https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

SOIL Discussions Open Access EGU

Development of pedotransfer functions for tropical mountain soilscapes: Spotlight on parameter tuning in machine learning

Anika Gebauer[1], Monja Ellinger[1], Victor M. Brito Gomez[2], Mareike Ließ[1]

[1]Department Soil System Science, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), Germany
[2]Departamento de Recursos Hídricos y Ciencias Ambientales, Facultad de Ciencias Agropecuarias, Universidad de Cuenca, Cuenca, Ecuador

Correspondence to: Anika Gebauer (anika.gebauer@ufz.de)

Abstract. Machine learning algorithms are good in computing non-linear problems and fitting complex composite functions, which makes them an adequate tool to address multiple environmental research questions. One important application is the development of pedotransfer functions (PTF). This study aims to develop water retention PTFs for two remote tropical mountain regions of rather different soil-landscapes, dominated by (1) organic soils under volcanic influence, and (2) tropical mineral soils. Two tuning procedures were compared to fit boosted regression tree models: 1) tuning by grid search, which is the standard approach in pedometrics and 2) tuning by differential evolution optimization. A nested cross-validation approach was applied to generate robust models. The developed area-specific PTFs outrival other more general PTFs. Furthermore, the first PTF for typical soils of Páramo landscapes, i.e. organic soils under volcanic influence, is presented. Overall, results confirmed the differential evolution algorithm's high potential for tuning machine learning models. While models based on tuning by grid search roughly predicted the response variables' mean for both areas, models applying the differential evolution algorithm for parameter tuning explained up to 22 times more of the response variables' variance.

1 Introduction

Machine learning algorithms are good at fitting highly complex non-linear functions (Witten et al., 2011). Major application fields in soil science investigate the soils' spatial variability (Heung et al., 2016), relate data from soil sensing to soil properties (Viscarra Rossel et al., 2016), or develop pedotransfer functions (PTFs, Botula et al., 2014; Van Looy et al., 2017). McBratney et al. (2019) give a timeline on developments in pedometrics, that refer to machine learning in multiple applications.

Pedotransfer functions derive laborious and complex soil parameters from more readily available soil properties. Most PTFs predict soil hydraulic properties. A review on the involved methodology is provided by Pachepsky and Rawls, (2004), Shein and Arkhangel'skaya (2006), Vereecken et al. (2010). Machine learning algorithms applied for PTF development include e.g. support vector machines (Lamorski et al., 2008), artificial neural networks (Haghverdi et al., 2012) and regression trees (Tóth et al., 2015). According to the review by Van Looy et al. (2017), most PTFs are developed for mineral soils, while PTFs applicable to organic soils or to soils with specific properties like volcanic ash soils are highly underrepresented. Patil

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

and Singh (2016) and Botula et al. (2014) provide reviews of hydrological PTFs for mineral soils of certain tropical and temperate regions. With particle size distribution (PSD) being the basic input parameter to derive soil hydrologic properties, most PTFs also use bulk density (BD) and soil organic carbon content (SOC) as predictors. As summarized in the review of Patil and Singh (2016), the application of existing hydrological PTFs is often restricted by two reasons: 1) The majority of PTFs are developed on soils that developed under certain conditions. Often these PTFs cannot be applied sufficiently in other regions as the site-specific soil-forming conditions can cause considerable differences in physical and chemical soil properties. This is for example demonstrated by the studies of Botula et al. (2012) and Moreira et al. (2004) who were able to show that, when applied to independent tropical soil data, existing temperate PTFs perform worse than existing tropical PTFs. 2) On the other hand, the applicability of existing PTFs is further restricted by the required input data. As stated by Morris et al. (2019) hydraulic PTFs developed on mineral soils are often inapplicable to organic soils. The measurement of the predictor variable PSD may be hampered by high organic matter contents, other organic soils may not include sufficient mineral soil material to justify PSD analysis at all. In general, only very few PTFs were developed for organic soils and most of them are based on data from specific temperate regions and rely on very specific predictor variables. Korus et al. (2007) for example related the water retention of polish peat soils to the predictors ash content, specific surface area, BD, pH, and iron content. In Finnland semiempirical water retention PTFs were developed on different predictors including BD, sampling depth and botanical residues (Weiss et al., 1998). Even though it was never intended to be used for predictions, Rocha Campos et al. (2011) provide the only regression model known to us, which relates the soil hydrologic parameters of tropical organic soils to independent variables (fiber content, mineral material, BD, organic matter fractionated into humin as well as fulvic and humic acids).

Finally, the application of machine learning algorithms requires them to be adjusted to the specific modelling problem by parameter tuning. Tuning parameter values cannot be calculated analytically, so in soil science applications, grid search is often used as a standard technique (e.g. Babangida et al., 2016; Khlosi et al., 2016; Twarakavi et al., 2009). It works by testing a number of predefined parameter values or combinations of parameter values to finally choose the best. Accordingly, the predominant part of the multivariate parameter space cannot be searched in the case of real-valued parameters and the optimum might not be found. To overcome this limitation, mathematical optimisation is a promising alternative. Commonly applied optimisation algorithms include artificial bee colony, simulated annealing, particle swarm optimisation, the nelder-mead method or evolutionary and genetic strategies. Their applications range from pattern recognition (e.g. Jayanth et al., 2015; Liu and Huang, 1998), through solving combinatorial problems (e.g. Kang-Ping Wang et al., 2003; Reeves, 1993) to parameter tuning in machine learning (e.g. Imbault and Lebart, 2004; Ozaki et al., 2017). We would like to particularly emphasize the differential evolution algorithm. It has been applied to diverse optimisation problems including the prediction of stable metallic clusters (Yang et al., 2018), the navigation of robots (Martinez-Soltero and Hernandez-Barragan, 2018), the classification of microRNA targets (Bhadra et al., 2012), parity-P problems (Slowik and Bialko, 2008), the parameter tuning of machine learning models trained to e.g. predict landslides (Tien Bui et al., 2017), and last but not least the first successful application in pedometrics (Gebauer et al., 2019). Compared to the other algorithms

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

mentioned above, it usually leads to better results and a comparatively low computing time (Price et al., 2005).

This study first of all aims to develop water retention PTFs for two tropical soil-landscapes dominated by organic soils under volcanic influence, like they are commonly occurring in Páramo regions, and tropical soils of a dry climate. Currently, PTFs suitable for the soils of these regions are at best very few, if any, at all. As we assume the parameter tuning technique to affect the performance of the machine learning based PTFs, our second and equally important aim is to test the power of differential evolution optimization for tuning by comparing it to the grid search approach usually applied in pedometrics. We assume that the pre-eminence of optimization for parameter tuning in machine learning will particularly show when applying it to a machine learning algorithm that requires the fitting of numerous real-valued parameters. This is why we have chosen to fit boosted regression tree models. According to our knowledge, this is the first time both parameter tuning techniques are directly compared for a machine learning application in soil science.

2 Material and methods

2.1 Research areas

The two investigated soil-landscapes are situated in southern Ecuador (Fig. 1). The Quinuas Catchment encompasses an area of about 93 km^2, including parts of the Cajas National Park (Fig. 1c). Being located between 3000 and 4400 m above sea level (a.s.l.), the mean annual temperature ranges between 5.3 and 8.7 °C with no seasonality (Carrillo-Rojas et al., 2016). With peaks in March/May and October (Celleri et al., 2007) mean annual precipitation varies between 900 and 1600 mm (Crespo et al., 2011). Due to volcanic ash deposits and the cold and wet climate, soils with a low bulk density and high SOC contents are typical (Buytaert et al., 2007). The Quinuas Catchment can be allocated to the Páramo ecosystem (Guio Blanco et al., 2018), which plays a major role in the water supply of the inter-Andean region (Buytaert et al., 2006a, 2006b, 2007). The Laipuna dry forest region is part of the “Laipuna Conservation and Development Area” and covers approximately 16 km^2 (Fig. 1d). Its temperature profile shows little seasonal variability, while there is a rain period from January to May. Depending on the altitude ranging between 400 and 1500 m a.s.l., the mean annual temperature varies between 16 and 23 °C and mean annual precipitation between 540 mm and 630 mm (Peters and Richter, 2011b, 2011a). Additionally, El Niño-Southern Oscillation influences the area (Bendix et al., 2003, 2011). Laipuna is part of an ecosystem with high biodiversity and many endemic species (Best and Kessler, 1995; Linares-Palomino et al., 2009), which are strongly adapted to the ecosystem and may be threatened by possible climate-induced changes of the water supply.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

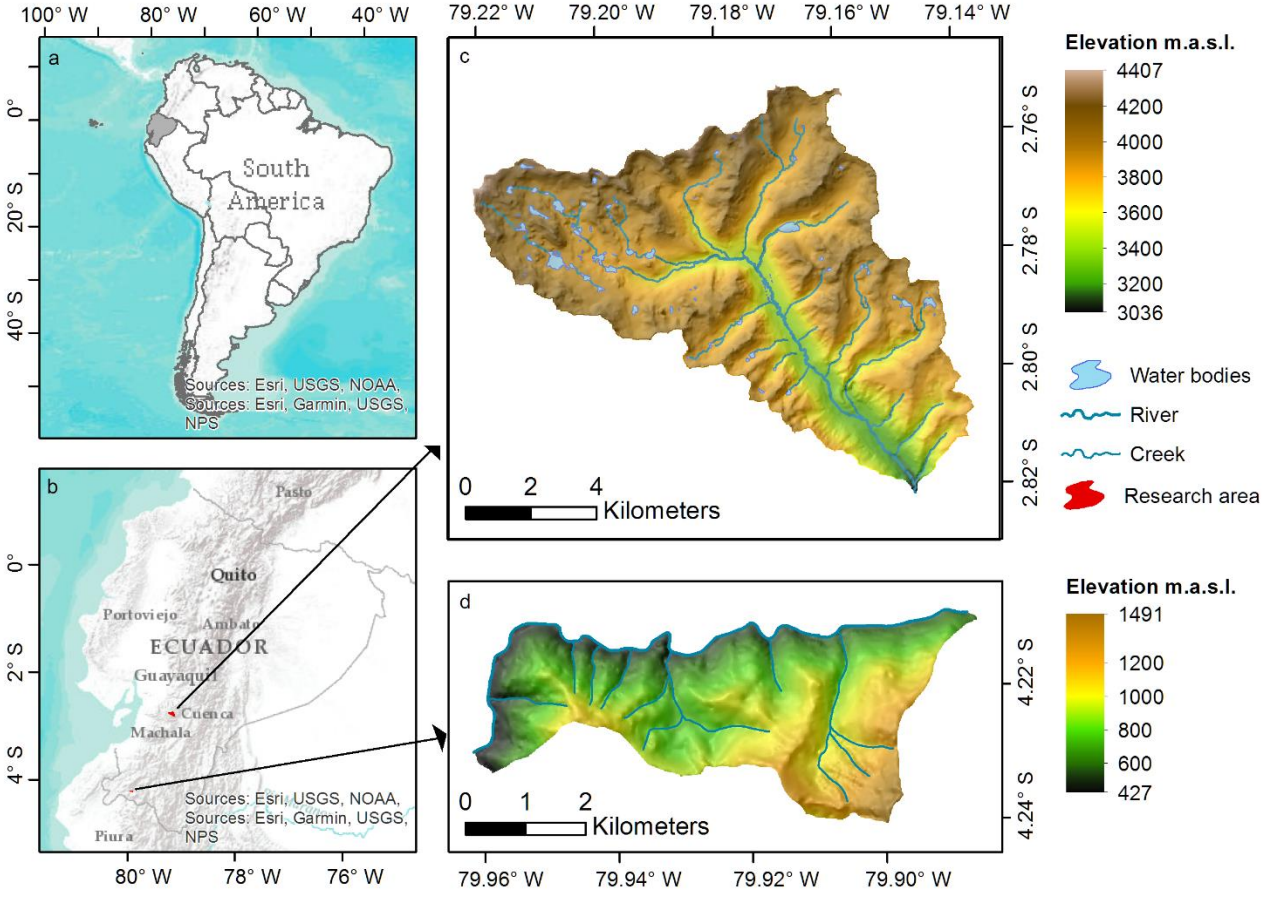


Fig. 1. Research areas. a) Ecuador within South America; b) Research areas within Ecuador; c) Quinuas. d) Laipuna (overlaid hillshading with a light source from north). Adapted from Ließ (2015). Topographical data use with permission from the Ecuadorian Geographical Institute (2013, national base, scale 1:50.000), further GIS data provided by NCI and ETAPA.

2.2 Soil data

To select representative datasets for the two areas, random stratified sampling was applied using the algorithm "QC-arLUS" (Ließ, 2015). It allows representing a research area`s complete landscape structure in sampling site selection while actual site selection is limited to the accessible area. Two sites were sampled per landscape stratum, resulting in 46 sites for Quinuas and 55 for Laipuna. Water retention at pF 0, 0.5, 1.5 and 2.5 was measured applying hanging water columns of increasing length to undisturbed samples of 100 cm^3. For BD measurements, the samples were oven-dried at 105 °C for three days. For Laipuna, disturbed samples were oven-dried at 40 °C, sieved to 2 mm and PSD was determined according to DIN ISO 11277:2002-08 in two (sand fractions) and three (clay and silt fractions) replicate samples. Measurements distinguish the following particle size classes: clay (< 2 µm), fine silt (2 – 6.3 µm), medium silt (6.3 – 20 µm), coarse silt (20 – 62 µm), fine sand (62 – 200 µm), medium sand (200 – 630 µm) and coarse sand (630 – 2000 µm). Sieved samples were ground and tested

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

105 for carbonates prior to the SOC determination using dry combustion. Multivariate outlier removal was carried out using the "hclust" function from R-package "fastcluster" (Müller, 2018), version 3.4.4.

2.3 Boosted regression trees

The BRT algorithm combines the machine learning techniques regression trees and boosting. Tree models use decision rules, which involve the predictor variables to recursively partition the response variable data into increasingly similar subgroups
110 until terminal nodes are reached (Kuhn and Johnson, 2013). The values of the terminal nodes are averaged to be used for prediction (James et al., 2017). The boosting machine learning technique improves the overall model accuracy by combining a number of simple models (Witten and Frank, 2005). To develop the PTFs, BRT models were trained using the "gbm" R-package, version 2.1.3 (Ridgeway, 2017), which is based on Friedman's (2002) stochastic gradient boosting. This boosting technique iteratively fits a number of simple regression tree models to random data subsets to finally make them form a
115 linear combination. Thereby, each new regression tree is fitted to explain most of the previous model's residuals. To apply a BRT model, usually up to four parameters are tuned: number of trees (n.trees), shrinkage, interaction depth, and bag fraction (e.g. Ottoy et al., 2017; Wang et al., 2017; Yang et al., 2016). Elith et al. (2008) provide a detailed analysis of their function: The n.trees parameter describes the number of regression tree models to be iteratively fitted. Shrinkage defines the model's learning rate by scaling the outcome of each simple regression tree, thereby controlling their contribution to the final
120 prediction. The interaction depth parameter controls the number of splits in each tree to divide the response variable data into subgroups. The bag fraction parameter determines the size of the randomly selected data subsets. This is able to reduce the risk of overfitting (Friedman, 2002), but may lower the model robustness (Elith et al., 2008). To develop PTFs for Quinuas and Laipuna, these four parameters were tuned following the steps described in Section 2.4 and 2.5.

2.4 Parameter tuning

125 Parameter tuning was done in two different ways: (1) by grid search and (2) by optimization applying the differential evolution algorithm. In order to reduce computing time, the number of values to be compared by grid search was limited to five for each of the k parameters. The selected values were based on the recommendations of Elith et al. (2008) and Ridgeway (2012). They are summarised in Table 1. Finally, k^5 different combinations of tuning parameter values, i.e. 625 four-dimensional vectors were compared. In contrast to this, the mathematical optimization algorithm "differential
130 evolution" by Storn and Price (1995) is able to search the multivariate space between defined upper and lower parameter limits. The parameter values are optimized by minimizing an objective function, which defines their suitability. The objective function is allowed to be stochastic and noisy and does not need to be differentiable or continuous (Mullen et al., 2011). Following evolutionary theory, this is done by repeating three steps for i iterations: mutation, crossover, and selection (Fig. 2). At first, an initial parent population of v k-dimensional parameter vectors is generated randomly. With each
135 iteration i, these vectors are changed by mutation and randomly mixed by crossover to generate a new population. Selection compares the objective function values belonging to the parent and the new vector to decide whether a new vector replaces

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

its parent vector. Differential evolution was applied using the R-package "DEoptim", version 2.2.4 (Ardia et al., 2016). For each tuning parameter, optimization limits correspond to the smallest and largest grid search values (Table 1). The number of vectors of size $k = 4$ tuning parameters was set to $v = 100$. The R-package's default mutation strategy was used, which changes each parent vector by adding two summands: (1) the difference between two random parent vectors and (2) the difference between the vector to be perturbed and the best vector found in the parent population. Summands were scaled by the factor 0.8. During crossover the probability of randomly mixing the parent and the mutated vectors' elements was set to 50%. To reduce computing time, the optimisation process was stopped after i_{max} = 10 iterations without improving the objective function or a maximum number of 200 iterations, respectively. Prior to the selection step, the discrete tuning parameter values (n.trees, interaction depth) were rounded, as the differential evolution algorithm treats all values as real numbers during mutation and crossover. To select the final tuning parameter values, grid search and differential evolution both minimised the cross validated $RMSE_T$ as objective function. $RMSE_T$ calculation is explained in Section 2.5.

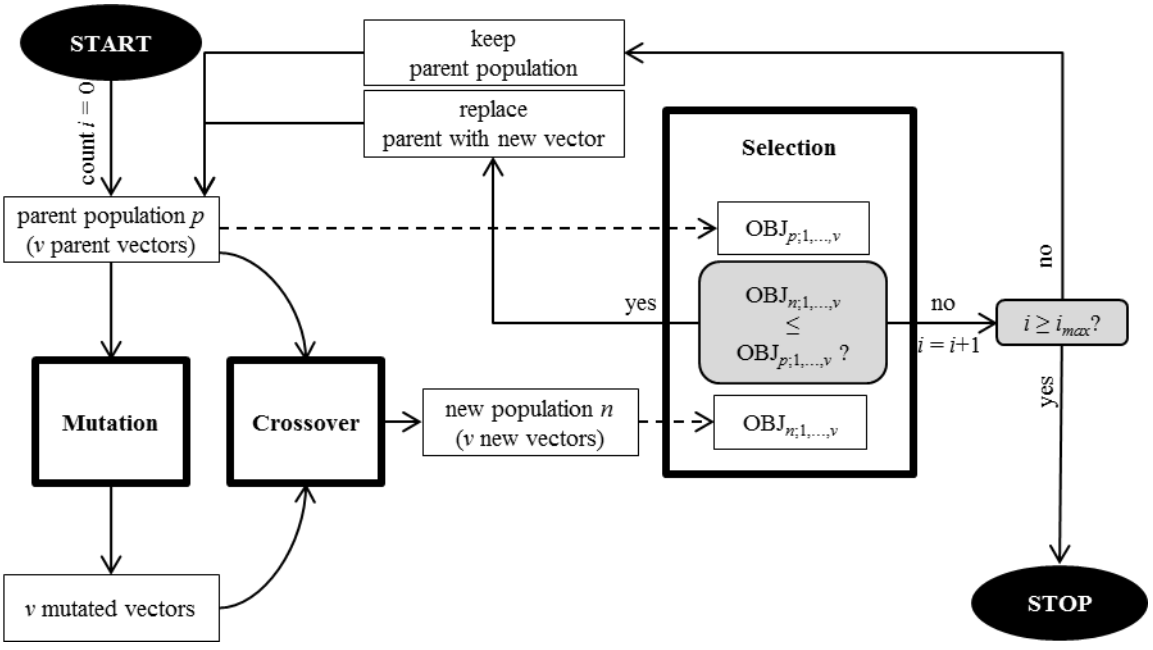


Fig. 2. Flowchart of the differential evolution algorithm. OBJ = objective function, p = parent population, n = new population, i = iteration, i_{max} = maximum number of iterations, v = number of vectors, BRT = boosted regression trees. Adapted from Gebauer et al. (2019).

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

Table 1. Tuning parameter values to be tested by grid search and optimization limits required by the differential evolution algorithm.

Tuning parameter	Grid search values	Lower and upper differential evolution limits	
n.trees	100; 1000; 2000; 3000; 4000	100	4000
shrinkage	0.001; 0.005; 0.01; 0.05; 0.1	0.001	0.1
interaction depth	1; 2; 3; 4; 5	1	5
bag fraction	0.5; 0.6; 0.7; 0.8; 0.9	0.5	0.9

2.5 Performance evaluation

In order to build robust models, we followed a nested cross-validation approach. Stratified five-fold cross-validation (CV) was applied for two purposes: (1) to conduct robust parameter tuning on resampled data subsets by either grid search or the differential evolution algorithm and (2) to evaluate the final performance of models built on tuned parameter values. CV provides error metrics with good bias and variance properties, is beneficial for small datasets and avoids overfitting (Arlot and Celisse, 2009; James et al., 2017). Following the steps shown in Fig. 3, the stratified five-fold CV was implemented with five repetitions for model evaluation and one repetition for parameter tuning. In Step 1, the complete dataset (n = 100%) was split into five folds with each of them (n = 20%) once used as the test set, leaving the remaining folds as the model training set. For resampling in parameter tuning (Step 2), each model training set was again subdivided, similar to Step 1. Each tuning parameter vector in grid search and the differential evolution algorithm was evaluated by the cross-validated $RMSE_T$ (Step 3, Step 4). By comparing the $RMSE_T$, the best vector of tuning parameter values for each model evaluation training set was selected and applied (Step 5, Step 6). To assess model performance, the coefficient of determination (R^2_E) and the root mean squared error ($RMSE_E$) of model evaluation were calculated by predicting the associated test set data (Step 7). To divide the datasets into folds, the function "partition_cv_strat" from R-package "sperrorest", version 2.0.0 (Brenning et al., 2017) was applied, with three equal probability strata of the response variable's density function.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

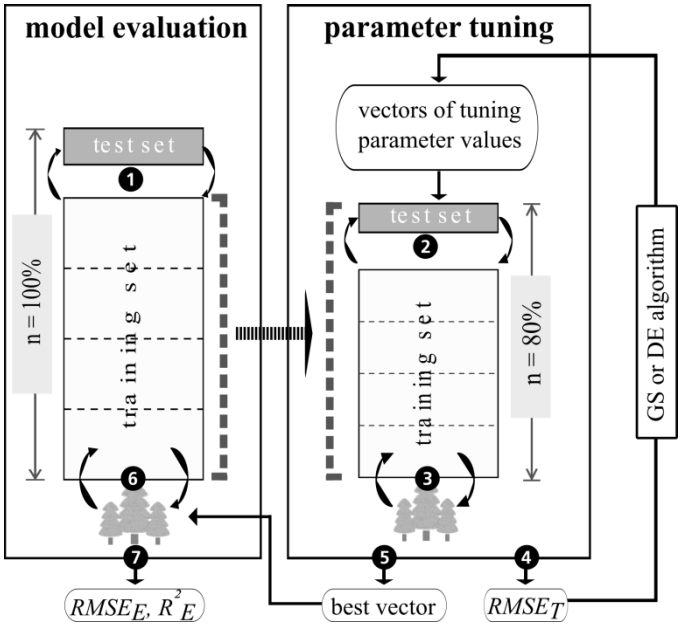


Fig. 3. Nested cross-validation approach comprising model evaluation and parameter tuning. Adapted from Guio Blanco et al. (2018). The tree icons symbolize BRT models, which are repeatedly (circular arrows) trained and tested on different data sets. The numbers within black circles belong to the modeling steps' description. $RMSE_T$ = root mean squared error of parameter tuning, $RMSE_E$ = root mean squared error of model evaluation, R^2_E = coefficient of determination of model evaluation, GS = grid search, DE = differential evolution.

3 Results and discussion

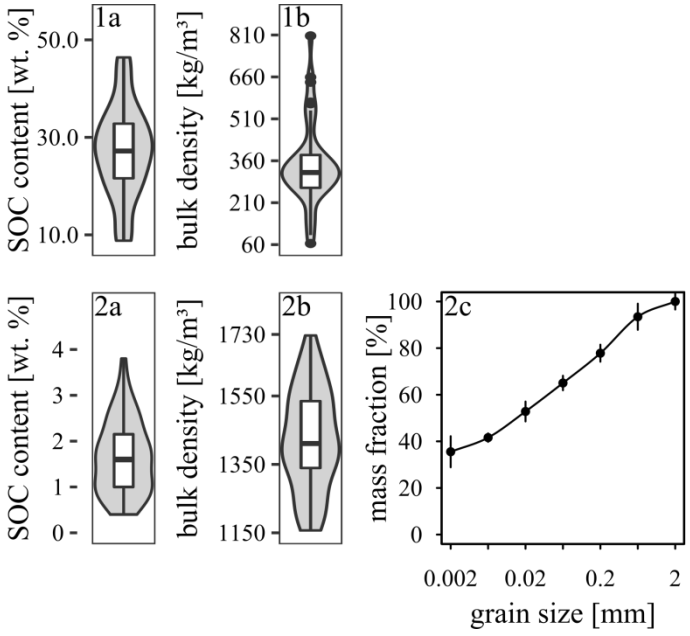
3.1 Model input

After removing multivariate outliers, the dataset contained 51 and 46 values for Laipuna and Quinuas, respectively. A summary of the data is shown in Fig. 4 and 5. As expected both areas show huge differences regarding the values of response and predictor variables. BD values in Quinuas range from 64 to 807 kg m^{-3}, while SOC values vary between 8.8 and 46.4 wt. %. SOC values are normally distributed, while BD data is skewed right. Averagely decreasing by 22%, water retention ranges from 0.25 cm^3 cm^{-3} (pF 2.5) to 0.94 cm^3 cm^{-3} (pF 0). While the data is skewed to the right for pF 0, the data distribution for the other pF values is left-skewed. For Laipuna, BD ranges between 1157 and 1727 kg m^{-3}, displaying a right-skewed distribution. The SOC content is normally distributed and varies between 0.4 and 3.8 wt. %. Clay content ranges between 17 and 48 %, silt between 24 and 45 % and sand between 14 and 50 %. Especially fine and medium silt show a skewed distribution. Water retention values are ranging between 0.25 cm^3 cm^{-3} (pF 2.5) and 0.61 cm^3 cm^{-3} (pF 0). On average, they decrease by 37 ± 0.09 % with increasing water tension. Data is skewed right for pF 0 and skewed left for pF 0.5.

Quinuas soils go along with the low density, porous soils, rich in organic material, that are found throughout the Paute river basin (Buytaert et al., 2007; Poulenard et al., 2003). Loosely bedded volcanic ash deposits further explain the low bulk

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

195 density values (Buytaert et al., 2007). High SOC contents are caused by low redox potential and the presence of organometallic complexes inhibiting degradation processes (Buytaert et al., 2006a). Comparatively high water retention values can be attributed to the porous structure of Páramo soils being able to retain a lot of water (Buytaert et al., 2007). The relatively small decrease in water retention with increasing water tension is assumed to be caused by the high amounts of organic matter being characterized by a high water holding capacity (Buytaert et al., 2007). Measured BD and SOC content
200 are in accordance with data observed for other Páramo regions (e.g. Buytaert et al., 2007, 2006b). The water retention values are also comparable to data obtained in other Páramo areas (Buytaert et al., 2005) and soils with high organic matter content (Schwärzel et al., 2002, 2006). Extreme values in BD and water retention (Fig. 4 and Fig. 5) correspond to soil samples with a higher proportion of mineral components or andic properties (Guio Blanco et al., 2018). Expecting these values to be reliable, they were not removed from the model input. The BD and SOC data measured for the Laipuna soil samples
205 correspond to other dry forest ecosystems (e.g. Conti et al., 2014; de Araújo Filho et al., 2017; Singh et al., 2015), whereas the PSD shows higher clay content compared to the dry forest soils investigated by Cotler and Ortega-Larrocea (2006), Jha et al. (1996) and Sagar et al. (2003). Measured water retention values are higher than those obtained in a tropical dry forest in Brazil (Vasques et al., 2016), probably caused by the higher clay content enhancing the water holding capacity.



210 **Fig. 4. Predictor variables.** 4.1 Quinuas, 4.2 Laipuna, a) SOC content, b) bulk density, c) particle size distribution displayed as a cumulative distribution function.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

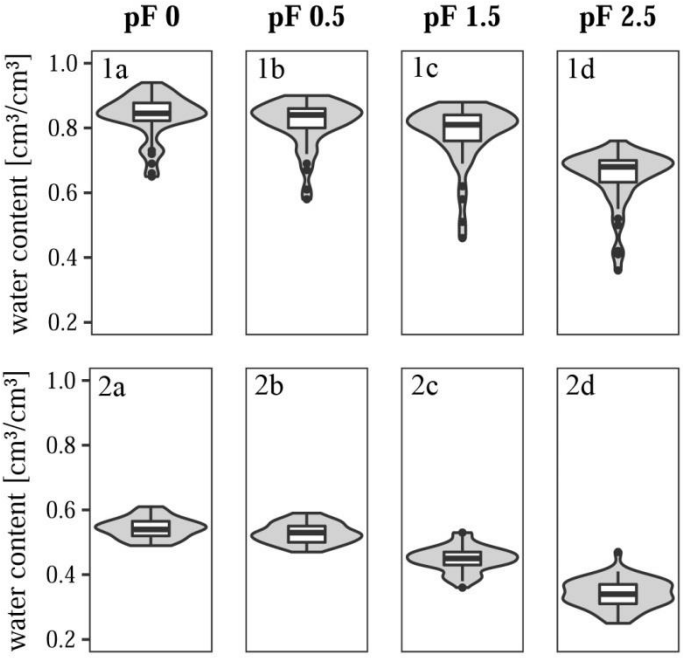


Fig. 5. Response variables. 5.1 Quinuas, 5.2 Laipuna, water retention at a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5.

3.2 Model performance

The performance of the final models built on parameters selected by grid search and the differential evolution algorithm is demonstrated by the error metrics R^2_E and $RMSE_E$ in Fig. 6 (Quinuas) and 7 (Laipuna) as well as by scatterplots comparing observed and predicted water retention values in Fig. 8 (Quinuas) and 9 (Laipuna). Models optimized by the differential evolution algorithm show a better predictive performance than models tuned by grid search. Based on scaled water retention values, all grid search models resulted in very similar $RMSE_E$ values (mean 0.20 ± 0.01) for Quinuas, while for Laipuna mean $RMSE_E$ range between 0.19 ± 0.00 (pF 2.5) and 0.25 ± 0.00 (pF 0 and pF 0.5). Looking at both error metrics, the performance of the differential evolution based Quinuas models worsens with increasing pF values, whereas for Laipuna model performance improves: final models with parameter tuning by differential evolution correspond to mean $RMSE_E$ values ranging from 0.11 ± 0.01 (pF 0) to 0.17 ± 0.01 (pF 2.5) for Quinuas and from 0.15 ± 0.01 (pF 2.5) to 0.26 ± 0.01 (pF 0) for Laipuna. Mean R^2_E values resulting from grid search, are varying between 0.04 ± 0.03 (pF 0) and 0.09 ± 0.07 (pF 1.5) for Quinuas and between 0.03 ± 0.04 (pF 0.5) and 0.05 ± 0.05 (pF 1.5) for Laipuna. The differential evolution algorithm resulted in mean R^2_E values ranging from 0.58 ± 0.09 (pF 2.5) to 0.79 ± 0.04 (pF 0) for Quinuas and from 0.27 ± 0.08 (pF 0) to 0.61 ± 0.04 (pF 2.5) for Laipuna. In comparison to grid search, models tuned by differential evolution show a higher predictive performance for all Quinuas models and the Laipuna models for pF 1.5 and pF 2.5: mean R^2_E values are up to 22 (Quinuas, pF 0) and 18 (Laipuna, pF 2.5) times higher, while $RMSE_E$ values are up to 1.9 (Quinuas, pF 0) and 1.3 (Laipuna, pF 2.5) times lower than those obtained by grid search. This corresponds to the scatterplots (Fig. 11 and 12): the largest

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

difference between grid search and differential evolution can be recognized for the pF 0 (Quinuas) and pF 2.5 (Laipuna) models. Additionally, as demonstrated by the scatterplots, the grid search models roughly predict the water retention values' mean, while the models with parameter tuning by differential evolution are able to explain more of the observations' variance. However, the five grid search predictions for each observation (Fig. 9 a-d and 10 a-d), cover a smaller range than the differential evolution predictions. Especially the differential evolution results of the Laipuna pF 0 and pF 0.5 models are characterized by comparatively high variance.

Probably caused by the better adjustment to the modeling problem, differential evolution models are able to explain more of the water retention's variance than the models tuned by grid search. The higher variability of the differential evolution predictions corresponds to the differential evolution tuning parameter values covering a wider range than those achieved by applying grid search (Section 3.2). For Laipuna, the improvement of the differential evolution models with increasing pF values is assumed to be caused by the used predictors. With increasing pF values the influence of PSD on water retention increases. At the same time the influence of BD decreases, which improves model performance as BD measurements were not corrected for stone content and are, therefore, not as good a predictor as they could be. For Quinuas the decreasing predictive performance with increasing pF values can probably be attributed to the lack of further predictors related to the soil matrix. PSD measurements were not conducted due to the prevalence of soils with organic properties in this area. Overall, the predictive power of all differential evolution based Quinuas models and the Laipuna pF 2.5 models are comparable to other studies. Botula et al. (2013) for example obtained R^2 ranging from 0.32 to 0.68 (pF 0) and from 0.60 to 0.68 (pF 1.5) by using k-nearest neighbor for soil data originating from the Lower Congo. Keshavarzi et al. (2010) used an artificial neural network to predict water retention at different pF values for soils from the Qazvin province in Iran. Haghverdi et al. (2012) used the same machine learning technique on soils from northeastern and northern Iran. While Keshavarzi et al. (2010) gained R^2 of 0.77 (pF 2.5) and 0.72 (pF 4.2), Haghverdi et al. (2012) reached R^2 ranging from 0.81 to 0.95.

Applying existing PTFs on the Laipuna dataset confirmed the good performance of the differential evolution BRT models. PTFs were selected based on four criteria: (1) developed for tropical soils, (2) similar predictor variables, (3) regression equation provided and (4) included in the peer-reviewed Clarivate Analytics' Web of Science database. Following the approach of Shang (2013), soil texture was previously converted to the USDA classification system by spline interpolation. Mean $RMSE_E$ values of the differential evolution tuned BRT models were between 1.5 times (pF 2.5, Minasny and Hartemink (2011) and Tomasella et al. (2000)) and 8.8 times (pF 1.5, Barros et al. (2013) better (Table 2). Because of differences in the predictors, it is difficult to find organic PTFs applicable to the Quinuas dataset. An exhaustive literature search revealed only the PTF of Boelter (1969), who related water retention at pF 0 to BD for temperate peat soils in northern Minnesota. Water retention values predicted by the existing PTF differed a lot from measured values. For BD higher than 370 kg m^{-3} predictions became even negative. While applying differential evolution BRT models resulted in a mean $RMSE_E$ of 0.032 (unscaled), applying the Minnesota PTF resulted only in an RMSE of 1.860. The high RMSE value is

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

assumed to be caused by large differences between the Minnesota and Quinuas soils and underlines the necessity of developing water retention PTFs specifically for tropical organic soils.

In general, we expect BRT model performance to improve further by removing extreme values in the model input or by using larger datasets. Especially for Quinuas, low water retention values are underrepresented in the dataset. According to Guio Blanco et al. (2018), these values are primarily observed in the lower part of the river valley and include measurements from mineral soils.

Table 2. Unscaled root mean squared errors for the tested PTFs. The best results for each matric potential are underlined. BRT PTF results are averaged.

PTF [data origin]	Theta 0	Theta 0.5	Theta 1.5	Theta 2.5
BRT PTF [Laipuna]	0.031	0.029	0.032	0.030
Barros et al. (2013) [Brazil] *	0.176	0.118	0.280	0.071
Gaiser et al. (2000) [Brazil and Niger]	-	-	-	0.062
Minasny and Hartemink (2011) [various tropical regions]	-	-	-	0.044
Nguyen et al. (2014) [Vietnam]	-	-	0.047	0.070
Obalum and Obi (2012) [Nigeria]	-	-	-	0.150
Pollacco (2008) [USA]	-	-	-	0.068
Tomasella et al. (2000) [Brazil] *	0.080	0.071	0.053	0.044

* applied to predict parameters of the Van Genuchten equation first.

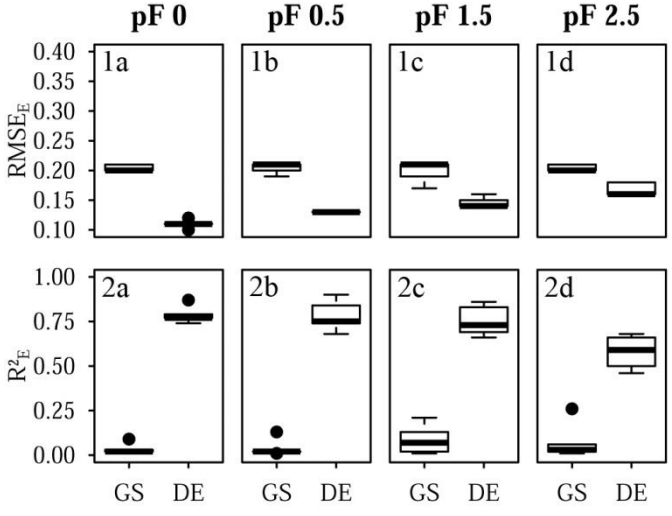


Fig. 6. Error metrics of the Quinuas BRT models. 6.1 $RMSE_E$, 6.2 R^2_E, a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5. Each boxplot is based on five values resulting from five CV repetitions. GS = grid search, DE = differential evolution algorithm.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

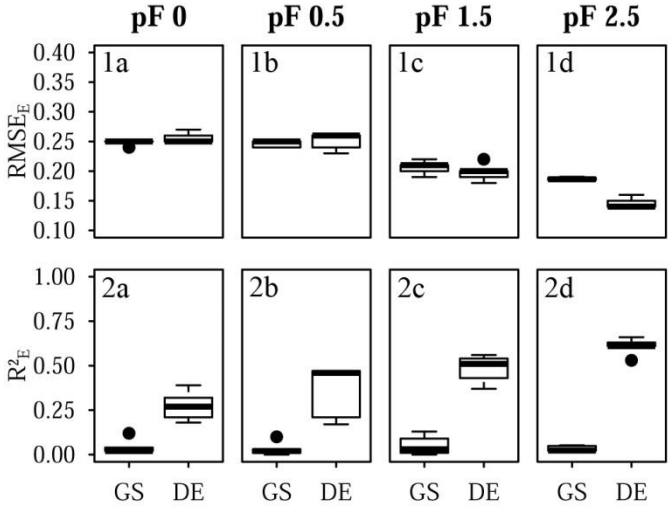


Fig. 7. Error metrics of the Laipuna BRT models. 7.1 $RMSE_E$, 7.2 R^2_E, a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5. Each boxplot is based on five values resulting from five CV repetitions. GS = grid search, DE = differential evolution algorithm.

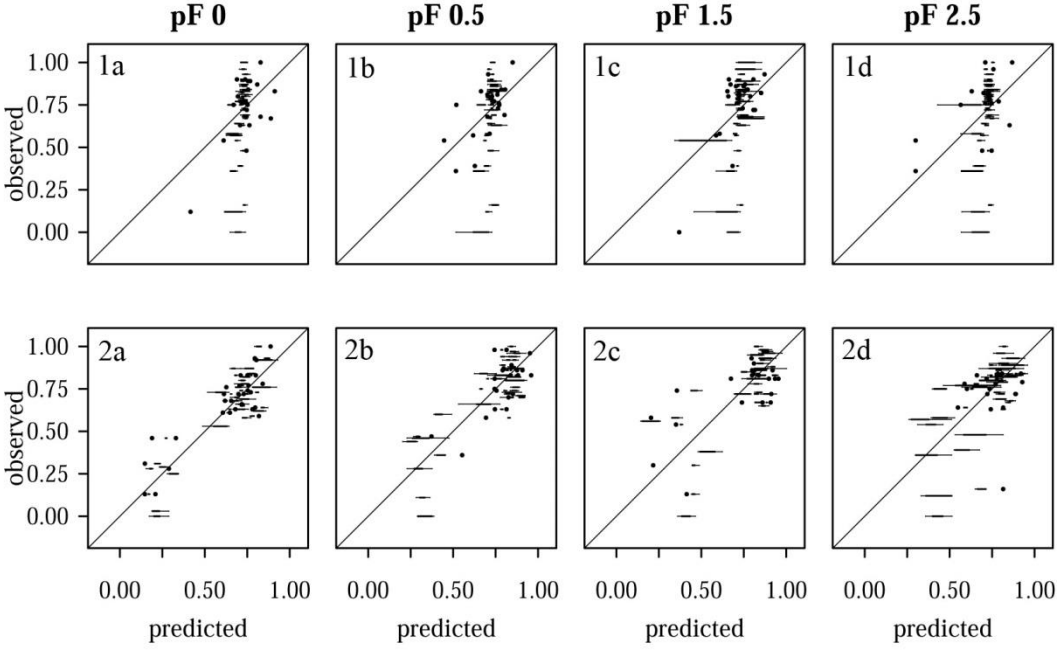


Fig. 8. Comparison of predicted and observed water retention values for Quinuas. 8.1 Models with tuning by grid search, 8.2 Models with parameter tuning by differential evolution, a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5. Each boxplot is based on five values resulting from five CV repetitions.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

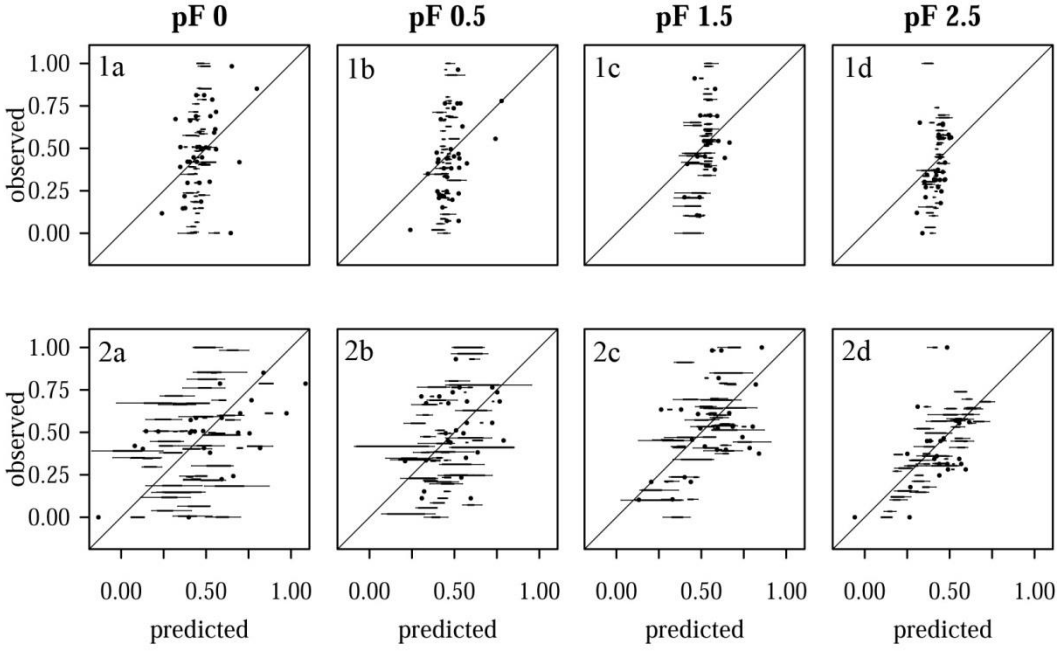


Fig. 9. Comparison of predicted and observed water retention values for Laipuna. 9.1 Models with tuning by grid search, 9.2 Models with parameter tuning by differential evolution, a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5. Each boxplot is based on five values resulting from five CV repetitions.

3.3 Model parameters

The final tuning parameter values obtained by grid search and the differential evolution algorithm are summarised in Fig. 10 (Quinuas) and 11 (Laipuna). The displayed boxplots show the selected 25 tuning parameter values corresponding to the models built by five-fold CV with five repetitions. Grid search compared 625 previously defined parameter vectors, while the differential evolution algorithm tested at least 1000 parameter vectors in the defined ranges of the four-dimensional parameter space. Differences between the two parameter tuning techniques are most distinct for n.trees and shrinkage for both research areas. Neglecting outliers, values obtained by the differential evolution algorithm cover a wider range than those resulting from grid search: while n.trees was in most cases set to the lowest tested value (100) by grid search, the differential evolution algorithm resulted in mean n.trees values ranging from 310 ± 321 (pF 0) to 810 ± 1132 (pF 1.5) for Quinuas and from 727 ± 851 (pF 0) to 1688 ± 1345 (pF 2.5) for Laipuna. Thereby, the mean n.trees values obtained by differential evolution parameter tuning are more than five (Quinuas) and more than ten times (Laipuna) higher than the mean grid search values. Neglecting extreme values, the shrinkage values resulting from the differential evolution algorithm also cover a wider range than the values obtained by the grid search tuning technique. For both areas, the shrinkage values were usually set to 0.001 or 0.01 by grid search, while applying the differential evolution algorithm resulted in mean shrinkage values ranging from 0.040 ± 0.028 (pF 0.5) to 0.047 ± 0.030 (pF 2.5) for Quinuas and from 0.034 ± 0.03 (pF 0) to 0.062 ± 0.027 (pF 2.5) for Laipuna. On average, the differential evolution shrinkage values are approximately 14 (Quinuas) and 17 (Laipuna) times higher than those obtained by grid search. The observed pattern is more complex for the other two tuning

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

parameters: interaction depth and bag fraction. Although the selected parameter ranges differ for most pF values, the median interaction depth values are the same for half of the cases for grid search and tuning by the differential evolution algorithm. The median of the selected bag fraction is at the upper limit for the Quinuas models that were tuned by the differential evolution algorithm, while grid search resulted in median bag fraction values at the lower limit in two cases. Laipuna bag fraction values do not show this pronounced difference between grid search and tuning by the differential evolution algorithm.

The selected tuning parameter values correspond to the differential evolution based models having more predictive power than those adapted by the common grid search approach. Usually higher n.trees values, as received from the differential evolution algorithm, are known to improve model performance (Elith et al., 2008). However, according to the results of Elith et al. (2008), by using more trees the shrinkage parameter gets smaller. The comparatively high differential evolution shrinkage values might be an indication of the n.trees values still being too small. For both areas, the differential evolution values for n.trees and shrinkage, covering a wider range than the grid search results, might be explained by the differential evolution optimization not being completed or being stuck in a local optimum. This corresponds to the high prediction variability of the final differential evolution models (Section 3.2). As n.trees and shrinkage control how precisely the model learns the input data's structure, these parameters are assumed to be more important than interaction depth and bag fraction. In this case, there might not even be an optimum for the latter two parameters. Especially for Laipuna, this might explain the interaction depth and bag fraction values of both tuning techniques covering the whole range of possible values. For Laipuna the bag fraction differences between differential evolution and grid search tuning remain unexplained. The results of both tuning techniques might be improved by testing more parameter values. For grid search, this can be realized by increasing the number of values to be compared for each tuning parameter. Increasing the number of iterations and starting with larger and thereby more heterogeneous initial populations is expected to do the same for differential evolution. This might also result in less variable differential evolution results. Besides increasing the number of iterations and the number of initial vectors, the risk of getting stuck in a local optimum can also be reduced by changing the parameters "crossover probability" and the "mutation scaling factor" as well as applying another mutation strategy (Das and Suganthan, 2011). To overcome the problem of choosing the right control parameters as well as the mutation strategy, self-adaptive differential evolution algorithms (e.g. Nahvi et al., 2016; Pierezan et al., 2017; Qin et al., 2009), which are able to automatically adjust their settings during the optimisation process, could be applied in future studies. In general, the superiority of differential evolution needs to be verified by applying it to further machine learning algorithms and applications.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

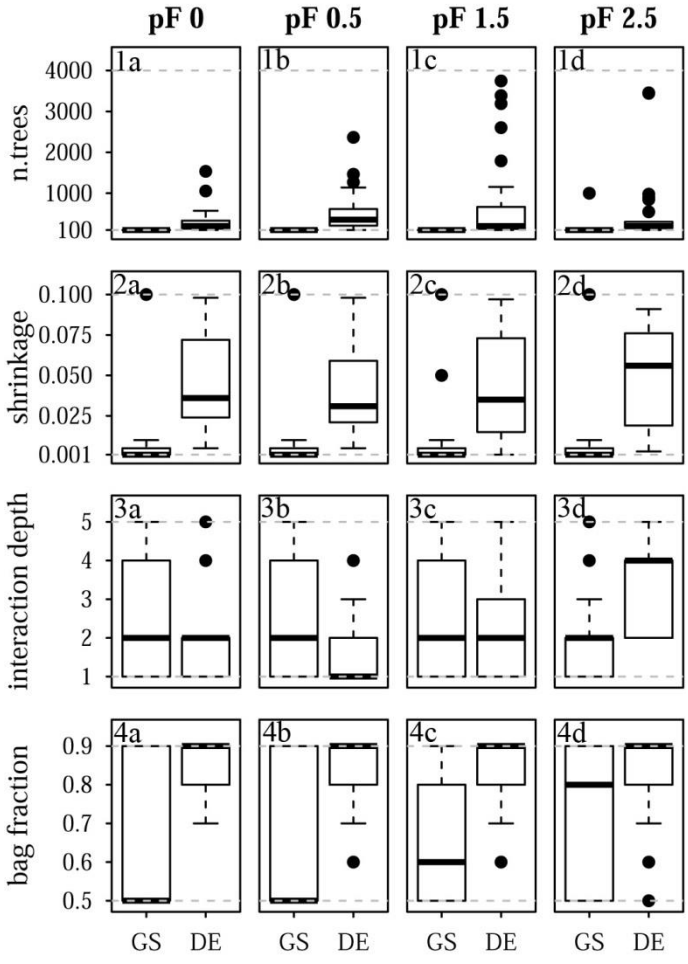


Fig. 10. Selected tuning parameter values for Quinuas. 10.1 n.trees, 10.2 shrinkage, 10.3 interaction depth, 10.4 bag fraction, a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5. Each boxplot is based on 25 values corresponding to the five-fold CV with five repetitions. Dashed grey lines are indicating the chosen optimization limits. GS = grid search, DE = differential evolution algorithm.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

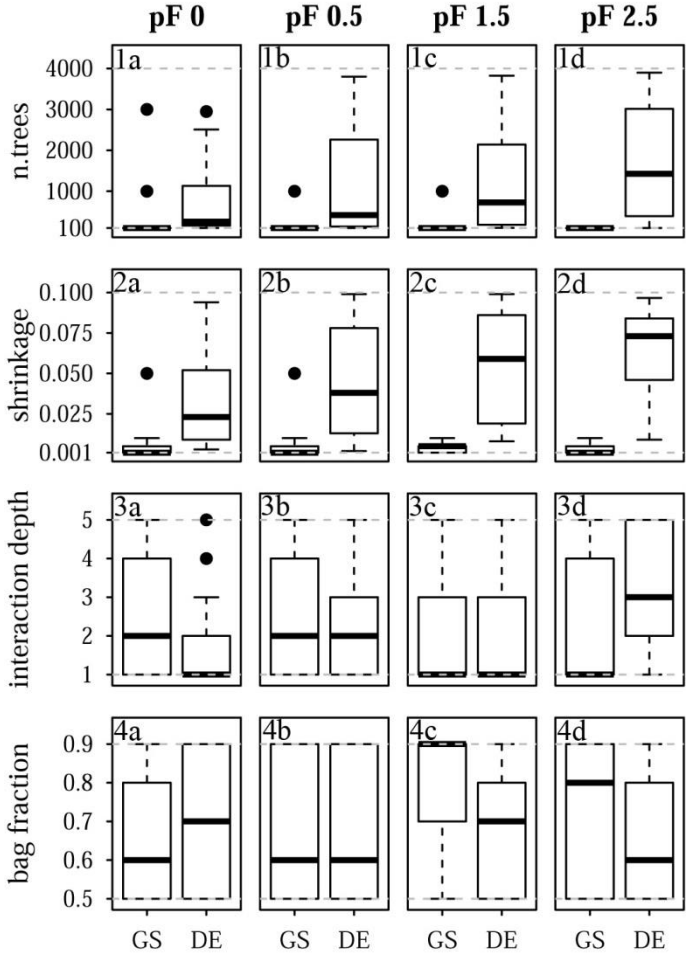


Fig. 11. Selected tuning parameter values for Laipuna. 11.1 n.trees, 11.2 shrinkage, 11.3 interaction depth, 11.4 bag fraction, a) pF 0, b) pF 0.5, c) pF 1.5, d) pF 2.5. Each boxplot is based on 25 values corresponding to the five-fold CV with five repetitions. Dashed grey lines are indicating the chosen optimization limits. GS = grid search, DE = differential evolution algorithm.

4 Conclusions

We successfully developed new PTFs for two tropical mountain regions. The comparison with ready available PTFs showed their high performance to predict soil water retention for the soils in these areas. This is of particular importance for soil process and hydrological modeling. Whether the two PTFs may also be applied to other areas of similar soils still has to be tested. The developed PTF for the Páramo area provides a novelty since PTFs for tropical organic soils under volcanic influence were unavailable until now. Furthermore, our study presents the first successful application of parameter tuning by differential evolution in PTF development. The comparison with the standard grid search technique revealed the superiority of the differential evolution algorithm and emphasizes the importance of parameter tuning for the successful application of machine learning models. Of course, this finding has to be confirmed by further applications in pedometrics including

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

different types of machine learning algorithms. We, therefore, hope to promote the implementation of optimization for parameter tuning within the pedometrics community by this study.

To sum up, especially for parameter tuning of real-valued machine learning algorithms future applications of the differential evolution algorithm are highly recommended.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This research was funded by the German Research Foundation (DFG) as part of the Platform for Biodiversity and Ecosystem Monitoring and Research in South Ecuador (PAK 825, LI 2360/1-1). Logistic support by the NGO Nature and Culture International (NCI) and the municipal public agency ETAPA is gratefully acknowledged.

References

de Araújo Filho, R. N., dos Santos Freire, M. B. G., Wilcox, B. P., West, J. B., Freire, F. J. and Marques, F. A.: Recovery of carbon stocks in deforested caatinga dry forest soils requires at least 60 years, For. Ecol. Manage., doi:10.1016/j.foreco.2017.10.002, 2017.

Ardia, D., Mullen, K., Peterson, B., Ulrich, J. and Boudt, K.: R package "DEoptim": Global Optimization by Differential Evolution., [online] Available from: https://cran.r-project.org/package=DEoptim, 2016.

Arlot, S. and Celisse, A.: A survey of cross-validation procedures for model selection, Stat. Surv., 4, 40–79, doi:10.1214/09-SS054, 2009.

Babangida, N. M., Ul Mustafa, M. R., Yusuf, K. W., Isa, M. H. and Baig, I.: Evaluation of low degree polynomial kernel support vector machines for modeling Pore-water pressure responses, MATEC Web Conf., 59, 0–5, doi:10.1051/matecconf/20165904003, 2016.

Barros, A. H. C., Lier, Q. de J. van, Maia, A. de H. N. and Scarpare, F. V.: Pedotransfer functions to estimate water retention parameters of soils in northeastern Brazil, Rev. Bras. Ciência do Solo, 37(2), 379–391, doi:10.1590/S0100-06832013000200009, 2013.

Bendix, J., Gämmerler, S., Reudenbach, C. and Bendix, A.: A case study on rainfall dynamics during El Niño/La Niña 1997/99 in Ecuador and surrounding areas as inferred from GOES-8 and TRMM-PR observations, Erdkunde, 57(2), 81–93, doi:10.3112/erdkunde.2003.02.01, 2003.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

Bendix, J., Trachte, K., Palacios, E., Rollenbeck, R., Göttlicher, D., Nauss, T. and Bendix, A.: El Niño meets La Niña-anomalous rainfall patterns in the "traditional" El Niño region of Southern Ecuador, Erdkunde, 65(2), 151–167, doi:10.3112/erdkunde.2011.02.04, 2011.

Best, B. J. and Kessler, M.: Biodiversity and Conservation in Tumbesian Ecuador and Peru, BirdLife International, Cambridge, U.K., 1995.

Bhadra, T., Bandyopadhyay, S. and Maulik, U.: Differential Evolution Based Optimization of SVM Parameters for Meta Classifier Design, Procedia Technol., 4, 50–57, doi:10.1016/j.protcy.2012.05.006, 2012.

Boelter, D. H.: Physical Properties of Peats as Related to Degree of Decomposition, Soil Sci. Soc. Amer. Proc., 33, 606–609, 1969.

Botula, Y.-D., Nemes, A., Mafuka, P., Van Ranst, E. and Cornelis, W. M.: Prediction of Water Retention of Soils from the Humid Tropics by the Nonparametric -Nearest Neighbor Approach, Vadose Zo. J., 12(2), 0, doi:10.2136/vzj2012.0123, 2013.

Botula, Y. D., Cornelis, W. M., Baert, G. and Van Ranst, E.: Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo), Agric. Water Manag., 111, 1–10, doi:10.1016/j.agwat.2012.04.006, 2012.

Botula, Y. D., Van Ranst, E. and Cornelis, W. M.: Pedotransfer Functions to Predict Water Retention for Soils of the Humid Tropics: A Review, Rev. Bras. Cienc. Do Solo, 38(3), 679–698, doi:10.1590/S0100-06832014000300001, 2014.

Brenning, A., Schratz, P. and Hermann, T.: R package "sperrorest": Perform Spatial Error Estimation and Variable Importance in Parallel., [online] Available from: https://cran.r-project.org/web/packages/sperrorest, 2017.

Buytaert, W., Wyseure, G., De Bièvre, B. and Deckers, J.: The effect of land-use changes on the hydrological behaviour of Histic Andosols in south Ecuador, Hydrol. Process., 19(20), 3985–3997, doi:10.1002/hyp.5867, 2005.

Buytaert, W., Deckers, J. and Wyseure, G.: Description and classification of nonallophanic Andosols in south Ecuadorian alpine grasslands (páramo), Geomorphology, 73(3–4), 207–221, doi:10.1016/j.geomorph.2005.06.012, 2006a.

Buytaert, W., Celleri, R., Willems, P., Bièvre, B. De and Wyseure, G.: Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes, J. Hydrol., 329(3–4), 413–421, doi:10.1016/j.jhydrol.2006.02.031, 2006b.

Buytaert, W., Deckers, J. and Wyseure, G.: Regional variability of volcanic ash soils in south Ecuador: The relation with parent material, climate and land use, Catena, 70(2), 143–154, doi:10.1016/j.catena.2006.08.003, 2007.

Carrillo-Rojas, G., Silva, B., Córdova, M., Célleri, R. and Bendix, J.: Dynamic mapping of evapotranspiration using an energy balance-based model over an andean páramo catchment of southern ecuador, Remote Sens., 8(160), doi:10.3390/rs8020160, 2016.

Celleri, R., Willems, P., Buytaert, W. and Feyen, J.: Space-time rainfall variability in the Paute Basin, Ecuadorian Andes, Hydrol. Process., 21(1), 3316–3327, doi:10.1002/hyp.6575, 2007.

Conti, G., Pérez-Harguindeguy, N., Quètier, F., Gorné, L. D., Jaureguiberry, P., Bertone, G. A., Enrico, L., Cuchietti, A. and Díaz, S.: Large changes in carbon storage under different land-use regimes in subtropical seasonally dry forests of southern South America, Agric. Ecosyst. Environ., 197, 68–76, doi:10.1016/j.agee.2014.07.025, 2014.

Cotler, H. and Ortega-Larrocea, M. P.: Effects of land use on soil erosion in a tropical dry forest ecosystem, Chamela watershed, Mexico, Catena, 65(2), 107–117, doi:10.1016/j.catena.2005.11.004, 2006.

Crespo, P. J., Feyen, J., Buytaert, W., Bücker, A., Breuer, L., Frede, H. G. and Ramírez, M.: Identifying controls of the rainfall-runoff response of small catchments in the tropical Andes (Ecuador), J. Hydrol., 407(1–4), 164–174, doi:10.1016/j.jhydrol.2011.07.021, 2011.

Das, S. and Suganthan, P. N.: Differential Evolution : A Survey of the State - of - the - Art, Ieee Trans. Evol. Comput., 15(1), 4–31, doi:10.1109/TEVC.2010.2059031, 2011.

Elith, J., Leathwick, J. R. and Hastie, T.: A working guide to boosted regression trees, J. Anim. Ecol., 77(4), 802–813, doi:10.1111/j.1365-2656.2008.01390.x, 2008.

Friedman, J. H.: Stochastic gradient boosting, Comput. Stat. Data Anal., 38, 367–378, doi:10.1016/S0167-9473(01)00065-2, 2002.

Gebauer, A., Brito Gómez, V. M. and Ließ, M.: Optimisation in machine learning: An application to topsoil organic stocks prediction in a dry forest ecosystem, Geoderma, 354(May), 113846, doi:10.1016/j.geoderma.2019.07.004, 2019.

Guio Blanco, C. M., Brito Gomez, V. M., Crespo, P. and Ließ, M.: Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest, Geoderma, 316(November 2017), 100–114, doi:10.1016/j.geoderma.2017.12.002, 2018.

Haghverdi, A., Cornelis, W. M. and Ghahraman, B.: A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data, J. Hydrol., 442–443, 46–54, doi:10.1016/j.jhydrol.2012.03.036, 2012.

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E. and Schmidt, M. G.: An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping, Geoderma, 265(March 2016), 62–77, doi:10.1016/j.geoderma.2015.11.014, 2016.

James, G., Witten, D., Hastie, T. and Tibshirani, R.: An Introduction to Statistical Learning, edited by G. Casella, S. Fienberg, and I. Olkin, Springer, New York, Heidelberg, Dordrecht, London., 2017.

Jayanth, J., Koliwad, S. and Ashok Kumar, T.: Classification of remote sensed data using Artificial Bee Colony algorithm, Egypt. J. Remote Sens. Sp. Sci., 18(1), 119–126, doi:10.1016/j.ejrs.2015.03.001, 2015.

Jha, P. B., Singh, J. S. and Kashyap, A. K.: Dynamics of viable nitrifier community and nutrient availability in dry tropical forest habitat as affected by cultivation and soil texture, Plant Soil, 180(2), 277–285, doi:10.1007/BF00015311, 1996.

Kang-Ping Wang, Lan Huang, Chun-Guang Zhou and Wei Pang: Particle swarm optimization for traveling salesman problem, in International Conference on Machine Learning and Cybernetics, pp. 1583–1585., 2003.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

Keshavarzi, A., Sarmadian, F., Sadeghnejad, M. and Pezeshki, P.: Developing Pedotransfer Functions for Estimating some Soil Properties using Artificial Neural Network and Multivariate Regression Approaches, Int. J. Environ. Earth Sci., 1(1), 31–37, 2010.

Khlosi, M., Alhamdoosh, M., Douaik, A., Gabriels, D. and Cornelis, W. M.: Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil, Eur. J. Soil Sci., 67(3), 276–284, doi:10.1111/ejss.12345, 2016.

Korus, M., Sławiński, C. and Witkowska-Walczak, B.: Attempt of water retention characteristics estimation as pedotransfer function for organic soils, Int. Agrophysics, 21(3), 249–254, 2007.

Kuhn, M. and Johnson, K.: Applied Predictive Modeling, Springer, New York, Heidelberg, Dordrecht, London., 2013.

Lamorski, K., Pachepsky, Y., Sławiński, C. and Walczak, R. T.: Using Support Vector Machines to Develop Pedotransfer Functions for Water Retention of Soils in Poland, Soil Sci. Soc. Am. J., 72(5), 1243, doi:10.2136/sssaj2007.0280N, 2008.

Ließ, M.: Sampling for regression-based digital soil mapping: Closing the gap between statistical desires and operational applicability, Spat. Stat., 13, 106–122, doi:10.1016/j.spasta.2015.06.002, 2015.

Linares-Palomino, R., Kvist, L. P., Aguirre-Mendoza, Z. and Gonzales-Inca, C.: Diversity and endemism of woody plant species in the Equatorial Pacific seasonally dry forests, Biodivers. Conserv., 19(1), 169–185, doi:10.1007/s10531-009-9713-4, 2009.

Liu, H. C. and Huang, J. S.: Pattern recognition using evolution algorithms with fast simulated annealing, Pattern Recognit. Lett., 19(5–6), 403–413, doi:10.1016/S0167-8655(98)00025-7, 1998.

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y. A., Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y. and Vereecken, H.: Pedotransfer Functions in Earth System Science: Challenges and Perspectives, Rev. Geophys., 55(4), 1199–1256, doi:10.1002/2017RG000581, 2017.

Martinez-Soltero, E. G. and Hernandez-Barragan, J.: Robot Navigation Based on Differential Evolution, IFAC-PapersOnLine, 51(13), 350–354, doi:10.1016/j.ifacol.2018.07.303, 2018.

McBratney, A., de Gruijter, J. and Bryce, A.: Pedometrics timeline, Geoderma, 338(December 2018), 568–575, doi:10.1016/j.geoderma.2018.11.048, 2019.

Minasny, B. and Hartemink, A. E.: Predicting soil properties in the tropics, Earth-Science Rev., 106(1–2), 52–62, doi:10.1016/j.earscirev.2011.01.005, 2011.

Moreira, L. F. F., Righetto, A. M. and Medeiros, V. M. D. A.: Soil Hydraulics Properties Estimation by Using Pedotransfer Functions in a Northeastern Semiarid Zone Catchment , Brazil, in International congress on environmental modelling and software, Osnabrück., 2004.

Morris, P. J., Baird, A. J., Eades, P. A. and Surridge, B. W. J.: Controls on Near-Surface Hydraulic Conductivity in a Raised Bog, Water Resour. Res., 1531–1543, doi:10.1029/2018WR024566, 2019.

Mullen, K., Ardia, D., Gil, D., Windover, D. and Cline, J.: DEoptim : An R Package for Global Optimization by Differential Evolution, J. Stat. Softw., 40(6), 1–26, doi:10.18637/jss.v040.i06, 2011.

Müller, D.: Fast Hierarchical Clustering Routines for R and "Python" Copyright, 2018.

Nahvi, B., Habibi, J., Mohammadi, K. and Shamshirband, S.: Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature, Comput. Electron. Agric., 124, 150–160, doi:10.1016/j.compag.2016.03.025, 2016.

Ottoy, S., Van Meerbeek, K., Sindayihebura, A., Hermy, M. and Van Orshoven, J.: Assessing top- and subsoil organic carbon stocks of Low-Input High-Diversity systems using soil and vegetation characteristics, Sci. Total Environ., 589, 153–164, doi:10.1016/j.scitotenv.2017.02.116, 2017.

Ozaki, Y., Yano, M. and Onishi, M.: Effective hyperparameter optimization using Nelder-Mead method in deep learning, IPSJ Trans. Comput. Vis. Appl., 9(1), 20, doi:10.1186/s41074-017-0030-7, 2017.

Pachepsky, Y. A. and Rawls, W. J., Eds.: Development of Pedotransfer Functions in Soil Hydrology, Elsevier Science. Developments in Soil Science, Amsterdam., 2004.

Patil, N. G. and Singh, S. K.: Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review, Pedosphere, 26(4), 417–430, doi:10.1016/S1002-0160(15)60054-6, 2016.

Peters, T. and Richter, M.: Climate Station Data Reserva Laipuna mountain peak, [online] Available from: http://www.tropicalmountainforest.org/data_pre.do?citid=963 (Accessed 6 September 2017a), 2011.

Peters, T. and Richter, M.: Climate Station Data Reserva Laipuna valley, [online] Available from: http://www.tropicalmountainforest.org/data_pre.do?citid=964 (Accessed 6 September 2017b), 2011.

Pierezan, J., Freire, R. Z., Weihmann, L., Reynoso-Meza, G. and dos Santos Coelho, L.: Static force capability optimization of humanoids robots based on modified self-adaptive differential evolution, Comput. Oper. Res., 84, 205–215, doi:10.1016/j.cor.2016.10.011, 2017.

Poulenard, J., Podwojewski, P. and Herbillon, A. J.: Characteristics of non-allophanic Andisols with hydric properties from the Ecuadorian páramos, Geoderma, 117(3–4), 267–281, doi:10.1016/S0016-7061(03)00128-9, 2003.

Price, K., Storn, R. and Lampinen, J.: Differential Evolution. A Practical Approach to Global Optimization, edited by G. Rozenberg, T. Bäck, A. E. Eiben, J. N. Kok, and H. . Spaink, Springer, Berlin, Heidelberg, New York., 2005.

Qin, A. K., Huang, V. L. and Suganthan, P. N.: Adaptation for Global Numerical Optimization, IEEE Trans. Evol. Comput., 13(2), 398–417, 2009.

Reeves, C. R., Ed.: Modern Heuristic Techniques for Combinatorial Problems, John Wiley & Sons, Inc., New York, NY, USA., 1993.

Ridgeway, G.: Generalized Boosted Models: A guide to the gbm package, , 1–15, doi:10.1111/j.1467-9752.1996.tb00390.x, 2012.

Ridgeway, G.: R package 'gbm': Generalized Boosted Regression Models, [online] Available from: https://cran.r-project.org/web/packages/gbm/index.html, 2017.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

Rocha Campos, J. R. da, Silva, A. C., Fernandes, J. S. C., Ferreira, M. M. and Silva, D. V.: Water retention in a peatland with organic matter in different decomposition stages, Rev. Bras. Ciência do Solo, 35(4), 1217–1227, doi:10.1590/s0100-06832011000400015, 2011.

Sagar, R., Raghubanshi, A. S. and Singh, J. S.: Tree species composition, dispersion and diversity along a disturbance gradient in a dry tropical forest region of India, For. Ecol. Manage., 186(1–3), 61–71, doi:10.1016/S0378-1127(03)00235-4, 2003.

Schwärzel, K., Renger, M., Sauerbrey, R. and Wessolek, G.: Soil physical characteristics of peat soils, J. Plant Nutr. Soil Sci., 165, 479–486, 2002.

Schwärzel, K., Šimůnek, J., Stoffregen, H., Wessolek, G. and van Genuchten, M. T.: Estimation of the Unsaturated Hydraulic Conductivity of Peat Soils: Laboratory versus Field Data, Vadose Zo. J., 5(2), 628, doi:10.2136/vzj2005.0061, 2006.

Shang, S.: Log-Cubic Method for Generation of Soil Particle Size Distribution Curve, , 2013, 2013.

Shein, E. V. and Arkhangel'skaya, T. A.: Pedotransfer functions: State of the art, problems, and outlooks, Eurasian Soil Sci., 39(10), 1089–1099, doi:10.1134/S1064229306100073, 2006.

Singh, M. K., Astley, H., Smith, P. and Ghoshal, N.: Soil CO2-C flux and carbon storage in the dry tropics: Impact of land-use change involving bioenergy crop plantation, Biomass and Bioenergy, 83, 123–130, doi:10.1016/j.biombioe.2015.09.009, 2015.

Slowik, A. and Bialko, M.: Training of artificial neural networks using differential evolution algorithm, in Conference on Human System Interactions, pp. 60–65, Krakow., 2008.

Storn, R. and Price, K.: Differential evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces., 1995.

Tien Bui, D., Nguyen, Q. P., Hoang, N. D. and Klempe, H.: A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS, Landslides, 14(1), 1–17, doi:10.1007/s10346-016-0708-4, 2017.

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G. and Tóth, G.: New generation of hydraulic pedotransfer functions for Europe, Eur. J. Soil Sci., 66(1), 226–238, doi:10.1111/ejss.12192, 2015.

Twarakavi, N. K. C., Šimůnek, J. and Schaap, M. G.: Development of Pedotransfer Functions for Estimation of Soil Hydraulic Parameters using Support Vector Machines, Soil Sci. Soc. Am. J., 73(5), 1443, doi:10.2136/sssaj2008.0021, 2009.

Vasques, G. M., Coelho, M. R., Dart, R. O., Oliveira, R. P. and Teixeira, W. G.: Mapping soil carbon, particle-size fractions, and water retention in tropical dry forest in Brazil, Pesqui. Agropecu. Bras., 51(9), 1371–1385, doi:10.1590/S0100-204X2016000900036, 2016.

Vereecken, H., Javaux, M., Weynants, M., Pachepsky, Y., Schaap, M. G. and Genuchten, V.: Using pedotransfer functions to estimate the van genuchten- mualem soil hydraulic properties: A review, Vadose Zo. J., 9(4), 795–820, doi:10.2136/vzj2010.0045, 2010.

https://doi.org/10.5194/soil-2019-72
Preprint. Discussion started: 4 November 2019

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C. B., Knadel, M., Morrás, H. J. M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. M. R., Sanborn, P., Sellitto, V. M., Sudduth, K. A., Rawlins, B. G., Walter, C., Winowiecki, L. A., Hong, S. Y. and Ji, W.: A global spectral library to characterize the world's soil, Earth-Science Rev., 155(January), 198–230, doi:10.1016/j.earscirev.2016.01.012, 2016.

Wang, S., Zhuang, Q., Wang, Q., Jin, X. and Han, C.: Mapping stocks of soil organic carbon and soil total nitrogen in Liaoning Province of China, Geoderma, 305(120), 250–263, doi:10.1016/j.geoderma.2017.05.048, 2017.

Weiss, R., Alm, J., Laiho, R. and Laine, J.: Modeling moisture retention in peat soils, Soil Sci. Soc. Am. J., 62(2), 305–313, doi:10.2136/sssaj1998.03615995006200020002x, 1998.

Witten, I. H. and Frank, E.: Practical machine learning tools and techniques, 2nd ed., edited by J. Gray, Elsevier, San Francisco., 2005.

Witten, I. H., Frank, E. and Hall, M. A.: Data Mining Practical Machine Learning Tools and Techniques, 3rd ed., Morgan Kaufmann., 2011.

Yang, R. M., Zhang, G. L., Liu, F., Lu, Y. Y., Yang, F., Yang, F., Yang, M., Zhao, Y. G. and Li, D. C.: Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem, Ecol. Indic., 60, 870–878, doi:10.1016/j.ecolind.2015.08.036, 2016.

Yang, Y. H., Xu, X. Bin, He, S. B., Wang, J. B. and Wen, Y. H.: Cluster-based niching differential evolution algorithm for optimizing the stable structures of metallic clusters, Comput. Mater. Sci., 149(March), 416–423, doi:10.1016/j.commatsci.2018.03.055, 2018.