Thank you for your time and valuable comments.
Please find our replies below each comment. Numbers were added to each comment. Replies are displayed in blue. Line numbers in the replies refer to the revised manuscript with tracked changes. Line numbers in the reviewer comments refer to the first manuscript version.

Kind regards,
Anika Gebauer

General comments

The authors derived pedotransfer functions also for soils which rarely have information on soil hydraulic properties: organic soils under volcanic influence with low bulk density and high organic carbon content. Soil hydraulic behaviour of these soils are unique, pedotransfer functions derived on mineral, non volcanic soils cannot be successfully applied for them. The presented study fills a gap related to describing soil hydraulic properties of an underrepresented soil system. For those, who are non-experts in the topic of the manuscript, it would be important to explain why separate pedotransfer functions were developed for the two studied sites.
The manuscript is well structured, the methods used to derive and optimize the predictions are adequate. Machine learning methods are common tools to predict soil properties. The selection of method to optimize the parameters of a particular machine learning method depends on the size of the dataset (number of samples in the training dataset), number of predictors, type of algorithm (how many parameters have to be optimized) and computation capacity. It would be important to mention and discuss these factors when performance of grid search and differential evolution algorithm is compared.
It is very positive that the newly derived PTFs are compared with existing PTFs – derived for tropical soils and available from the literature –, but it is not clear how many samples of the Laipuna site were used for it. The number of samples used for the analysis are generally not clear in the text, suggestions for clarification are included in the specific comments below.
It would be more informative to show results based on unscaled values, i.e.: Figures 6-8 and lines 186-191 and 219-223. It would enhance comparison of the results with the literature.
Authors could put more focus on soil physical interpretation of the results. Prediction of soil hydraulic properties of such specific soils as presented in the study is particular and very useful. The novelty of the paper could be connected to this. It is not mentioned how the derived pedotransfer functions can be accessed.
Thank you for your general comments. Our answers can be found below, next to the corresponding specific comments.

Specific comments

C1: L1-2: please specify in the title somehow that you derived PTFs to predict soil hydrological properties.
We changed the title to "Development of pedotransfer functions for water retention in tropical mountain soilscapes: Spotlight on parameter tuning in machine learning"

C2: L15: please add name of the country.
The name of the country was added (line 16).

C3: L82, 88: could you please add the WRB (IUSS Working Group WRB: World Reference

Base for Soil Resources 2014. International soil classification system for naming soils and creating legends for soil maps, Rome, 121 pp., 2014.) name of the most typical soils occurring in the studied sites?

Due to monetary short cuts in the project grants we do not have all the required laboratory data in order to classify the soil profiles according to WRB. This is why we refrain from mentioning specific soil reference groups.

C4: L98-99: please give number of soil profiles and soil samples, instead of number of sites and sampling depth.

Lines 116 f. were adapted.

C5: L99: Please add e.g. suction applied in hPa or matric potential head in cm. Why did you choose to measure water retention at pF 0, 0.5, 1.5 and 2.5? What is the reason for not determining water retention at pF 4.2?

The suction applied in hPa was added to lines 118 f.. Water retention at pF 4.2 was not determined because of high soil organic matter contents in Quinuas (added to lines 121 ff.). To be able to compare the model input of both research areas, water retention was measured at the same pF values.

C6: L100: please add reference for the determination of soil water retention and BD.

References were added in line 119 and line 123.

C7: L101-103: please add if it is the standard method in Germany for the determination of PSD.

There are several methods to determine soil texture. We refrain from calling one of them a standard method. The applied methodology is given in lines 126ff.

C8: L104: if I have understood it well, you didn't measure PSD for the Quinuas site because of the high SOC content. Please mention it here and shortly describe the reason for it.

Please refer to line 129 and lines 95-96.

C9: L105: please add reference to $CaCO_3$ determination.

Adapted accordingly in lines 124f.

C10: L126: it would be helpful to shortly summarize what happens in 1) grid search and 2) with the differential evolution algorithm.

Lines 164 ff. were adapted accordingly. The differential evolution algorithm is explained in lines 169 ff..

C11: L129: … were compared in grid search… Please add the name of R package you used to apply the grid search for tuning the parameters.

We did not use an R package for grid search and programmed it on our own.

C12: L134: please define here the meaning of v. Is the meaning of v = 100 the same in L139?

Yes, adapted accordingly in line 174.

C13: L147: add somewhere in the materials and methods section which soil variables you use as predictors by sites. In the present manuscript reader gets information about it only from Fig. 4. under Results and discussion section.

Soil variables that were used as predictors were explained in Section 2.2 (added to lines 122 f.).

C14: L177: please add that the description is in the text e.g.: … to the modelling steps

described in text …
The caption of Fig. 3 was adapted accordingly.

C15: L182: what do you mean that number of samples was 51 and 46? Please rephrase the sentence accordingly.
The sentences in lines 237 ff. were rephrased.

C16: L186-191: please add unscaled RMSE value with unit as well already here, because readers are familiar with that.
Lines 186-191 (now lines 235ff.) describe the dataset shown in Figure 4 and 5. No RMSE values are shown.
The BRT models were trained on scaled values, which is why we provided scaled RMSE values in Section 3.2. A text section to explain this aspect was added in lines 135 ff..

C17: L199-201: The two sentence could be concatenated: the one starting with "Measured BD …" and the other starting with "The water …".
The two sentences are separated as both refer to different literature references.

C18: L203: please explain what you mean by "correspond to soil samples with a higher proportion of mineral components or andic properties".
The sentence in lines 260 f. was rephrased.

C19: L208: you could highlight here why it is an interesting dataset for deriving a new hydraulic pedotransfer function.
Please see lines 36 ff. and lines 79 f..

C20: L210: Figure 4: - it would help comparison of Quinuas and Laipuna data if the min. and max. values of y-axis would be the same, you could include violin plot of both sites in one plot: one plot for OC and another for BD,
- add in caption that PSD of Quinuas was not measured, and shortly add reason for it,
- instead of showing the cumulative distribution of the PSD (Fig 4. c)) texture triangle diagrams separately would be more informative,
- please add number of samples to the figure, e.g: in title or caption.
The violin plots of both sides were not included into one plot because BD and SOC ranges are rather different for Laipuna and Quinuas.
The cumulative PSD distribution was used instead of separate texture triangle diagrams, as it allows to show mean values and standard deviations per texture class (added to the caption of Fig. 4). The caption of Fig. 4 was adapted to explain why PSD was not measured for Quinuas and to add the number of samples.

C21: L213: Figure 5: - please add number of samples to the figure, e.g: in title or caption.
The number of samples was added to the caption of Fig. 5.

C22: L218: you didn't mention in Materials and methods that you use scaled water retention values in the algorithm, please add it there and the reason for it there.
Please see comment C16.

C23: L219-223: please add unscaled RMSE value with unit as well.
Please see comment C16.

C24: L232: … models, regarding $RMSE_E$ and R_E values …
The sentence relates to the difference between the scatterplots in Fig. 8 and 9.

C25: L234: please consider to delete “However,”.
“However” was deleted (line 292).

C26: L237-238: please consider if number of samples can influence the performance of parameter tuning in sentence starting with “Probably”. Maybe it could be discussed how performance of tuning methods would change if you could include other predictors as well, e.g.: pH, CEC, etc.
The result of every tuning method (i.e. the parameter values) indirectly depends on the dataset, as parameter tuning means adjusting a model to the specific modelling problem / dataset (lines 52 f.). As discussed in e.g. lines 312 ff. or lines 322 ff. the number of samples and the number of predictors (and their information content) influence the performance of the BRT models. In general, a larger dataset of high quality results in more explicit relationships between response and predictor variables that can be detected and reproduced more easily by a model. This might reduce the required differential optimum iterations and the probability of getting stuck into a local optimum (added to lines 440 ff.).

Concerning the general questions:
The optimization methods should be chosen based on the type of model algorithm. Models with real valued parameters require a different tuning technique than those with discrete valued parameters (lines 55 ff.). The selection of the tuning technique does not depend on the number of parameters to be tuned. Grid search and optimization algorithms allow the tuning of only one parameter as well as the tuning of several parameters.
The computation capacity is still a limiting factor when it comes to optimization. But with computers becoming more efficient and the possibility of parallelization, a parameter tuning technique should be judged based on the predictive power of the resulting machine learning model and not on the required computing resources.

C27: L242-243: please mention under materials and methods the mean stone content of the Laipuna samples, if stoniness is characteristic for those.
The stone content was not determined for all samples.

C28: L245: It is not clear what predictors were used to predict water retention of Quinuas. Please add it as mentioned before. It could be explained which suction heads can be covered by the predictors you have for Quinuas. Sentence starting with “PSD” should be moved under Materials and methods section, please see previous comments.
Soil variables used as predictors are mentioned in lines 122 f. (please see comment C13). Reasons, why PSD measurements were not possible for Quinuas, are explained in lines 129 f. (please see comment C8). The sentences in lines 313 ff. were adapted.

C29: L246: Why performance of Laipuna PTFs for pF0, pF 0.5 and pF 1.5 is lower that that of Quinuas? Please discuss how those could be improved.
We detected an error in Fig. 7 that was corrected: R^2 values of the differential evolution pF 0 Laipuna models range from 0.64 to 0.78 and are similar to those achieved in Quinuas. No significant difference between models of both research areas can be detected based on the R^2 values and the scatter plots shown in Fig. 8 and 9.
It was discussed how to improve model performance in different sections of the paper (please see comment C36).

C30: L253-273: Please add title to that section, to highlight that you applied existing PTFs on the sites to compare the performance of the newly derived PTFs to those.
A new Section (3.3) was added under Results and discussion.

C31: L253: please add number of samples of the Laipuna dataset. Did you use the test set for the comparision?
Please see comment C37.

C32: L254: … PTFs from the literature were selected … Or add something similar.
"from the literature" was added to line 223.

C33: L254-256: please move it under Materials and methods.
The selection of existing PTFs to be applied on the Quinuas and Laipuna datasets was explained in a new Section (2.6) under Materials and methods.

C34: L256: it might be more precise to write that silt and sand content was converted to 2-50 _m and 50-2000 _m fractions by spline interpolation to calculate the USDA texture classes.
The sentences in lines 226 ff. were adapted.

C35: L263: add unit of RMSE.
Please see comment C37.

C36: L266-267: please mention other factors as well which could increase the performance of the PTFs.
The predictive performance of a model is affected by three factors: 1) The adjustment of the algorithm to the modelling problem by parameter tuning, which is discussed in Section 3.3. 2) The quality and size of the dataset forming the model input, which is discussed in lines 312 ff., 322 ff. and 440 f.. And 3) the chosen model algorithm, which was added to lines 326 f..

C37: L271: Table 2: - add number of samples – by pF values – used to test the newly derived and existing PTFs, did you use the test set of Laipuna dataset? - please use also here pF 0, 0.5, 1.5 and 2.5 instead of Theta 0, 0.5, 1.5 and 2.5., - add unit of the RMSE.
The same test sets were used to evaluate the performance of the newly developed PTFs and the existing PTFs. Using cross validation each sample was used for testing (please see Section 2.5). In Table 2 the numbers of response – predictor variable data pairs were added and "Theta" was changed to "pF". The RMSE values are shown without the unit (please see comment C16).

C38: L275-278: for easier comparison Figures 6 and 7 could be concatenated by using grouped boxplots.
The boxplots of Fig. 6 and Fig.7 were not concatenated as models built on datasets of two research areas with very different properties cannot be compared directly.

C39: L280-285: based on Figure 5 observed pF values of Quinuas site is greater then 0.30 cm3/cm3, for Laipuna those are greater than 0.20 cm3/cm3. Please check in calculations why you have observed pF values close to 0 cm3/cm3 on Figures 8 and 9. Or are those scaled observed and scaled predicted variables? It would be more informative to show the scatterplot for not scaled observed and predicted values. Please revise Figures 8 and 9.
Fig. 8 and 9 show scaled values, which was added to the figure captions. As predicted and observed values were scaled in the same way, the not scaled scatterplots would look like the scaled ones – except for the axis labels. Scaled values were used to be able to use the same axis (between 0 and 1) for each plot. It enhances comparability without losing information.

C40: L293-294: Sentence starting with "Difference": there is difference between GS and DE in case of the bag fraction as well. Is it possible to show which parameter – among

number of trees, shrinkage, interaction depth, bag fraction – has the most dominant influence on the performance of BRT?
As mentioned in lines 412 ff. there is a difference in bag fraction, which cannot be explained (lines 430 f.). There might not even be an optimum for bag fraction (lines 428 f.).
We made an assumption about the parameter importance in lines 427 f.. A sensitivity analysis could be used to estimate the importance of each parameter. As the model performance depends on the interaction of all parameter values (added to lines 426 f.), we recommend tuning the number of trees, shrinkage, interaction depth and bag fraction simultaneously.

C41: L318: … for the final differential evolution models derived for Laipuna site (Fig. 9) …
The sentence in lines 425 f. was adapted.

C41: L334, 339: In the caption of Figure 10 and 11: add number of tested parameter vectors for both method.
The number parameter vectors tested by grid search is mentioned in lines 167 f.. In accordance with the review of Mr. Wadoux the number of differential evolution iterations was added to lines 397 ff.. One iteration compared 100 parameter vectors (line 179).

C42: L343-344: please note that in most of the cases local PTFs perform better than PTFs trained on dataset originating from elsewhere with different soil forming factors. Please revise the sentence.
Please see lines 35 ff.). The comparison to other existing PTFs that were developed under conditions as similar as possible (lines 222 f.) is important to legitimate the developed BRT PTF and should be mentioned in the conclusions.

C43: L351-354: please consider to concatenate the last two sentence of the conclusions to better balance highlight both on the newly derived PTFs and results of comparing parameter tuning methods.
The conclusion was divided into two new paragraphs to highlight the two achievements. The last two sentences were concatenated (lines 464 f.)

C44: L354: please consider to provide availability of derived PTFs – which you recommend to use – for users.
The developed PTFs, as well as the underlying datasets, will be uploaded to the Open Science Framework (OSF) upon acceptance (added to lines 468 ff.).

Technical comments

TC1: L67: Please add the country after Páramo.
"Ecuador" was added after "Páramo" (line 79).

TC2: L151: The acronym of BRT is not included in the flowchart.
"BRT = boosted regression trees" was removed from line 192.

TC3: L343: … readily available …
„ready" was changed to „readily" (line 455).