

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Soil-Landscape Modelling – Reference Soil Group Probability Prediction in Southern Ecuador

Mareike Ließ<sup>1</sup>, Bruno Glaser<sup>2</sup> and Bernd Huwe<sup>1</sup>

<sup>1</sup>University of Bayreuth, Department of Geosciences/ Soil Physics

<sup>2</sup>Martin-Luther University Halle Wittenberg, Soil Biogeochemistry  
Germany

## 1. Introduction

Since long, soils are understood as a function of their genetic factors: parent material, relief, climate, organisms and time, a concept first described by Dokutschajew (1883) and better known from Jenny (1941). The complex interaction of these factors activates particular soil forming processes, which in dependence of their intensity and duration, lead to characteristic soil properties. The resulting profile reflects the balance of these processes in its properties (Grunwald, 2006). Soil-landscape modelling uses the knowledge about soil genesis to predict soil distribution in a landscape based on continuously available environmental parameters by statistical models. The early conceptual models have resulted into quantitative soil-landscape models, which do not only make the spatial prediction of continuous soil properties possible, but include model uncertainty. Being at first an unwelcome nuisance that reduced map reliability, gradually soil variation and its unpredictability was seen as a key soil attribute by itself (Burrough et al., 1994).

Soil research within the area of the scientific research station San Francisco in the southern Ecuadorian Andes has been carried out for many years. A first preliminary soil map was provided in 2009 (Liess et al., 2009). This soil map, based on Reference Soil Groups (RSGs) from the World Reference Base for Soil Resources (WRB) (FAO, IUSS Working Group WRB, 2007) does not include prediction uncertainty and neglects Cambisols. However, Cambisols are part of the dataset and have also been described by Yasin (2001) and Wilcke et al. (2002, 2003). Other soils that occur under natural vegetation within the area are Histosols, Stagnosols, Umbrisols and Regosols (FAO, IUSS Working Group WRB, 2007). Histosols were described by Yasin (2001) and Schrumpf et al. (2001) as Haplosaprists according to Soil Taxonomy classification (Soil Survey Staff, 2006). Yasin (2001) investigated forest soils only between 1900 – 2240 m a.s.l., whereas Schrumpf et al. (2001) explored soils along an altitudinal gradient from 1850 – 3050 m a.s.l. Thus, Histosols were found on slope angles varying from 10 – 50° at 1850 – 2700 m a.s.l.; Stagnosols were described between 2080 – 2850 m a.s.l. (Yasin, 2001; Schrumpf et al., 2001; Liess et al., 2009). Umbrisols were assigned by Schrumpf et al. (2001) and Liess et al. (2009).

Prediction of soil types from terrain factors by statistical models is a standard approach within the field of soil-landscape modelling. Lagacherie & Holmes (1997) as well as Moran & Bui (2002) assigned soil classes by CTs based on parameters calculated from a digital

elevation model (DEM). Skidmore et al. (1996), Thomas et al. (1999) and Dobos et al. (2000) spatially predicted soil types from terrain analysis. Furthermore, Gessler et al. (1995), Moore et al. (1993) and Odeh et al. (1994) predicted soil attributes from terrain parameters. Bourennane et al. (2000) and Hengl et al. (2004) regionalised soil horizon and topsoil thickness from a DEM. Several statistical models are available to relate soils to environmental predictors. Bishop & Minasny (2006) compared some of them: Linear, generalized linear (GLM) and generalized additive (GAM) models, classification and regression trees (CART) and artificial neural networks (ANN). Among the considered model types, only ANN were assigned a better predictive power than CART, but lack the ease of use, parsimony, interpretability and computational efficiency that applies for CART.

By extending the dataset of Liess et al. (2009) and constructing various classification trees (CT), we expect to develop a more precise RSG map and include prediction uncertainty by displaying the RSG probability. The investigated soils will be related to terrain parameters by a CT (Breimann et al., 1984) that organises the dataset according to the respective RSG. The tree model can then be used to assign the RSG probabilities to the whole area covered by a DEM.

## 2. Material and methods

### 2.1 Research area

The research area is situated between the provincial capitals Loja and Zamora (Figure 1) in the southern Ecuadorian Andes from 1670 to 3160 m a.s.l. It extends in UTM-Zone 17M from west to east between 710500 and 716000, and from north to south between 9561500 and 9557000 (Figure 1). The San Francisco River divides the area into two parts: The north-west facing slopes south of the river are covered by montane rain forest and subpáramo vegetation above the tree line. Within this area, Homeier et al. (2002) differentiated various

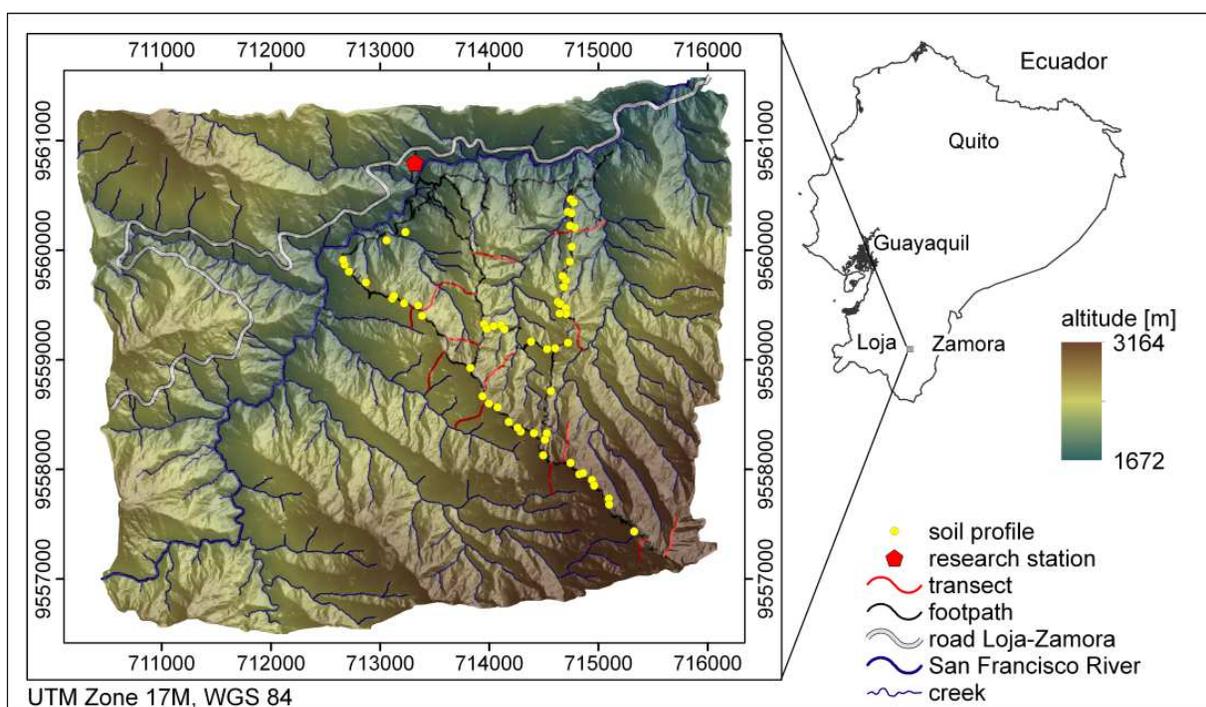


Fig. 1. Research area. Overlaid hill shading with light source from north-east (adapted from Liess et al., 2009)

forest types according to their altitude and position on the ridge or in the valley. The south-eastern facing slopes north of the river are mainly covered by pastures and succession vegetation after fire clearance when sites were left unused. For soil model development, only sites under natural vegetation were considered.

As part of the Chiguinda unit, the research area is lithologically covered by metasiltstones, siltstones and quartzites which are intermixed with layers of phyllite and clay schists (Litherland et al., 1994). Furthermore, it is influenced by the regular occurrence of landslides. Average total annual rainfall increases from 2050 mm at an altitude of 1960 m a.s.l. to approximately 4400 mm at 3100 m a.s.l. (Rollenbeck, 2006). Average air temperature decreases with increasing altitude from 19.4 to 9.4 °C (Fries et al., 2009).

## 2.2 Classification trees

Classification trees (CTs), a method first described by Breimann et al. (1984), were used to relate the RSGs to terrain parameters. It was conducted with the rpart library of the R-Project for Statistical Computing (Therneau & Atkinson, 2003).

In CTs subdivision is based on a categorical response variable, i. e. RSG. The final subsets, also called end nodes, should be as pure as possible. This is done by trying to assign them to only one category in the response variable, e.g. to Histosol. The Gini criterion (Equation 1) is applied as a measure of purity (Breiman et al., 1984). It serves as a decision criterion, to determine which terrain parameter best separates the dataset continuously into always two subsets to create the purest end nodes.

$$gini(t) = 1 - \sum_{i=1}^k P_i^2 \quad (1)$$

The Gini-Index (Equation 1) reaches its maximum in a particular node  $t$  if all categories  $k$  within this node are equally represented. On the other hand, when the probability  $P_i$  is equal to zero for all but one category within any node, the Gini-Index reaches its minimal value. The categorical value accounting for the majority within each end node is then assigned to the corresponding parameter values, indicating the typical position within the landscape (e.g. Liess et al., 2009). However, another option is to assign the percentage of each categorical value within an end node as occurrence probability to the corresponding landscape position. This is of course only justified if the applied sampling scheme guarantees sampling of all landscape positions to an equal extent.

The CT is pruned to avoid overfitting and obviate random variation. To assess model performance, the cross validation error (CV) is calculated. The dataset is subdivided into 10 subsets, and the process is repeated 10 times with 9 parts for model training and the 10<sup>th</sup> part as the evaluation dataset. Eventually, among all trees considered for the final model, the tree with the lowest cross validated error rate is chosen. CV and model pseudo  $R^2$  are calculated. Pseudo stability indices are constructed to satisfy the different interpretations, e.g. explained variance or square of correlation. They are similar to  $R^2$  in that they also range between 0 and 1 and a higher value represents a better adaptation to the data.

## 2.3 Dataset and GIS methodology to gain terrain data

Topographic data for the research area is available on a continuous landscape level. The DEM used to obtain terrain parameters for the establishment of a prediction model of RSG occurrence has 2 m cell size (Liess et al., 2009). For model application, this accuracy was

reduced to 10 m to decrease calculation time. The used terrain parameters include altitude a.s.l., aspect, slope angle, terrain curvature, upslope contributing catchment area and overland flow distance to the channel network (OFD).

Slope angle, aspect and curvature were computed with a 2<sup>nd</sup> degree polynomial fit from Zevenbergen and Thorne (Zevenbergen & Thorne, 1987; Cimmery, 2007). The contributing area was calculated with two methods; (1) based on the Kinematic Routing Algorithm (KRA CA) (Lea, 1992) and (2) based on the Braunschweiger Digital Relief Model (BS CA) (Bauer et al., 1985). In addition to the OFD, the horizontal (HOFD) and vertical (VOFD) overland flow distances were also calculated. The channel network itself was assessed applying the Strahler stream order  $\geq 5$  as initiation threshold (Strahler, 1957). Terrain curvature was computed using directly adjacent cells. Finally, the terrain parameters were calculated and the RSGs were predicted for each individual raster grid cell. The free and open source GIS software, SAGA, was used (Böhner et al., 2006).

The research area was sampled at 367 sites, including 311 auger points and 56 soil profiles. Soil sampling covered 24 sampling classes produced by an overlay of four altitudinal, three slope angle and two aspect classes to guarantee representative area coverage. Transects for auger sampling (Figure 1) were laid according to the catena concept (Milne, 1935) from hilltop to valley bottom. For more detailed information on the applied sampling design, see Liess et al. (2009).

Two methods were used to assign terrain parameters to the soil dataset. On the one hand, the nearest neighbour (n. n.) value was allocated to each soil profile or auger point. On the other hand, a buffer representing the radius of GPS accuracy was placed around the sampled location, and the calculated mean value of the corresponding area was assigned. This assignment was completed for each of the described parameters apart from the slope angle and aspect. These were directly measured in the field. The slope angle and aspect which were computed from the DEM were solely used for model application.

## 2.4 Probability calculation

The probability of each RSG was predicted via a CT which grouped the soil sampling points regarding the existence or absence of that RSG. Thus, the percentage of sampling points assigned to the corresponding RSG in each end node of the tree was used to predict the probability of that RSG. Thereby, the diagnostic properties necessary for assigning the particular RSG were used, whereas the necessary absence of other properties was neglected. This was done in particular to establish a good prediction scheme for Stagnosols. It was decided that the occurrence and thickness of a sufficient stagnic colour pattern and/ or albic horizon is more important than the limitation in organic layer thickness. As a consequence, soils with a 40 cm organic layer displaying also a thick stagnic horizon were classified as Histosols and Stagnosols. Any other proceeding would have made the development of a Stagnosol prediction scheme incomplete and complex.

To sum the individual probabilities and standardize them by relating each RSG to the total probability sum, is one option. This option neglects WRB (FAO, IUSS Working Group WRB, 2007) hierarchy, because all RSGs are competing on an equal level and no soil process is given dominance over another. As a consequence, the probabilities refer to the probability of the diagnostic property necessary for RSG assignation. Later we will refer to these as WRB independent probabilities.

Figure 2 shows the probability calculation scheme based on WRB (FAO, IUSS Working Group WRB, 2007) hierarchy.

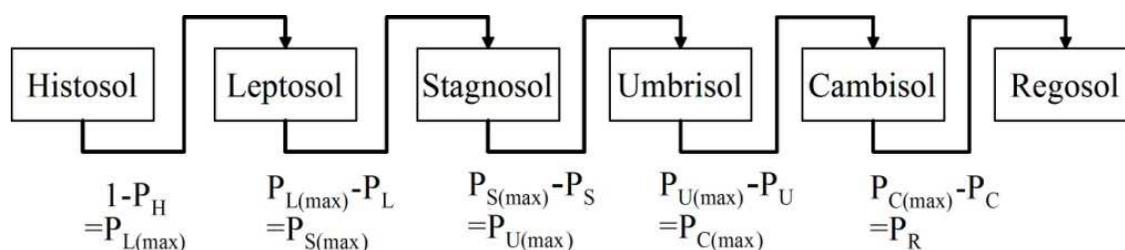


Fig. 2. Hierarchical calculation scheme for the maximum possible probability of each RSG according to WRB hierarchy.  $P_X$  is the actual probability of the respective RSG: H Histosol, L Leptosol, S Stagnosol, U Umbrisol, C Cambisol, R Regosol.  $P_{X(max)}$  is the maximum possible probability of the RSG

It is used to calculate the maximal possible probability for each RSG from the probability predicted by the CTs. Maximal Leptosol probability is left after subtracting Histosol probability from 1. Maximal Stagnosol probability is left after also subtracting the actual Leptosol probability and so on. Equation 2 shows the calculation of the actual probability,  $P_X$ , according to the CT probability,  $P_{X(tree)}$ , and the maximal possible probability,  $P_{X(max)}$ .

$$P_{x(max)} \cdot P_{x(tree)} = P_x \quad (2)$$

### 3. Results and discussion

#### 3.1 Classification tree models and digital soil maps

Figure 3 presents the CT models to predict Histosol, Leptosol and Stagnosol occurrence probability from nearest neighbour (n. n.) and mean terrain values.

The RSG Histosol is assigned to soils with an organic layer  $\geq 40$  cm (FAO, IUSS Working Group WRB, 2007). Its probability within the research area was found to depend on two hydrological parameters (Figures 3a and 3d): KRA CA and VOFD. Probability is predicted with at least 0.2 (Figure 3a, d) throughout the research area. The highest probability (0.87) as predicted by n. n. relief values (Figure 3a) was obtained for small catchments (KRA CA < 258 m<sup>2</sup>) within a distance of 14 – 23 m from the channel network. Though, probabilities are also high, 0.65, for small catchments (KRA CA < 258 m<sup>2</sup>) within a VOFD of 54 – 176 m. The latter is a more conservative prediction, since it is based on 206 sampled sites and not only 15 as for the first differentiation criteria (Figure 3a). Sites seem to coincide in some parts with upper slope areas and ridges (Figure 4a).

Prediction by mean terrain values (Figure 3d) again shows high probabilities in similar landscape positions, i.e. for small catchments < 254 m<sup>2</sup> from 54 – 175 m VOFD (0.65) and < 26 m VOFD (0.70). The former is the safest prediction similar to the Histosol prediction from n. n. terrain values (188 sampled sites). Areas likely to be covered by Histosols with this 0.65 probability are again found along ridges. In contrast to the CT from n. n. terrain values, the highest probabilities, 0.85, by mean relief values (Figure 3d) are assigned to large catchments ( $\geq 254$  m<sup>2</sup>) with a VOFD from 103 – 145 m, dominating in dark colours as broad belts at 103 m distance around the creeks (Figure 4b). This also accounts for the major difference between the two models (Figure 4c). But since the corresponding end node in the tree model (Figure 3d) is only supported by 13 sampled sites, this finding is not representative for the research area.

Leptosols refer to soils limited to 25 cm depth by continuous rock (FAO, IUSS Working Group WRB, 2007). During soil sampling continuous rock was rarely attained, and refusal typically occurred at the C horizon. This made the establishment of a model predicting soil

depth to continuous rock impossible. Therefore, to calculate Leptosol occurrence probability expert knowledge was applied in addition to the CT methodology. From field work and data review it was known that Leptosols are found on steep slopes  $\geq 50^\circ$  and close to the creeks at approximately  $< 20$  m HOFD. Other soils, which occurred at the same landscape positions with even higher probability, were excluded for model development. Afterwards, they were included again to calculate the probabilities of the tree end nodes. This explains the rather untypical appearance of the Leptosol CTs (Figures 3b and 3e). Usually, for any final subdivision into two end nodes, one of them would always display a probability  $> 0.5$  and the other  $< 0.5$ . However, for the reason of adding more datasets after tree development this is not the case. This procedure was necessary in order to develop a reasonable model and account for true probabilities. Leptosol CTs established with n. n. and mean terrain values are very similar. In the already mentioned positions, Leptosol probability was assumed 0.30 – 0.36 (Figure 3b and 3e).

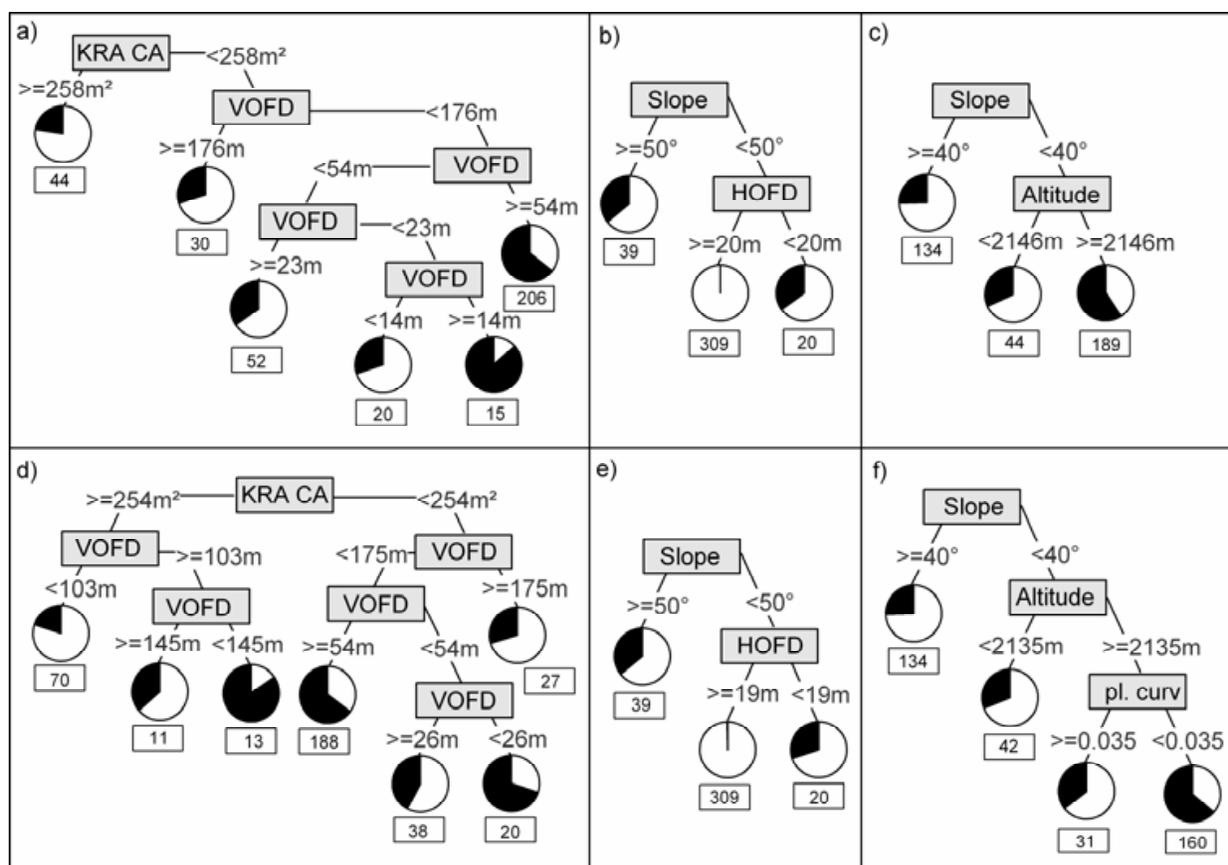


Fig. 3. Classification trees predicting RSG probability. The pie charts' black parts represent the occurrence probability in the corresponding landscape positions. The numbers in the boxes underneath the charts refer to the number of sampling sites used for the probability prediction in each end node. Prediction by n. n. terrain values: a) Histosol probability, b) Leptosol probability and c) Stagnosol probability. Prediction by mean terrain values: d) Histosol probability, e) Leptosol probability and f) Stagnosol probability. (KRA CA = upslope contributing catchment area according to the Kinematic Routing Algorithm, VOFD = vertical overland flow distance, HOFD = horizontal overland flow distance, pl. curv = plan curvature)

The lighter colours in Figure 4b compared to Figure 4a are due to the fact that a probability of 0.20 ( $< 103$  m VOFD, Figure 3d) falls into a smaller mapping class than 0.23 (Figure 3a) in

the map layout. The similarity between the two models for the mentioned sites is indicated by yellow colours in Figure 4c. The sites mapped in red colours refer to a 0.1 – 0.3 higher probability as predicted by mean relief values. Comparison of the two tree models (Figures 3a and 3d) shows that differences are not higher than 0.13. The models differ only by a probability of 0.03 – 0.13, neglecting the mentioned 13 sites.

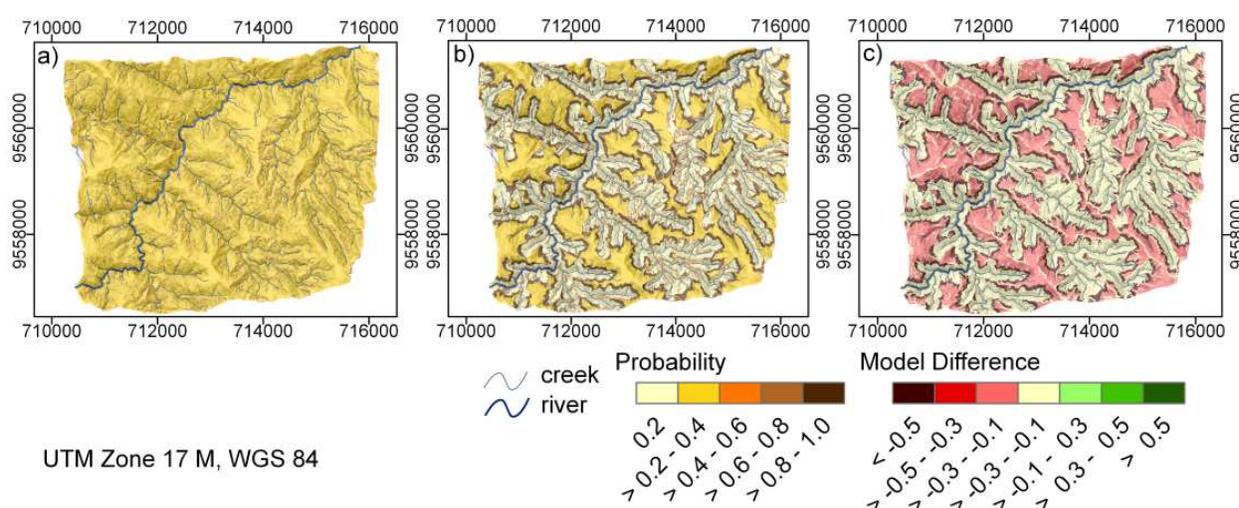


Fig. 4. Maps of Histosol occurrence probability (Overlaid hill shading with light source from north-east): a) Prediction by n. n. terrain values, b) prediction by mean terrain values and c) model difference

Figures 5a and 5b show the Leptosol probability distribution within the research area after model application. With the inclusion of WRB (FAO, IUSS Working Group WRB, 2007) hierarchy, Leptosol probability also depends on Histosol probability. But since Histosol probability close to the creeks ( $< 103$  m VOFD) is predicted with only 0.2, model from mean relief values (Figure 3d), and 0.3, model from n. n. relief values (Figure 3a), it does not influence Leptosol probability much for those sites. Model difference regarding prediction by n. n. and mean terrain values (Figure 5c) is always  $\leq \pm 0.1$  (0.05); including WRB hierarchy (FAO, IUSS Working Group WRB, 2007), model difference (Figure 5f) is increasing (hardly recognisable in the map). Model difference regarding probability predicted directly by the CTs and probability being calculated based on WRB hierarchy (Figures 5g and 5h) shows a similar picture. The difference between the WRB independent and dependent prediction by n. n. values (Figure 5g) is  $\leq \pm 0.1$ , but higher regarding the prediction difference by mean terrain values (Figure 5h).

Stagnosols are “soils exhibiting hydromorphic features for some time during the year in some part within 50 cm of the mineral soil surface and show a stagnic colour pattern and/or an albic horizon in half or more of the soil volume” (FAO, IUSS Working Group WRB, 2007). Planosols are classified by similar diagnostic properties, but in addition display an abrupt textural change, which could not be confirmed for the investigated soils. Stagnosol probability is predicted throughout the research area with at least 0.25 (Figures 3c and 3f). The probability in both models depends on slope angle and altitude. It is higher on slopes  $< 40^\circ$ . Above 2146 m a.s.l. for the prediction by n. n. and above 2135 m a.s.l. by mean terrain values, the probability increases even further. While curvature is of no importance for Stagnosol probability prediction by n. n. relief values, mean terrain values assign an even

higher probability for concave plan curvature with 0.64. Landscape positions  $< 2146$  m a.s.l. for prediction by n. n. and  $< 2135$  m a.s.l. by mean terrain values, and high slope angles account for the lowest probability of Stagnosols.

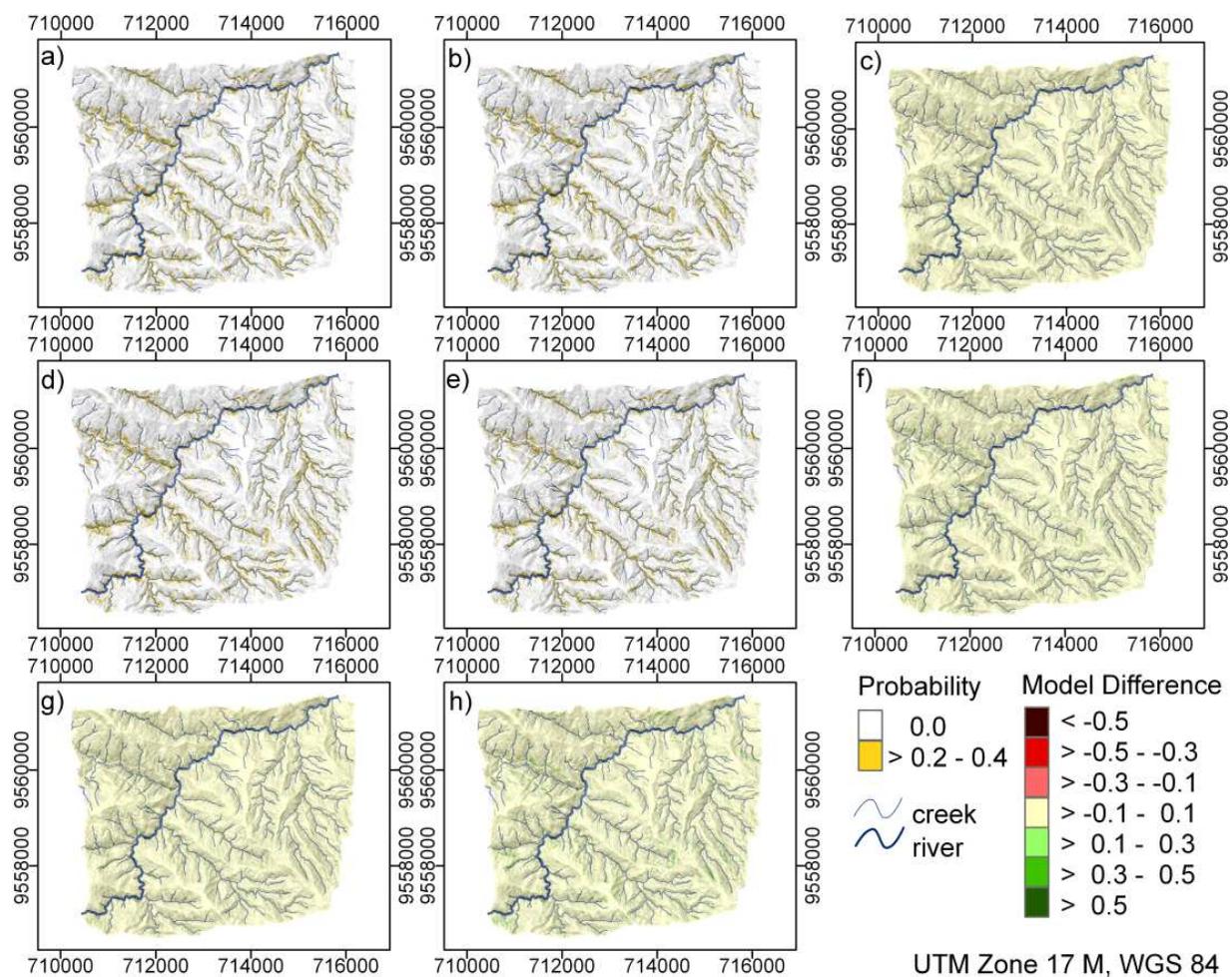


Fig. 5. Maps of Leptosol occurrence probability (Overlaid hill shading with light source from north-east). Independent on WRB hierarchy: a) prediction by n. n. terrain values, b) prediction by mean terrain values and c) model difference. Dependent on WRB hierarchy: d) prediction by n. n. terrain values, e) prediction by mean terrain values and f) model difference. Difference between independent and WRB hierarchy dependent prediction: g) n. n. terrain values and h) mean terrain values

Model application to the research area is shown in Figure 6. Stagnosols reach higher probabilities by the mean terrain values model (Figure 6b) compared to the prediction from n. n. terrain values (Figure 6a). Figures 3c and 3f show that the difference between the probability prediction by n. n. and mean relief values (Figure 6c),  $+ 0.1 - 0.3$ , is not due to this higher Stagnosol probability on high altitudes as predicted by mean terrain values on concave sites. This difference accounts for only 0.05. However, it is due to the reduced probability assigned to convex sites  $\geq 2135$  m a.s.l. (0.24 difference). As a conclusion to this, the two models are quite similar, mainly differing by the dependence on curvature, which is not included in the model from n. n. relief values. Including WRB (FAO, IUSS Working Group WRB, 2007) hierarchy in the probability prediction, a site classified as Histosol or

Leptosol cannot be classified as Stagnosol. Accordingly, Histosol probability reduces Stagnosol probability to a perceptible extent (Figures 6d and 6e). Figures 6g and 6h show that these differences account for 0.1 to 0.3 for most of the research area with the prediction by n. n. terrain values (Figure 6g) still yielding less differences in the lower altitudes compared to the prediction by mean terrain values (Figure 6h). Differences between the two models are extended while including WRB hierarchy (Figure 6f), compared to that being independent of WRB hierarchy (Figure 6c).

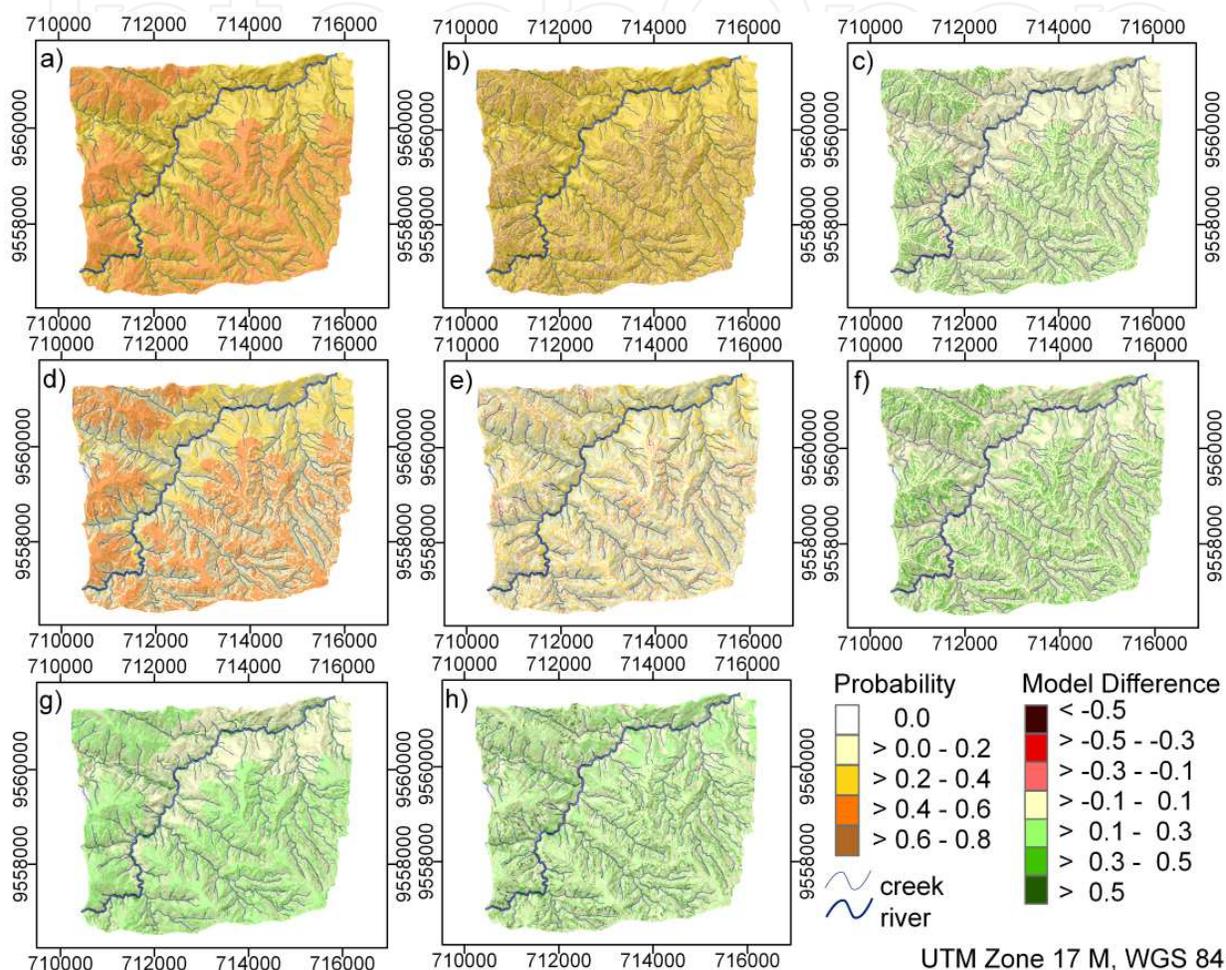


Fig. 6. Maps of Stagnosol occurrence probability (Overlaid hill shading with light source from north-east). Independent on WRB hierarchy: a) prediction by n. n. terrain values, b) prediction by mean terrain values and c) model difference. Dependent on WRB hierarchy: d) prediction by n. n. terrain values, e) prediction by mean terrain values and f) model difference. Difference between independent and WRB hierarchy dependent prediction: g) n. n. terrain values and h) mean terrain values

CTs for Umbrisols, Cambisols and Regosols cannot be provided. Umbrisol prediction was impossible, since the used dataset contains only 7 Umbrisols among 367 sampled sites and is not enough to gain a clear prediction scheme. Furthermore, not all but some of the determined Umbrisols are situated within the accumulation zone of former landslides so that an additional variable to predict their occurrence would be necessary. Cambisols and Regosols, on the other hand, are rather unspecific RSGs which makes their prediction difficult. Cambisols need a cambic horizon, but apart from that they are rather determined

by the absence of diagnostic criteria that would classify the soil for another RSG. Regosols are even worse, since they do not have any characteristic on their own, but refer to all soils that do not classify as another RSG.

### 3.2 Model performance and uncertainty

Overall CT model performance is limited (Table 1). Terrain attributes can likely only explain RSG distribution to a limited extent within this mountainous tropical landscape. Unfortunately, no information is available about parent material distribution, but rapid bedrock changes were discovered during field work. The profound influence of landslides causes shifts in soil material and mixes it with rock material, leading to quite different soil properties. Although there has been a landslide inventory based on visible landslide scars on a time series of aerial photographs from 1962 to 1998 (Stoyan, 2000), most former landslides remain hidden under the regrown dense forest cover as was experienced during field work.

RSG	Terrain Parameters	Model Pseudo R <sup>2</sup>	CV Pseudo R <sup>2</sup>
Histosol	nearest neighbour	0.34	0.22
	mean	0.35	0.21
Stagnosol	nearest neighbour	0.22	0.19
	mean	0.28	0.13

RSG= Reference Soil Group, CV = cross validation

Table 1. Model quality of classification trees to predict Histosol and Stagnosol probability

CTs in general have certain disadvantages: (1) They are very dependent on the dataset used, i.e. some sample points more or less may lead to rather different models and (2) they predict abrupt values due to the grouping into end nodes. A continuous probability distribution of the RSGs in reality therefore is replaced by some probability classes according to Figure 3.

What makes WRB RSG prediction in general problematic is the character of the WRB itself. Assignment of some RSGs requires exceeding an absolute (Histosols) and for others a relative (Stagnosols) thickness value of a diagnostic horizon. If a soil has an organic layer  $\geq 40$  cm, it is classified as Histosol independent of its mineral properties. If the organic layer is 1 cm less, these mineral properties abruptly become important. Relating the extent of the stagic horizon to soil depth obviously is not characteristic enough to allow for a good model relating the Stagnosol occurrence pattern to terrain parameters. This is probably the reason why model accuracy is limited. As a consequence the low R<sup>2</sup> are not considered as a problem, but as a natural phenomenon in predicting complex entities such as RSGs.

Furthermore, the calculated CT R<sup>2</sup> refers to a one value prediction. As was described earlier, a CT model usually assigns the category which forms the majority within each end node to the respective landscape position. It does not consider other categories assigned to that end node as classification possibility, but neglects them. Any soil map has a certain degree of uncertainty. Usually boundaries between soil units are drawn according to expert knowledge or GIS interpolations. However, the degree of uncertainty which is a logical phenomenon in any below ground investigation usually is not included within the soil map. The new generation of digital soil maps provides a new development in this area. Accordingly, our digital soil maps include this model uncertainty through assigning RSG occurrence probabilities instead of unique values. Other authors mainly used fuzzy-logic to include this uncertainty, e.g. McBratney & De Gruiter (1992), Hannemann (2010).

Another aspect to be considered, is that generally soil maps are gained on a much larger scale. Lagacherie & Holmes (1997) use a spatial resolution of 50 m, Moran & Bui (2002) use 250 m. Therefore, the small scale, 10 m resolution, in our soil maps might be another reason for the low  $R^2$ . The soils within the research area change within a few meters radius as typical for tropical soils. Accordingly, the highest possible resolution was used. This way low scale soil variability is included within the models, which would be neglected while working on a larger scale. To conclude, the size of the applied dataset is not enough to represent the investigated soil-landscape at this high precision.

### 3.3 Comparison with earlier soil map

A RSG probability prediction is also possible from a single CT which predicts all RSGs at once. Liess et al. (2009) established such a CT for the research area (Figure 7), but did not predict probabilities from it. The percentage of the RSGs within each end node of this tree was interpreted as occurrence probability for the RSGs according to the related landscape position and compared it to the findings from the various CTs of this study. The difference between RSG probability by the tree model from Liess et al. (2009) and our predictions is displayed in Figure 8. The first column maps the RSG probabilities according to Liess et al. (2009), the second column presents the differences between the latter and our prediction from n. n. relief values (WRB dependent), and the third column shows the differences regarding the prediction from mean relief values.

The model from Liess et al. (2009) (Figure 7) assigned a very high Histosol probability with 0.6 – 0.8 to about half of the research area. For some sites the predicted probability was even higher. In our new model, Histosol probability was less, 0.2 – 0.4 for most of the area (Figure 4b), but continuous on all sites with at least 0.2 (Figure 3a and 3d). It was shown that Histosol probability is high within some landscape positions and for a VOFD from 54 – 175 m this is supported by a high number of sampled sites. In contrast to this, the end nodes in the tree model from Liess et al. (2009) mostly contain only a very limited number of sampling sites, e.g. the end nodes that predict particularly high Histosol probabilities ( $\geq 0.8$ ) only contain 12 – 15 sampled sites. The end node with the most sites predicting Histosol probability with 0.78, refers to landscape positions in small catchments < 214 m HOFD, similar to our findings. The importance of the catchment size as first subdividing variable for model development was confirmed. For smaller catchment sizes, i.e. sites through which a smaller area discharges, Histosol occurrence is more likely.

Leptosols were predicted with low probability on steep slopes and close to the creeks (< 20, < 19 m HOFD). The latter is confirmed by Liess et al. (2009) who predicted Leptosols < 21 m HOFD, but with a high probability of 0.71 (Figure 8d). 0.71 of 7 sampled sites that are contained within the respective end node no. 7 (Figure 7) are 5 sampled sites. To use only five Leptosol sites to predict such a high probability seems unreasonable. On steep slopes, especially in an area influenced by landslides, soils have less chance to develop. Hence, it is no surprise to find Leptosols in these landscape positions. Close to the creeks soil material is probably removed downslope within the channel system during times of high rainfall; through these sites a high amount of water discharges due to a high contributing catchment area. On many sites the organic layer directly overlies continuous rock.

Stagnosols were predicted with a higher probability by n. n. relief values compared to the model from Liess et al. (2009) (Figure 8h). This is due to the fact that Stagnosols were predicted as all soils that display sufficient stagnic properties, but it was neglected that some of them carry a sufficiently thick organic layer to qualify as Histosols. Stagnic properties and



used a subset of our dataset, predicted the RSGs by mean relief values, too. Stagnosol probability increases above an altitude of 2146 m a.s.l. on slope angles < 40°. An increase in Stagnosol abundance with increasing altitude and decreasing slope angle was also described by Liess et al. (2009). Schrumpf et al. (2001) stated an increase in hydromorphic properties with increasing altitude and designated soils as Humaquepts (Soil Survey Staff, 2006). The increase with altitude can be attributed to the increasing rainfall (Rollenbeck, 2006). Lesser steep slope angles account for a slower discharge.

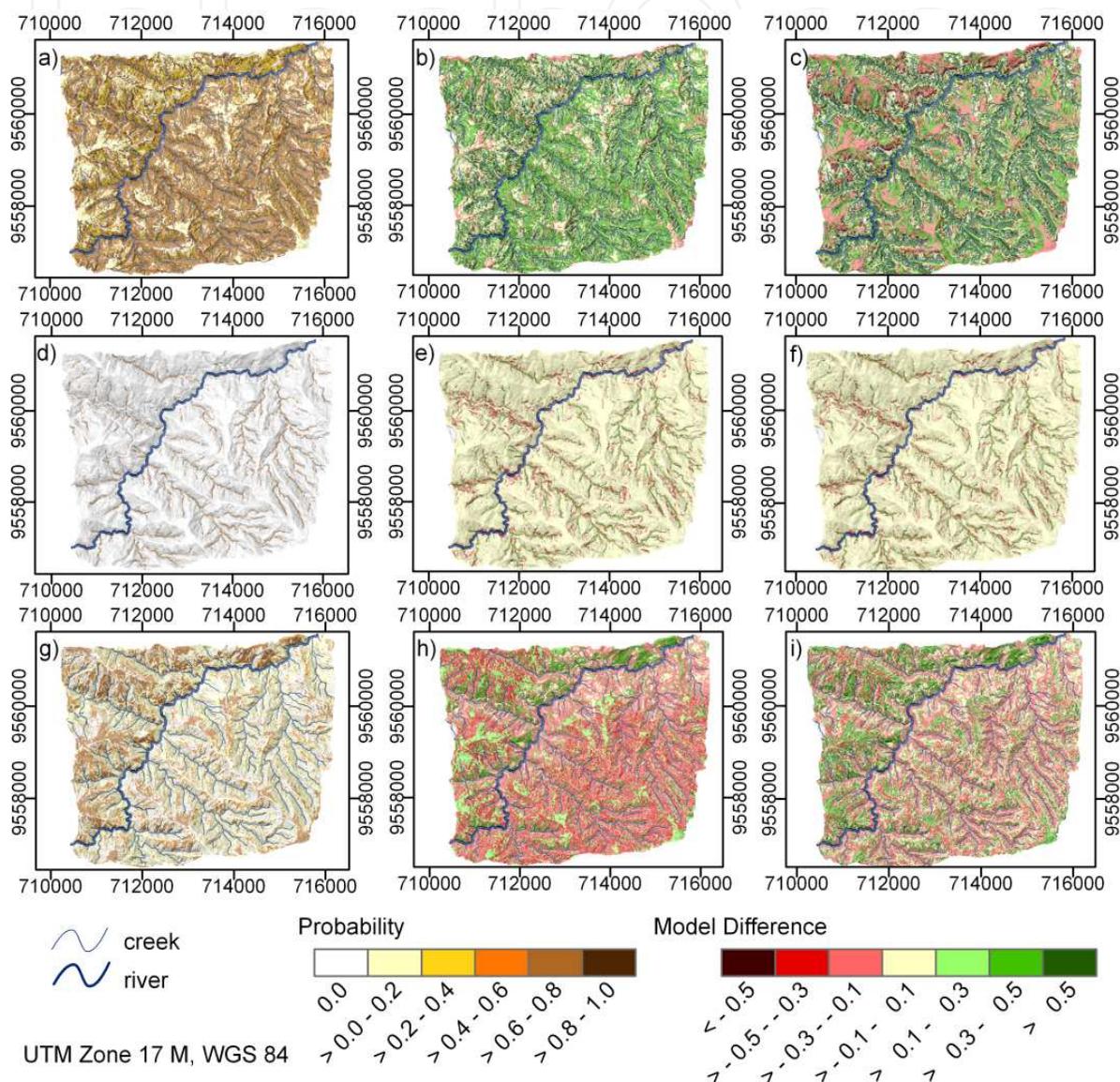


Fig. 8. RSG probability prediction by the simple tree model from Liess et al. (2009) (column 1: a, d and g) and calculated difference in probability prediction between that model and the WRB dependent model from n. n. (column 2: b, e and h) and mean terrain values (column 3: c, f and i). Histosol (1<sup>st</sup> row), Leptosol (2<sup>nd</sup> row) and Stagnosol (3<sup>rd</sup> row). Overlaid hill shading with light source from north-east

We assume the RSG probability predicted by various CTs, to better represent soil reality within the research area, since the dataset does not consist of all RSGs to an equal extent so that some are preferred over others during the tree subdivision process. Furthermore, the

multiple CTs rather predict probabilities of soil diagnostic properties, which can occur simultaneously at one site within the soil profile. Accordingly, the model from Liess et al. (2009) overestimated Histosol probability for most sites as can be seen by the mainly green colours in Figure 8b and c. However, at the same time it underestimated Stagnosols in most of the area as can be deduced from the prevailing red colours in Figures 8h and i. In a similar way, Leptosols are overestimated by the model from Liess et al. (2009).

#### 4. Conclusions

Models adapted for n. n. compared to those adapted for mean terrain values showed only minor differences. We conclude that predicting all RSGs at once is not as good as predicting each RSG on its own by a CT. The dataset does not consist of all RSGs to an equal extent, so some RSGs are preferred over others during the tree subdivision process.

Model performance might be improved by choosing a lower resolution to exclude small scale diversity, reducing model dependence on the dataset, applying a different statistical model or predicting soil properties instead of the complex RSG entities. However, further research is needed to prove these assumptions.

Model uncertainty in the digital soil maps is represented by the occurrence probabilities of the RSGs. Probabilities of various RSGs at the same landscape position can be understood as competing RSGs. But the probabilities of the various RSGs can also be interpreted as a soil composed of the various RSGs, i.e. various diagnostic horizons or various soil processes running simultaneously or successively as has been part of soil genesis theory for a long time (Simonsen, 1959; Schelling, 1970). Thereby, this provides a good means to acknowledge inter-relations between the RSGs. An even better chance to acknowledge this would be the prediction of the diagnostic properties necessary for WRB classification by themselves. In accordance with McBratney & De Gruiter (1992), who thought to improve the existing soil classification systems via fuzzy sets, we would like to contribute the above-mentioned ideas to the development of a continuous soil systematisation system.

#### 5. Acknowledgements

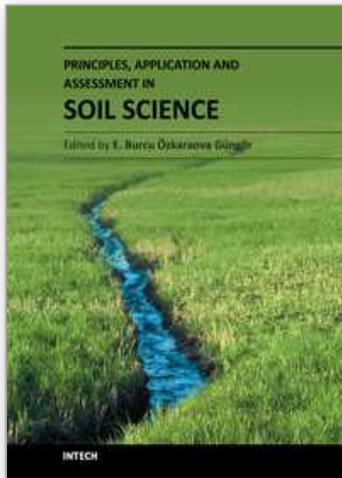
The authors are indebted to the German Research Foundation (DFG) for funding the study in the framework of the Research Unit FOR816. Logistic support of the foundation Nature and Culture International (NCI, San Diego - Loja) is gratefully acknowledged. Furthermore, we would like to thank Christopher L. Shope for English language revision.

#### 6. References

- Bauer, J., Rohdenburg, H., & Bork, H.-R. (1985). Ein digitales Reliefmodell als Voraussetzung für ein deterministisches Modell der Wasser- und Stoff-Flüsse. In: *Landschaftsgenese und Landschaftsökologie H. 10, Parametereaufbereitung für deterministische Gebiets-Wassermodelle, Grundlagenarbeiten zur Analyse von Agrar-Ökosystemen*, Bork, H.- R., Rohdenburg, H., pp. (1 - 15). Selbstverlag Abteilung für Physische Geographie und Landschaftsökologie der Technischen Universität Braunschweig, ISSN 0170-7299, Braunschweig
- Bishop, T. F. A., Minasny, B., 2006. Digital Soil-Terrain Modeling: The Predictive Potential and Uncertainty. In: Grunwald, S. (Ed.) *Environmental Soil-Landscape Modeling*. ISBN 0-8247-2389-9, CRC Press, Boca Raton.

- Böhner, J., McCloy, K.R., & Strobl, J. (2006). *Göttinger Geographische Abhandlungen 115: SAGA - Analysis and Modelling Application*, Geographisches Institut der Universität Göttingen, Verlag Erich Goltz GmbH & Co. KG, Göttingen
- Bourennane, H., King, D., Couturier, A., 2000. Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. *Geoderma* 97, pp.(255– 271).
- Breimann, L., Friedmann, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees*, CRC press, Wadsworth.
- Burrough, P. A., Bouma, J., Yates, S. R., 1994. The state of the art in pedometrics. *Geoderma*, Vol 62, pp.(311 – 326), ISSN 0016-7061
- Cimmery, V. (2007). *User guide for SAGA, version 2.0*, Available from: <http://sourceforge.net/projects/saga-gis/files/> (access: 25.11.2009)
- Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., & Helt, T. (2000). Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma*, Vol. 97, pp. (367–391), ISSN 0016-7061
- Dokutschajew, W. W., 1883. *Russkij Chernozem*, St. Petersburg.
- FAO, IUSS Working Group WRB (2007). *World Reference Base for Soil Resources*, ISRIC, Rome
- Fries, A., Rollenbeck, R., Göttlicher, D., Nauss, T., Homeier, J., Peters, T., & Bendix, J. (2009). Thermal structure of a megadiverse Andean mountain ecosystem in southern Ecuador, and its regionalization. *Erdkunde*, Vol. 63, pp. (321–335), ISSN 0014-0015
- Gessler, P., Moore, I., McKenzie, N., & Ryan, P. (1995). Soil-landscape modelling and spatial prediction of soil attributes. *International Journal of Geographical Information Systems*, Vol. 9, 4, pp. (421– 432), ISSN 1365-8816
- Grunwald, S., 2006. What Do We Really Know about the Space-Time Continuum of Soil Landscapes? In: Grunwald, S. (Ed.), *Environmental Soil-Landscape Modeling*, pp. (3 – 36), ISBN 0-8247-2389-9, CRC Press, Boca Raton.
- Hannemann, J. (2010). *Die Berücksichtigung inhaltlicher und räumlicher Unschärfe bei der GIS-gestützten Erstellung von Bodenkarten*. Dissertation Universität Bayreuth, Geowissenschaft. Shaker Verlag, Aachen
- Hengl, T., Heuvelink, G.B.M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, Vol. 120, No. 1-2, pp. (75– 93), ISSN 0016-7061
- Homeier, J., Dalitz, H., & Breckle, S.-W. (2002). Waldstruktur und Baumartendiversität im montanen Regenwald der Estación Científica San Francisco. Südecuador. *Ber. d. Reinh. Tüxen-Ges.*, Vol. 14, pp. (109–118), ISSN 0940-418X
- Jenny, H. (1941). *Factors of soil formation. A system of quantitative pedology*, McGraw-Hill, New York.
- Lagacherie, P. & Holmes, S. (1997). Addressing geographical data errors in a classification tree soil unit prediction. *International Journal of Geographic Information Science*, Vol. 11, pp. (183–198), ISSN 1365-8816
- Lea, N. L. (1992). An aspect driven kinematic routing algorithm. In: *Overland Flow Hydraulics and Erosion Mechanics*. Parsons, A. J. and Abrahams, A. D., pp. (393 – 407), London
- Liess, M., Glaser, B., & Huwe, B. (2009). Digital Soil Mapping in Southern Ecuador. *Erdkunde*, Vol. 63, 4, pp. (309–319), ISSN 0014-0015
- Litherland, M., Aspen, J.A., & Jemielita, R.A., 1994. The metamorphic belts of Ecuador. *Overseas Memoirs - British Geological Survey*, Vol. 11, pp. (1–147), ISSN 0951-6646
- McBratney, A.B. & DeGruiter, J.J. (1992). A continuum approach to soil classification by modified fuzzy k-means with extragrades. *Journal of Soil Science*, Vol. 43, pp. (159– 175), ISSN 0038-075X

- Milne, G. (1935). Some Suggested Units of Classification and Mapping, Particularly for East African Soils. *Soil Research*, Vol. 4, pp. (183-198), ISSN 0309-133
- Moran, C.J. & Bui, E.N. (2002). Spatial data mining for enhanced soil map modelling. *International Journal of Geographic Information Science*, Vol. 16, pp. (533-549), ISSN 1365-8816
- Moore, I., Gessler, P., Nielsen, G., Peterson, G., 1993. Soil attribute prediction using terrain analysis. In: *Soil Science Society of America Journal*, 57/ 2, pp.(443- 452).
- Odeh, I., McBratney, A., Chittleborough, D., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, 63/3-4, pp.(197- 214).
- Rollenbeck, R. (2006). Variability of precipitation in the Reserva Biológica San Francisco / Southern Ecuador. *Lyonia, A Journal of Ecology and Application*, Vol. 9, No. 1, pp. (43-51), ISSN 0888-9619
- Schelling, J. (1970). Soil genesis, soil classification and soil survey. *Geoderma*, 4, 3, pp. (165-193), ISSN 0016-7061
- Schrumpf, M., Guggenberger, G., Valarezo, C., & Zech, W., 2001. Tropical montane rainforest soils. Development and nutrient status along an altitudinal gradient in the South Ecuadorian Andes. *Die Erde*, Vol. 132, pp. (43-59), ISSN 0013-9998
- Simonson, W.R. (1959). Outline of a Generalized Theory of Soil Genesis. *Soil Science Society of America Journal*, Vol. 23, pp. (152-156), ISSN 0361-5995
- Skidmore, A.K., Watford, F., Luckananurug, P., & Ryan, P.J. (1996). An operational GIS expert system for mapping forest soils from a geographical information system. *International Journal of Geographic Information Science*, Vol. 5, pp. (431-445), ISSN 1365-8816
- Soil Survey Staff (2006). Keys to Soil Taxonomy. 10th ed. United States Department of Agriculture, Natural Resources Conservation Service, Available from: <http://soils.usda.gov/technical/classification/taxonomy/> (access: 22/10/2007)
- Stoyan, R., 2000. *Aktivität, Ursachen und Klassifikation der Rutschungen in San Francisco/Südecuador*. Diplomarbeit Universität Erlangen.
- Strahler, A.N. (1957). Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, Vol. 38, No. 6, pp. (913-920), ISSN 0002-8606
- Therneau, T.M. & Atkinson, B. (2003). The rpart Package, Available from: <http://cran.r-project.org/web/packages/rpart/rpart.pdf> (access: 28/02/2008)
- Thomas, A.L., King, D., Dambrine, E., Couturies, A., & Roque, A. (1999): Predicting soil classes with parameters derived from relief geologic materials in a sandstone region of the Vosges mountains (northeastern France). *Geoderma*, Vol. 90, pp. (291-205), ISSN 0016-7061
- Wilcke, W., Yasin, S., Abramowski, U., Valarezo, C., & Zech, W. (2002). Nutrient storage and turnover in organic layers under tropical montane rain forest in Ecuador. *European Journal of Soil Science*, Vol. 53, pp. (15-27), ISSN 1351-0754
- Wilcke, W., Valladarez, H., Stoyan, R., Yasin, S., Valarez, C., & Zech, W. (2003). Soil properties on a chronosequence of landslides in montane rain forest, Ecuador. *Catena*, Vol. 53, pp. (79-95), ISSN 0341-8162
- Yasin, S. (2001). *Water and Nutrient Dynamics in Microcatchments under Montane Forest in the South Ecuadorian Andes*. Bayreuther Bodenkundliche Berichte, Band 73.
- Zevenbergen L.W. & Thorne C.R. (1987). Quantitative Analysis of Land Surface Topography. *Earth Surface Processes Landforms*, Vol. 12, pp. (47 - 56), ISSN 0197-9337



## **Principles, Application and Assessment in Soil Science**

Edited by Dr. Burcu E. Ozkaraova Gungor

ISBN 978-953-307-740-6

Hard cover, 394 pages

**Publisher** InTech

**Published online** 22, December, 2011

**Published in print edition** December, 2011

Our dependence on soil, and our curiosity about it, is leading to the investigation of changes within soil processes. Furthermore, the diversity and dynamics of soil are enabling new discoveries and insights, which help us to understand the variations in soil processes. Consequently, this permits us to take the necessary measures for soil protection, thus promoting soil health. This book aims to provide an up-to-date account of the current state of knowledge in recent practices and assessments in soil science. Moreover, it presents a comprehensive evaluation of the effect of residue/waste application on soil properties and, further, on the mechanism of plant adaptation and plant growth. Interesting examples of simulation using various models dealing with carbon sequestration, ecosystem respiration, and soil landscape, etc. are demonstrated. The book also includes chapters on the analysis of areal data and geostatistics using different assessment methods. More recent developments in analytical techniques used to obtain answers to the various physical mechanisms, chemical, and biological processes in soil are also present.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mareike Ließ, Bruno Glaser and Bernd Huwe (2011). Soil-Landscape Modelling – Reference Soil Group Probability Prediction in Southern Ecuador, Principles, Application and Assessment in Soil Science, Dr. Burcu E. Ozkaraova Gungor (Ed.), ISBN: 978-953-307-740-6, InTech, Available from:  
<http://www.intechopen.com/books/principles-application-and-assessment-in-soil-science/soil-landscape-modelling-reference-soil-group-probability-prediction-in-southern-ecuador>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen