lea, "An aspect driven kinematic routing algorithm," in *Overland Flow Hydraulics and Erosion Mechanics*, A. J. Parsons and A. D. Abrahams, Eds., pp. 393–3407, London, UK, 1992.

[29] J. Böhner, R. Köthe, O. Conrad, J. Gross, A. Ringeler, and T. Selige, "Soil regionalisation by means of terrain analysis and process parameterisation," in *Soil Classification 2001*, E. Micheli, F. Nachtergaele, and L. Montanarella, Eds., Research Report No. 7, EUR, 20398 EN, pp. 213–222, European Bureau, Luxembourg, 2002.

[30] J. Böhner and O. Antonic, "Land surface parameters specific to topo-climatology," in *Geomorphometry—Concepts, Software, Applications*, T. Hengl and H. I. Reuter, Eds., pp. 195–226, Elsevier, Amsterdam, The Netherlands, 2009.

[31] T. R. Oke, *Boundary Layer Climates*, Taylor & Francis, London, UK, 1988.

[32] J. P. Wilson and J. C. Gallant, Eds., *Terrain Analysis—Principles and Applications*, John Wiley & Sons, New York, NY, USA, 2000.

[33] V. Olaya, "A gentle introduction to SAGA GIS. Edition 1. 1," 2004, http://sourceforge.net/projects/saga-gis/files/.

[34] J. Böhner, K. R. McCloy, and J. Strobl, *SAGA—Analysis and Modelling Application. Göttinger Geographische Abhandlungen*, vol. 115, Geographisches Institut der Universität Göttingen, 2006.

[35] V. Cimmery, "User guide for SAGA (version 2. 0)," 2007, http://sourceforge.net/projects/saga-gis/files.

[36] A. Fries, R. Rollenbeck, D. Göttlicher et al., "Thermal structure of a megadiverse Andean mountain ecosystem in southern Ecuador and its regionalization," *Erdkunde*, vol. 63, no. 4, pp. 321–336, 2009.

[37] J. Homeier, H. Dalitz, and S. W. Breckle, "Waldstruktur und Baumarten im montanen Regenwald der Estación Científica San Franscisco in Südecuador," *Berichte der Reinhold-Tüxen-Gesellschaft*, vol. 14, pp. 109–118, 2002.

[38] M. Oesker, H. Dalitz, S. Günter, and S. Matezki, "Spatial heterogeneity patterns—a comparison between gorges and ridges in the upper part of an evergreen lower montane forest," in *Gradients in a Tropical Mountain Ecosystem of Ecuador*, E. Beck, J. Bendix, I. Kottke, F. Makeschin, and R. Mosandl, Eds., Chapter 18, pp. 267–274, Springer, Berlin, Germany, 2008.

[39] T. Behrens, A.-X. Zhu, K. Schmidt, and T. Scholten, "Multi-scale digital terrain analysis and feature selection for digital soil mapping," *Geoderma*, vol. 155, no. 3-4, pp. 175–185, 2010.

[40] M. Ließ, *Soil-landscape modelling in an Andean mountain forest region in southern Ecuador [Ph.D. thesis]*, University of Bayreuth, Bayreuth, Germany, 2011.

[41] P. Lagacherie and S. Holmes, "Addressing geographical data errors in a classification tree for soil unit prediction," *International Journal of Geographical Information Science*, vol. 11, no. 2, pp. 183–198, 1997.

[42] A. B. McBratney, I. O. A. Odeh, T. F. A. Bishop, M. S. Dunbar, and T. M. Shatar, "An overview of pedometric techniques for use in soil survey," *Geoderma*, vol. 97, no. 3-4, pp. 293–327, 2000.

[43] C. J. Moran and E. N. Bui, "Spatial data mining for enhanced soil map modelling," *International Journal of Geographical Information Science*, vol. 16, no. 6, pp. 533–549, 2002.

[44] C. Strobl, T. Hothorn, and A. Zeileis, "Party on! A new conditional variable importance measure for Random Forests available in the party package. Technical Report Number 050, Department of Statistics, University of Munich," 2009, http://www.stat.uni-muenchen.de.

[45] L. Breiman, J. H. Friedmann, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, CRC Press, Wadsworth, Ohio, USA, 1984.

[46] L. Breiman, "Technical Report for Version 3," 2001, http://oz.berkeley.edu/users/breiman/randomforest2001.pdf.

[47] W. Rudnicki and B. Kursa, "Boruta—a tool for finding significant attributes in information systems. CRAN Reference Manual," 2012, http://cran.r-project.org/web/packages/Boruta/Boruta.pdf.

[48] M. Ließ, B. Glaser, and B. Huwe, "Making use of the World Reference Base diagnostic horizons for the systematic description of the soil continuum—application to the tropical mountain soil-landscape of southern Ecuador," *Catena*, vol. 97, pp. 20–30, 2012.

[49] E. A. Schuur and P. A. Matson, "Net primary productivity and nutrient cycling across a mesic to wet precipitation gradient in Hawaiian montane forest," *Oecologia*, vol. 128, no. 3, pp. 431–442, 2001.

[50] L. Roman, F. N. Scatena, and L. A. Bruijnzeel, "Global and local variations in tropical montane cloud forest soils," in *Tropical Montane Cloud Forests: Science for Conservation and Management*, L. A. Bruijnzeel, F. N. Scatena, and L. S. Hamilton, Eds., pp. 200–226, Cambridge University Press, 2010.

[51] P. Campling, A. Gobin, and J. Feyen, "Logistic modeling to spatially predict the probability of soil drainage classes," *Soil Science Society of America Journal*, vol. 66, no. 4, pp. 1390–1401, 2002.

[52] M. Ließ, B. Glaser, and B. Huwe, "Functional soil-landscape modelling to estimate slope stability in a steep Andean mountain forest region," *Geomorphology*, vol. 132, no. 3-4, pp. 287–299, 2011.

[53] S. J. Park and P. L. G. Vlek, "Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques," *Geoderma*, vol. 109, no. 1-2, pp. 117–140, 2002.

[54] H. Joosten and J. Clarke, "Wise use of mires and peatlands—Background and principles—Including a framework for decision making. International Mire Conservation Group and International Peat Society," 2002, http://www.gret-perg.ulaval.ca.

[55] Bayerisches Landesamt für Umweltschutz, *Moorentwicklungskonzept Bayern (MEK)—Moortypen in Bayern*, Projektgruppe Landschaftsentwicklung + Artenschutz, Kessler Verlagsdruckerei, Bobingen, Germany, 2005.

[56] R. A. Chimner and J. M. Karlberg, "Long-term carbon accumulation in two tropical mountain peatlands, Andes Mountains, Ecuador. Mires and Peat, 3/ article 4, 1–10," 2008, http://www.mires-and-peat.net.

[57] J. Benner, P. M. Vitousek, and R. Ostertag, "Nutrient cycling and nutrient limitation in tropical montane cloud forests," in *Tropical Montane Cloud Forests*, L. A. Bruijnzeel, F. N. Scatena, and L. S. Hamilton, Eds., International Hydrology Series, pp. 90–100, Cambridge University Press, Cambridge, UK, 2010.

Journal of
Waste Management

The Scientific
World Journal

Journal of
Environmental and
Public Health

Scientifica

International Journal of
Ecology

Journal of
Ecosystems

International Journal of
Oceanography

International Journal of
Atmospheric Sciences

Journal of
Marine Biology

Hindawi

Submit your manuscripts at
http://www.hindawi.com

International Journal of
Biodiversity

Journal of
Geological Research

International Journal of
Forestry Research

Applied &
Environmental
Soil Science

Journal of
Climatology

Journal of
Earthquakes

Journal of
Computational
Environmental Sciences

International Journal of
Geophysics

Advances in
Meteorology

Advances in
Environmental
Chemistry