

# Revision of manuscript on “Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (euptfv2)”

## Content of the document

I. POINT BY POINT AUTHORS' RESPONSE TO THE REVIEWERS.....	1
1. RESPONSE TO REFEREE #1.....	1
2. RESPONSE TO REFEREE #2.....	1
3. RESPONSE TO REFEREE #3.....	5
II. LIST OF AUTHOR’S CHANGES IN MANUSCRIPT AND SUPPLEMENTARY MATERIAL .....	17
III. MARKED-UP MANUSCRIPT AND SUPPLEMENTARY MATERIAL VERSION .....	18

## I. POINT BY POINT AUTHORS' RESPONSE TO THE REVIEWERS

### 1. RESPONSE TO REFEREE #1

Thank you for the positive review.

### 2. RESPONSE TO REFEREE #2

Thank you for the review and constructive comments. We addressed the comments in a revised version of the article. Below we give details on exactly how we addressed the concerns raised by anonymous referee 2. Please note the following during reading the responses:

- the responses are in blue regular font and follow the referee’s questions (RC2),
- new text parts that were added to the manuscript are in blue italic font,
- the reference to the lines (L) and pages (P) relates to the marked up version of the manuscript available in III. MARKED-UP MANUSCRIPT AND SUPPLEMENTARY MATERIAL VERSION section of this document.

#### RC2:

Soil pedotransfer functions are important when used for estimation of soil hydraulic parameters in catchment, regional, or continental scale applications. This manuscript improves the estimation of euptfv1 and provides information about prediction uncertainty, and can be applied for more predictor variable combinations than the euptfv1. Overall, the manuscript is interesting, important, well written, and organized in a logical well. Therefore, I recommend accepting this manuscript after minor revisions that are required to address the general and specific comments provided below.

A: Thank you for the positive general comment.

**RC2:**

1. The authors compared the estimation of water content at saturation, field capacity, wilting point, plant available water content, saturated hydraulic conductivity, etc., individually. I think these sections are somewhat lengthy. However, the most interesting part of the comparisons between point and parameter predictions and euptfv1 and v2 are very short. Is it possible to extend the comparisons and the discussion?

AGREED.

A1: Regarding comparison between point and parameter predictions, we added the following to be more specific:

P12 L30-31:

*„ ... more accurate and for further 8 cases RMSE were smaller.”*

P12 L31 – P13 L2:

*„The reason for higher RMSE in parameter estimation can be that the VG model does not always adequately describe the measured MRC data (Weber et al., 2019). Therefore, when THS, FC, FC\_2 and WP are computed with parameter estimation those are not only affected by the uncertainty of the prediction of VG parameters but by the goodness of VG model fit as well.”*

P13 L4-6: we rephrased the sentence to make it clearer, which now reads:

*„For THS point estimation performed better than parameter estimation. When the moisture retention curve is not needed, but only THS and/or FC/FC\_2 and/or WP, we recommend to compute those with ...”*

In order to include the suggested comparison, between euptfv1 and v2, we included the following sentences

P13 L9-13:

*“The most important reason for it can be that the interaction between the target variable and the predictors is more complex for the cases of predicting FC or VG parameters – to describe the MRC –, which can be untangled using random forest. This may provide a reasons the random forest algorithm performed significantly better than the PTFs derived with linear regression or a simple regression tree.”*

P13 L14-L21:

*„The RMSE of THS prediction was somewhat lower for euptfv1 than for euptfv2, but the difference was not significant. It could be due to the close to linear relationship between THS and BD and high relative importance of BD in THS prediction (84 %). This way their interaction can be efficiently described with the linear regression which is capable to extrapolate as well. Extrapolation with the random forest algorithm is not possible outside the training data, which can limit its performance. The general improvement of the PTFs in euptfv2 is threefold, it is due to i) using random forest instead of single regression tree or linear regression, ii) including more detailed information on soil sampling depth, not only distinguishing topsoils and subsoils and iii) providing information on prediction uncertainty.*

Regarding the description of the individual point and parameter estimations we kept the details because we think it is instructive to provide information about the importance of specific predictors.

**RC2:**

2. The authors listed so many PTFs. When I was reading the conclusion part, I cannot find which PTF I should use. Is it possible to make some concluding remarks regarding which PTFs should be used for corresponding predictors? I think this will be very helpful for future readers.

AGREED.

A2: Thank you very much for this very helpful comment. Indeed, it is very important that users should easily understand which PTF to select and apply. To achieve this, we i) added a dedicated paragraph on it above the Conclusion section, ii) highlighted in the abstract and short summary that this section was provided, iii) moved Table S3 from supplementary material to the manuscript as Table 11.

The new paragraph 4 reads P13 L25 – P14 L19:

***“4. Practical guidance on how to use the PTFs***

*The minimum input requirements for all PTFs are sand, silt and clay content, and soil depth. Soil depth needs to be considered in regard to the depth of the other input properties and soil hydraulic data needs, e.g. if the soil hydraulic properties of the top 20 cm (0-20 cm) is needed, then depth needs to be set at 10 cm in the input data of the prediction.*

*If only soil texture information is available for the predictions, the class PTFs from euptfv1 could be applied (Tóth et al., 2015).*

*We emphasise that:*

- 1. the units of input soil properties (predictors) have to be the same as indicated in the text and that the sand, silt, and clay are defined by the following particle diameters: clay < 2  $\mu\text{m}$ , silt between 2 and 50  $\mu\text{m}$ , and sand between 50 and 2000  $\mu\text{m}$ ,*
- 2. when only specific water content values at saturation, field capacity or wilting point are required (ie. THS, FC\_2, FC, WP) it is recommended to use point PTFs. This is also true for the prediction of KS,*
- 3. for AWC, the most accurate way is by first predicting FC and WP with the point predictions and then compute AWC using Eq. (1), and similarly for AWC\_2 using FC\_2 and Eq. (2),*
- 4. it is recommended to do the VG prediction if only moisture retention curve parameters are needed, and*
- 5. the MVG prediction when both moisture retention and hydraulic conductivity parameters are required.*

*The VG algorithms predict the following van Genuchten model parameters: the residual water content  $\theta_r$  ( $\text{cm}^3 \text{ cm}^{-3}$ ), the saturated water content  $\theta_s$  ( $\text{cm}^3 \text{ cm}^{-3}$ ), and shape parameters  $\alpha$  ( $\text{cm}^{-1}$ ) and  $n$  (-). Parameter  $m$  is provided based on  $m=1-1/n$  (van Genuchten, 1980), and for the hydraulic conductivity curve, the two additional parameters:  $K_0$  ( $\text{cm day}^{-1}$ ) the hydraulic conductivity acting as a matching point at saturation and  $L$ , the shape parameter related to pore tortuosity (-).*

*Table 11 shows the recommended PTFs for each predicted soil hydraulic property and available predictor variables. The users need to check which basic soil properties are available for the predictions, then look in Table 11 which PTF is recommended to use. The algorithms have been implemented in a web interface to facilitate the use of the PTFs, where the PTFs' selection is automated based on soil properties available for the predictions and required soil hydraulic property. The Code and data availability section provides information on how to access this resource.”*

The additional text in the short summary and the abstract is given by:

Short summary:

*"... The influence of predictor variables on predicted soil hydraulic properties is explored and practical guidance on how to use the derived PTFs is provided. ..."*

Abstract (P1 L25):

*"... for the prediction of water content at -100 cm matric potential head and plant available water content. A practical guidance on how to use the derived PTFs is provided."*

**Specific comments:**

**RC2:**

1. Figures 2, 5, and 6: Is it possible to include R2 in these figures? This will make the comparison between different figures easier.

AGREED

A1: We added R2 to Figures 2 (P34), 5 (P37), 6 (P38) and S1 (P2 in supplementary material).

**RC2:**

2. In the abstract and conclusion sections: -15.000 should be -15,000

AGREED

A2: Thank you for noting it, we corrected it in the entire text.

**RC2:**

3. Page 6, line 4: why did the authors utilize median values instead of mean values?

Nothing changed.

A3: Our aim was to provide information about the uncertainty of the predictions, therefore we applied quantile regression forests. This way the most probable predicted response value is at the 50<sup>th</sup> percentile, i.e. the median, which is considered more robust against the outliers than the mean. In this way we decided to use the median as the predicted value ( $\hat{y}$ ) rather than the mean.

**RC2:**

4. Page 7, line 19: "in the study of (Khodaverdiloo et al., 2011)" should be "in the study of Khodaverdiloo et al. (2011)"

AGREED

A4: Thank you, we corrected it (P7 L30).

**RC2:**

5. Page 10, line 4: "and RMSE" should be "an RMSE"

AGREED

A5: Thank you, we corrected it (P10 L19).

**RC2:**

6. Page 10, Line 27: ";" should be ","

AGREED

A6: Thank you, we corrected it (P11 L11).

**RC2:**

7. Page 12, line 14: add a connection/linking word before “it is due to”

**AGREED**

A7: The text is rephrased to (P13 L18-19): *“The general improvement of the PTFs in euptfv2 is threefold, the better performance is due to ...”*.

### 3. RESPONSE TO REFEREE #3

Thank you for the detailed review and suggestions for further improvements. In the following, we give a detailed presentation of how we addressed the questions and issues raised, following the referee’s questions (RC3). Please note the following during reading the responses:

- the responses are in blue regular font and inserted under the referee’s questions,
- new text parts that were added to the manuscript are in blue italic font,
- the reference to the lines (L) and pages (P) relates to the marked up version of the manuscript available in III. MARKED-UP MANUSCRIPT AND SUPPLEMENTARY MATERIAL VERSION section of this document.

**RC3:**

This manuscript aims to update the previously developed PTF for European soils called euptfv1. More importantly, euptfv2 contributes to the understudied issue of uncertainty in PTFs for potential users. Despite the existing large amount of results, the paper is easy to follow with some possibilities to improve.

A: Thank you for the positive general comment.

**RC3:**

The authors also provide a detailed and user-friendly website from euptfv2, however, no library called euptfv exists in R Repository, even the available zip file has problems to be run.

A: The R package of euptfv2 is under construction.  
The available zip files include the R scripts used to develop the predictions and the derived pedotransfer functions. The dataset which we used for training and testing the algorithms cannot be shared according to the agreement between the data holders. Regarding the model development the following information is included separately for point and parameter estimations: i) loading data, define path, input variables and function to compute performance of the PTFs (setupRF.R), ii) parameter tuning of the random forest (tuneRF.R), iii) building final random forest (buildfinalRF.R), iv) compute performance of the final random forest on the test set (testRF.R) .  
In a separate folder (<https://github.com/TothSzaboBrigitta/euptfv2/tree/master/help>) a sample input dataset (data\_sample.csv) and an R script (apply\_PTFs\_script.R) - which shows some examples on how to apply the PTFs in R – have been added to the repository.

**RC3:**

Many comparisons among the possibilities of PTFs for different soil hydraulic properties were done. These series of “euptfv(i)” will contribute to the modelling of soil processes. I recommend this paper

for publication, however, I outlined some questions and comments as below (L denotes line and P for page)

A: Thank you for considering the usability of the euptfs.

**RC3:**

L30, P1. variably saturated fluxes? do you mean flow through variably saturated soil media?

AGREED

A: Yes, thank you for noting it, we corrected it (P1 L30 – P2 L1): “Simulations of flow through variably saturated soil media either rely on ...”

**RC3:**

L31. P.2. Not necessary to machine learning-based methods are able to calculate uncertainty because the sampling effect can propagate parameter uncertainty, which can be implemented even in simple regression-based models. Tens of resamples for training and testing with different distributions can be drawn from the population (Tranter et al., 2010; Kotlar et al., 2019).

Do train and test datasets in bootstraps follow the same distributions?

Tranter, G., Minasny, B. and McBratney, A.B., 2010. Estimating pedotransfer function prediction limits using fuzzy k-means with extragrades. *Soil Science Society of America Journal*, 74(6), pp.1967-1975.

Kotlar, A.M., de Jong van Lier, Q., Barros, A.H.C., Iversen, B.V. and Vereecken, H., 2019. Development and Uncertainty Assessment of Pedotransfer Functions for Predicting Water Contents at Specific Pressure Heads. *Vadose Zone Journal*, 18(1).

AGREED

A: Thank you to highlight it with references, we added this information to P2 L17-19:

*“Tranter et al. (2010) developed an uncertainty estimation method using fuzzy k-means with extragrades classification that can be applied in any PTF prediction. Kotlar et al. (2019) presented uncertainty assessment of PTFs through deriving PTFs on tens of resamples for train and test sets.”*

and the following to P3 L1-3:

*“If PTFs are derived with these algorithms, the uncertainty of the predicted soil property can be directly estimated when applying the PTF (Szabó et al., 2019a), although this could also be achieved by applying the above mentioned uncertainty assessment methods without using machine learning methods (e.g. Kotlar et al., 2019; Tranter et al., 2010).”*

Using Kolmogorov–Smirnov tests, we tested whether training and test sets have the same frequency distributions, please find the results in Table 1. For THS, FC and WP the distribution of training and TEST\_BASIC set is equal in almost all the cases of the most important basic soil properties. For KS, the distribution of sand and organic carbon content is equal in the training and TEST\_BASIC set, in case of FC\_2 only the distribution of sand content is equal based on the statistical test. The distributions of training and TEST\_CHEM+ sets are equal only in case of FC. For the other sets, at least one soil property has equal distribution in the two sets.

Table 1. Results of the Kolmogorov–Smirnov test (p value of 0.05) computed to compare distribution of the most important basic soil properties of training and test datasets.

Soil hydraulic property	Input variable	p-value of Kolgomorov-Smirnov test	
		Training TEST_BASIC set	vs. Training TEST_CHEM+ set
THS	USSAND	0,137	0,000
	USCLAY	0,022	0,000
	OC	0,598	0,004
	BD	0,483	0,021
FC_2	USSAND	0,616	0,112
	USCLAY	0,004	0,000
	OC	0,018	0,000
	BD	0,023	0,000
FC	USSAND	0,019	0,157
	USCLAY	0,172	0,078
	OC	0,662	0,737
	BD	0,313	0,489
WP	USSAND	0,730	0,007
	USCLAY	0,372	0,003
	OC	0,649	0,000
	BD	0,047	0,074
KS	USSAND	0,396	0,000
	USCLAY	0,001	0,008
	OC	0,755	0,001
	BD	0,000	0,000

Train and test datasets in bootstraps are divided in the following way: in the random forest algorithm for each tree 63% of the data is selected with replacement to build the tree, i.e. number of selected data will be increased to reach the number of samples of the training set with the replacement, this way some samples will be used multiple times in a single tree. Each tree of the forest is trained on different samples. The forest includes 200 trees and the predicted value is the median of all 200 trees. However, it is difficult to compute the Kolgomorov-Smirnov test for all the 200 in-bag and out-of-bag samples by each predicted soil hydraulic properties, we could confirm based on the literature (Hastie et al., 2009), that the forest will neither be biased nor overfitted to the data because of the two step randomization – bagging process and split-variable randomization – implemented in the algorithm.

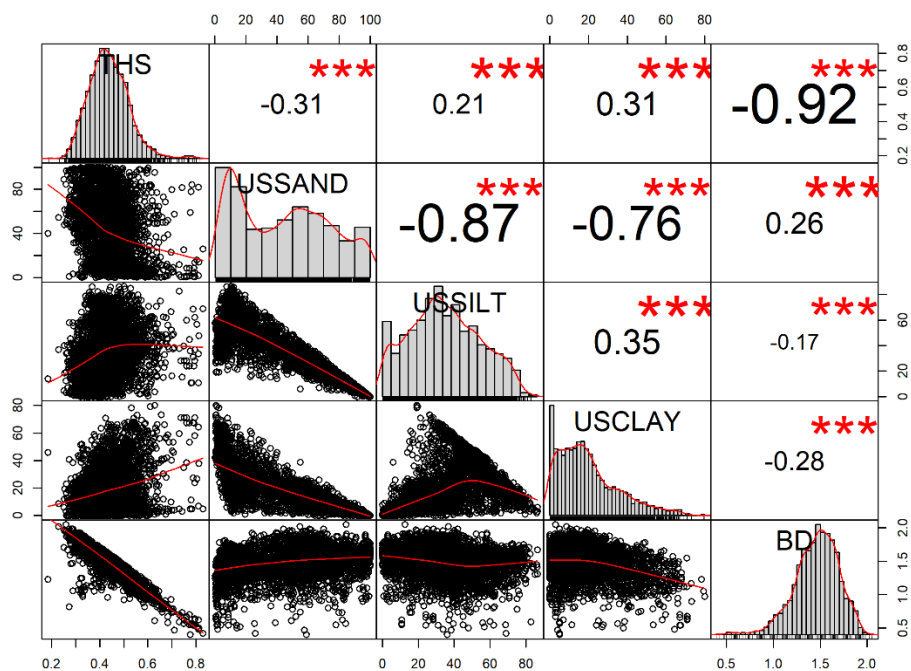
### RC3:

Table1. P.18. Numbers are not aligned exactly below the names.

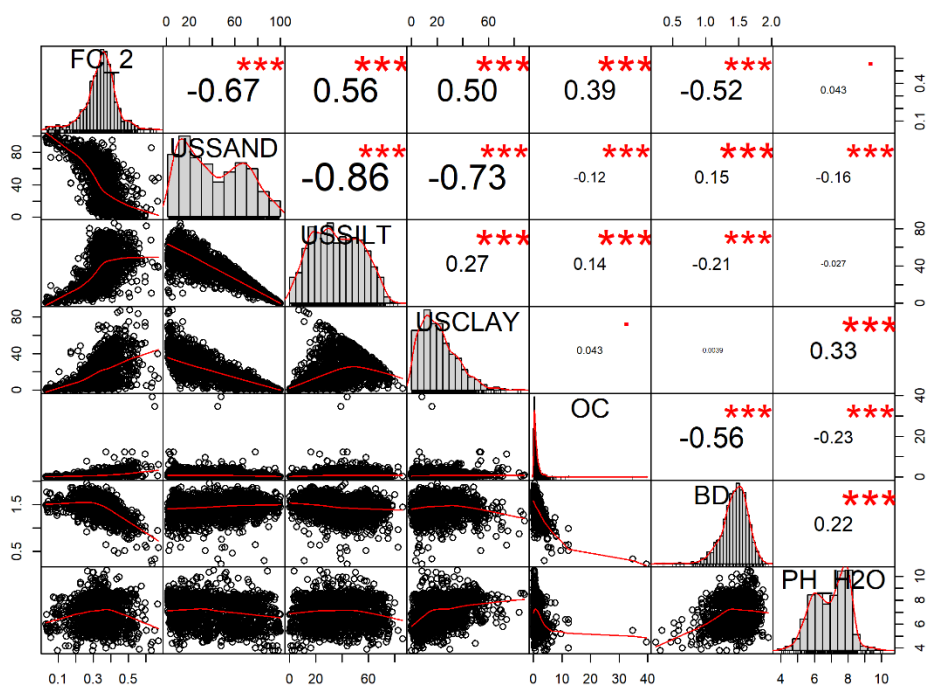
Correlation matrix of observations would be useful information (in appendix) at least for the dataset used for the best PTFs.

AGREED

- A: The columns' names were aligned with the numbers below (Table 1 P21). The correlation plots of the best PTFs are inserted below the answers (Fig\_responses\_1 – Fig\_responses\_7), however descriptive power of them are limited because the relationship between predicted parameters and predictors are not linear. This is the reason why PTFs are derived with a machine learning algorithm and partial dependence plots are shown in the manuscript. We feel that the correlation plots might not provide indispensable information.

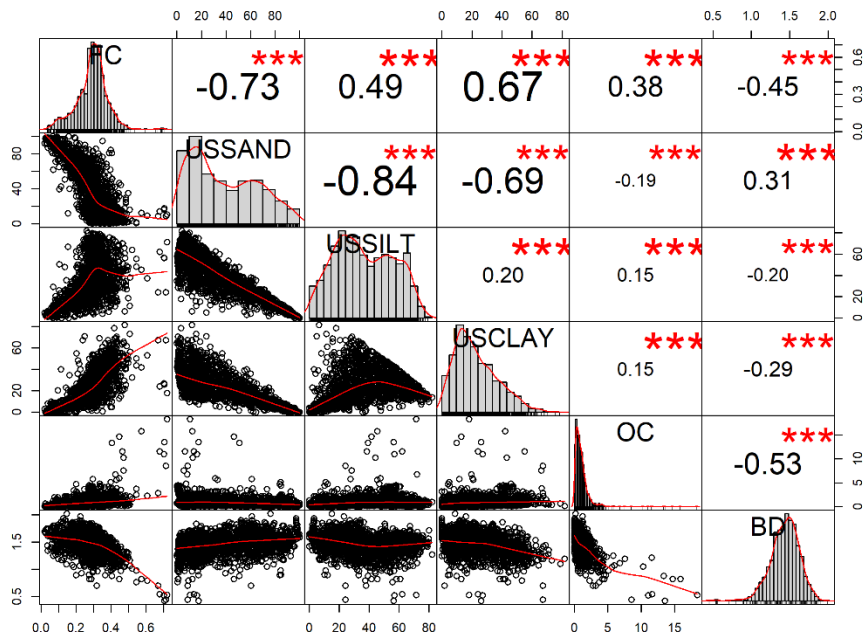


Fig\_responses\_1

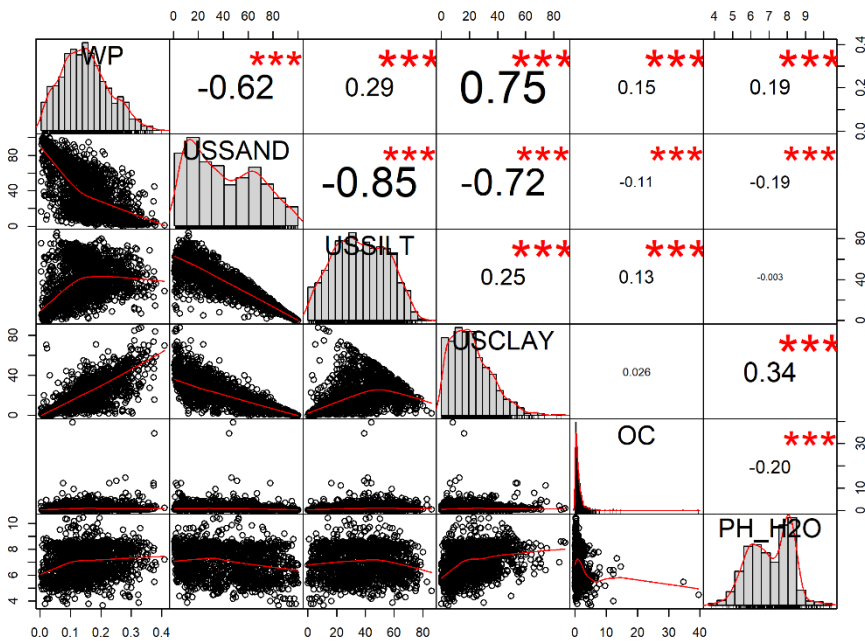


Fig\_responses\_2

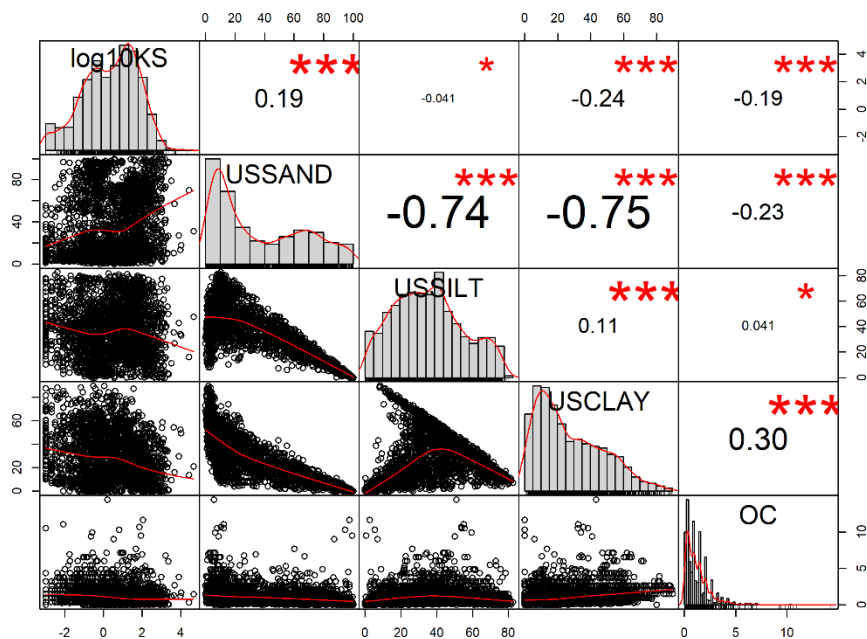




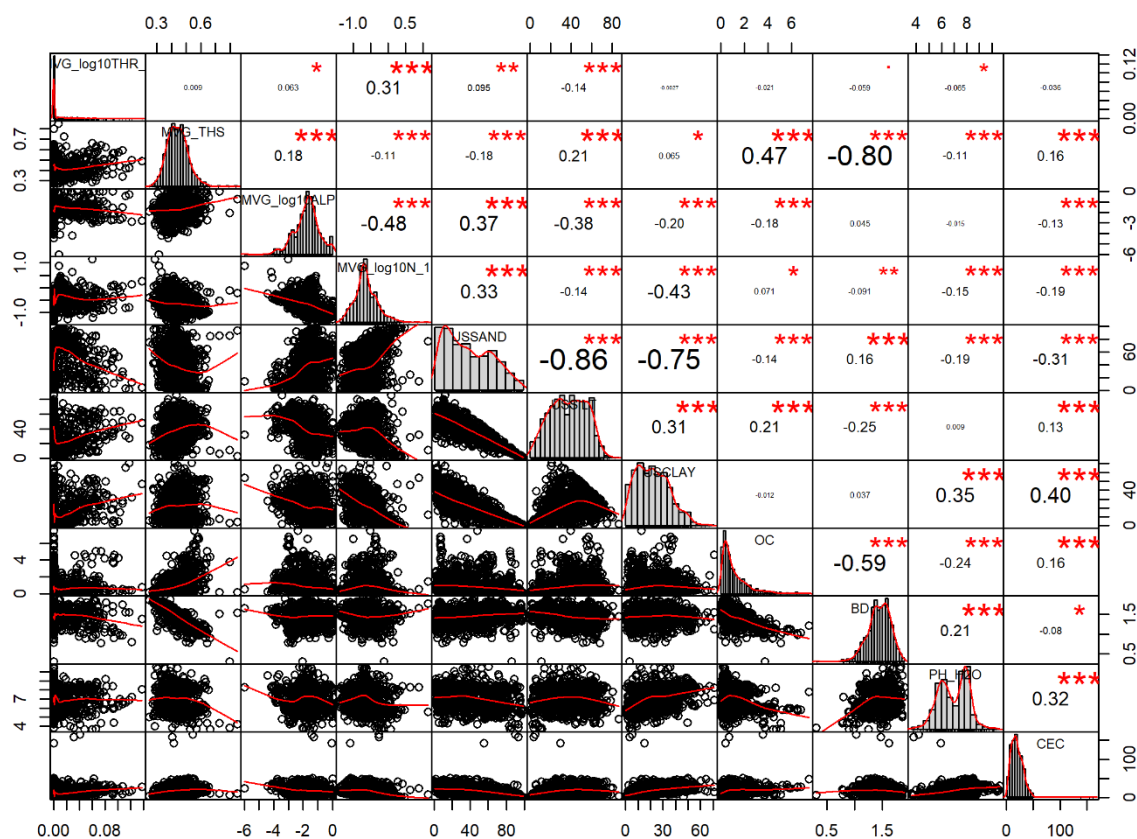
Fig\_responses\_3



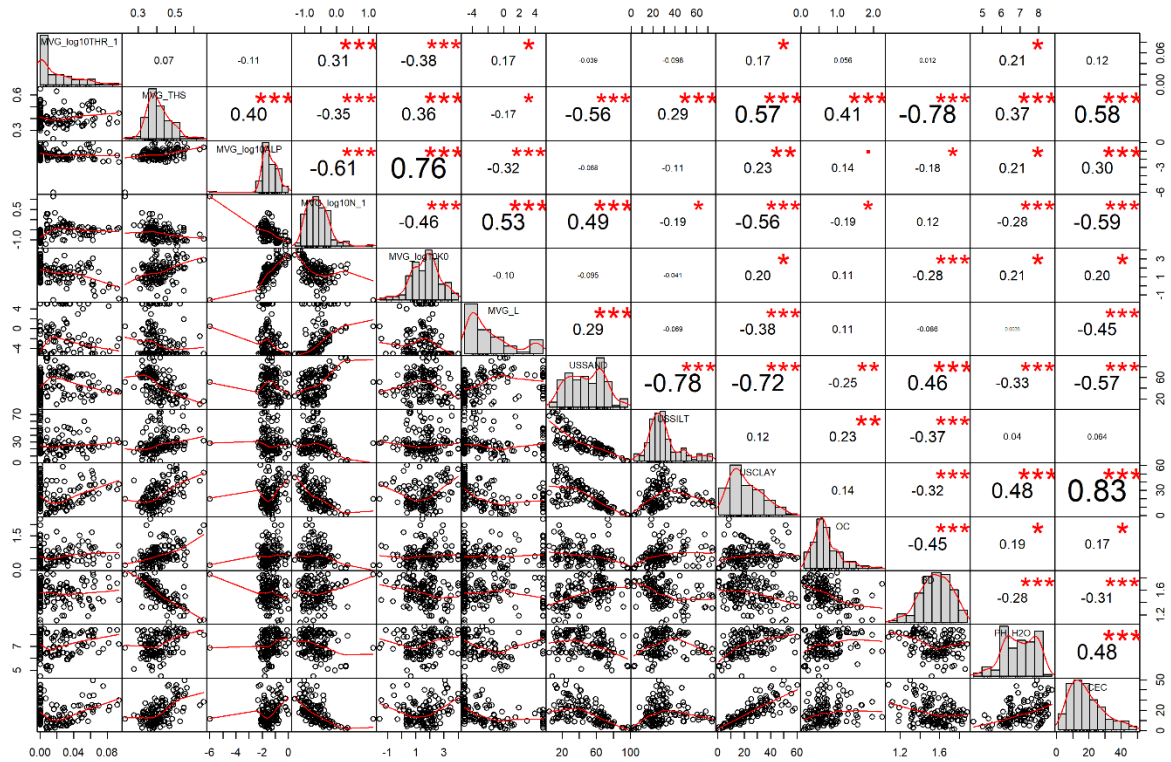
Fig\_responses\_4



Fig\_responses\_5



Fig\_responses\_6



Fig\_responses\_7

RC3:

L6, P5: Please calculate variable importance of parameters in PTFs as relative which makes summation of all 100%. (e.g. Figure 3)

AGREED

A: We replaced Figure 3 (P35) and 7 (P39) with relative importance plots and specified in the figures' caption that relative variable importance is shown.

We added the following text in P6 L1-2:

*"The relative importance was assessed by dividing the variable importance of each predictor by the sum of the importance of all the predictors after Kotlar et al. (2019)."*

RC3:

L3, P7: To give a better view of the performance of PTFs, compare the mean values of measured parameters with RMSE of predictions. Compared to Toth et al., (2015), improvement in the prediction of THS is less than FC and WP, why?

AGREED

A: Thank you for the suggestion. We added the normalized RMSE (RMSE/(ymax-ymin)), which was also suggested by the reviewer under "L1-8, P10".

The following texts was added in the revised manuscript:

P6 L15-19:

*"The different data range of the dataset influences the performance of the PTFs when that is compared to the studies in the literature. Therefore, normalized RMSE (NRMSE) was computed (Eq. 5.), where  $y_{max}$  and  $y_{min}$  are the maximum and minimum value of variable .*

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \quad (5)''$$

and in P7 L16-18 we add

*“Table S3 shows the NRMSE for the point predictions computed for the TEST\_BASIC and TEST\_CHEM+ sets to provide possibility for comparison with other PTFs available from the literature.”*

And a table on “Normalized root mean square error (NRMSE) of the point predictions by soil hydraulic properties computed on the test datasets in cm<sup>3</sup> cm<sup>-3</sup> for water retention and log<sub>10</sub> (cm day<sup>-1</sup>) for saturated hydraulic conductivity. In case of PTF01, 02, 03 and 07 TEST\_BASIC set was used for the analysis, for the rest of the PTFs TEST\_CHEM+ set was considered” was added to the supplementary material as Table S3 (P5 of the revised supplementary material). The original TableS3 of the supplementary material available from <https://gmd.copernicus.org/preprints/gmd-2020-36/gmd-2020-36-supplement.pdf> was moved to the manuscript as Table 11.

Comparison to Toth et al. (2015): thank you for the reviewer’s comment on THS and BD, which helps to clarify findings related to comparison of euptfv1 and v2. There was no significant difference between euptfv1 and v2 in case of THS when BD was available for the prediction and euptfv1 was derived with linear regression. The reason for it – which was mentioned by the reviewer as well – that the relative importance of BD is 84% in the prediction of THS and the relationship between THS and BD is close to linear. In this case random forest could not significantly improve the prediction. In case of FC and WP the interaction between the target variable and the predictors is more complex, this way the random forest algorithm performed significantly better than the PTFs derived with linear regression or a simple regression tree. We added the following information in P13 L9-18:

*“The most important reason for it can be that the interaction between the target variable and the predictors is more complex for the cases of predicting FC or VG parameters – to describe the MRC, which can be untangled using random forest. This may provide a reason the random forest algorithm performed significantly better than the PTFs derived with linear regression or a simple regression tree. For THS, WP, KS, and MVG only those PTFs did not improve significantly, for which comparisons on the TEST\_CHEM+ set were possible – which includes reduced number of samples. The RMSE of THS prediction was somewhat lower for euptfv1 than for euptfv2, but the difference was not significant. It could be due to the close to linear relationship between THS and BD and high relative importance of BD in THS prediction (84 %). This way their interaction can be efficiently described with the linear regression which is capable to extrapolate as well. Extrapolation with the random forest algorithm is not possible, which can limit its performance.”*

### RC3:

Figure S2, please replace SE by RMSE so the reader doesn’t lose the track of comparison criteria.

AGREED

A: We replaced Figure S2, S4, S6, S8, S10, S12, S14, S16, S19 showing SE with the one showing RMSE, please find the new figures on P6, 8, 10, 12, 14, 16, 18, 20, 23 of the revised supplementary material below (III. MARKED-UP MANUSCRIPT AND SUPPLEMENTARY MATERIAL VERSION).

**RC3:**

L1, P8. Please mention the correlation between THS and BD, lets arguably consider THS equal to total porosity, does the 1-BD/PD, assuming PD=2.65 give better RMSE than PTF03 for THS? or you might easily obtain the best PD to predict THS by this formula. In PTF 32, the relative importance of BD is almost 100%.

**DISAGREE**

A: Thank you for this idea, however, to remain consistent in methodology we would make use of the better performing PTF based on the random forest. The reason: the correlation between THS and BD is -0.92. We have computed the porosity on the test dataset of PTF03 (N = 1274) based on BD and PD (=2.65 g/cm<sup>3</sup>), then the RMSE of it. We found that the RMSE of PTF03 is smaller than that of porosity (POR\_calc), please find the performance of POR\_calc and PTF03 in the below table.

Method	ME (cm <sup>3</sup> cm <sup>-3</sup> )	RMSE (cm <sup>3</sup> cm <sup>-3</sup> )	R <sup>2</sup>	N
POR_calc	-0.007	0.038	0.789	1274
PTF03	0.000	0.031	0.862	1274

**RC3:**

L31, P10. Elaborate the range of Ks values used in training for PTF02, so reader can judge how low is RMSE of 0.94.

**AGREED.**

A: Thank you for giving this helpful viewpoint. We will added it in that sentence (P10 L11-12):  
*"In the case of KS prediction, the simplest best performing PTF – which was derived on a training dataset with KS ranging between -3.00 and 4.67 log<sub>10</sub>(cm day<sup>-1</sup>) – has an RMSE of 0.94 log<sub>10</sub>(cm day<sup>-1</sup>) ..."*

**RC3:**

L1-8, P10. You can compare the randomized RMSE by PTF02 (RMSE/(maxKs-minKs)) by some studies in the literature (preferably Europe or at least temperate soils)

**AGREED**

A: Thank you for this suggestion. We computed it for all the derived PTFs and highlighted this error measure in the case of KS and called it normalized RMSE (NRMSE). We also computed the NRMSE for

- the literature referred in the manuscript:
  - Zhang and Schaap (2017) (ROSETTA3): 0.11 log<sub>10</sub> (cm/day) (PSD+BD)
  - Lilly et al. (2018) (HYPRES) 0.18 log<sub>10</sub> (cm/day) (topsoil/subsoil distinction+USDA soil texture class+PSD+BD+OC),
  - Araya and Ghezzehei (2019) (USKSAT database) 0.06 log<sub>10</sub> (cm/day) (PSD+BD+OC),
- Nemes et al. (2005) (HYPRES) 0.15 log<sub>10</sub> (cm/day).

We added the information on computing NRMSE to P6 L15-19 in the manuscript as mentioned above, and the following:

P10 L12-13:

*"... has an RMSE of 0.94 log<sub>10</sub>(cm day<sup>-1</sup>) and NRMSE 0.14 log<sub>10</sub>(cm day<sup>-1</sup>) (Table S3)."*

P10 L17-22:

*"ROSETTA3 PTF with PSD and BD predictors had and RMSE of 0.68 log<sub>10</sub> (cm day<sup>-1</sup>) with an NRMSE of 0.11 log<sub>10</sub> (cm day<sup>-1</sup>) (Zhang and Schaap, 2017). Araya and Ghezzehei*

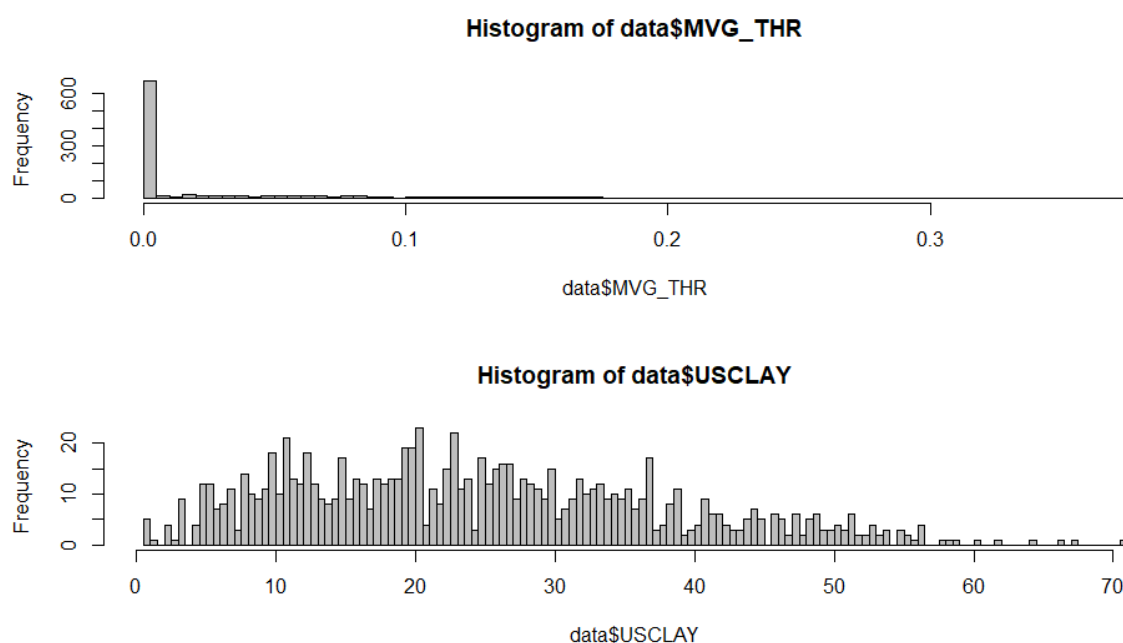
(2019) published PTF using PSD, BD and OC predictors with highest accuracy in the literature with an RMSE of  $0.34 \log_{10} (\text{cm day}^{-1})$  and NRMSE of  $0.06 \log_{10} (\text{cm day}^{-1})$ . In Lilly et al. (2008), the performance of the KS predictions and findings were similar to this study. They report an RMSE between 0.95 and  $1.08 \log_{10}(\text{cm day}^{-1})$  – with an NRMSE between 0.17 and  $0.20 \log_{10}(\text{cm day}^{-1})$  – for the KS prediction when analysed with several input combinations.”

### RC3:

L19, P10. I expect to see the high importance of clay in THETAr. It is not clear exactly how to estimate VG and MVG parameters.

### AGREE SOMEWHAT

A: It is right, expectation is not supported by the data, please see our answer above related to correlation plot: scatterplot of THR vs USCLAY. The reason for it can be that THETAr is a fitting parameter and for most of the samples it was close to 0. Please find here (Fig\_responses\_8) the histogram of THETAr and clay content based on all EU-HYDI samples that has measured chemical properties and fitted THETAr values:



Fig\_responses\_8

We also point out, that during the estimation of THR in the original model fitting of VG or MVG, THR is not only influenced by clay content, but also by pore connectivity, next to other soil structural properties. Importantly, THR is also influenced by the data range available during the fitting of the original data (Weber et al., 2020), which is a viable reason for the correlation between THR and USCLAY not to be as pronounced as one would expect.

Weber, T.K.D., Finkel, M., Conceição Gonçalves, M., Vereecken, H., Diamantopoulos, E., 2020. Pedotransfer function for the Brunswick soil hydraulic property model and comparison to the van Genuchten-Mualem model. Water Resour. Res. <https://doi.org/10.1029/2019WR026820>

Each VG and MVG parameters are predicted separately with random forest models.



**RC3:**

L23, P10. K0, matching point should be defined earlier.

**AGREED**

A: The following text was added to P4 L13-16:

*Similarly to euptfv1, for the description of the moisture retention curve (MRC), we predicted the VG model parameters: the residual water content ( $\vartheta_r$ ), the saturated water content ( $\vartheta_s$ ), and shape parameters  $\alpha$  and  $n$ . For the hydraulic conductivity curve, two additional parameters: the hydraulic conductivity acting as a matching point at saturation  $K_0$  and a shape parameter related to pore tortuosity ( $L$ ) are estimated too.*

**RC3:**

L25-30, P11. How many of K data are obtained from evaporation method, this method usually goes up to -1000 cm, is it why overestimation occurs in Fig S21 in drier conditions or another reason? Note that in this dry region K data is obviously small and mean error of about 0.8 is significant.

Moreover, comparing Fig s21 with Fig S1b (Toth et al., 2015), there less error in this dry region was observed.

**AGREED**

A: We deleted the sentence starting with “In parts, this is ...” (it was on P11 L32- P12 L2 in the discussion paper available from <https://gmd.copernicus.org/preprints/gmd-2020-36/gmd-2020-36.pdf>) and added the following text to P12 L19-27:

*“Samples with measurements of the HCC at pressure heads < -1000 cm are less frequent and are not as numerous within a dataset of a single sample, if it was measured. Since the dataset of estimated VG model parameters were identical in this study and in Tóth et al. (2015), differences between the two studies of the unsaturated HCC are related to the PTF methods involved. However, at pressure heads < -1000 cm, the HCC is dominated by non-capillary conductivity (Weber et al., 2019, Streck and Weber 2020), which is not included in the MVG model. The considerable data mismatch observable for the dry range (Fig. 6) can only be overcome by a different soil hydraulic property model and by a different PTF, because of compensatory effects in the VG. With this we mean that better data descriptions in the dry end, will lead to a larger mismatch in the wet end, as a consequence of the rigid model structure in the MVG model, which only accounts for capillary storage and conductivity. For better data description at < -1000 cm other more comprehensive models need to be adopted (Weber et al. 2020).”*

Streck, T., Weber, T.K.D., 2020. Analytical expressions for noncapillary soil water retention based on popular capillary retention models. *Vadose Zo. J.* 19, 1–5. <https://doi.org/10.1002/vzj2.20042>

Weber, T.K.D., Finkel, M., Conceição Gonçalves, M., Vereecken, H., Diamantopoulos, E., 2020. Pedotransfer function for the Brunswick soil hydraulic property model and comparison to the van Genuchten-Mualem model. *Water Resour. Res.* <https://doi.org/10.1029/2019WR026820>

**RC3:**

Fig2, 5. Explain the term “count” in legend

**AGREED**

A: The following was added to Figure 2 and Figure S1:

*“; Count: the number of cases in each rectangle.”*

Figures 5 and 6:

*“; Count: the number of cases in each hexagon.”*

**RC3:**

Table 7. RMSE is log10(cm/d) but this belongs to retention curve.

AGREED

A: Thank you for noting it, the unit was wrongly written in the title, we corrected it to  $\text{cm}^3 \text{cm}^{-3}$ .

**RC3:**

Table 8. this RMSE was computed only by K(h) data? Did you consider Lambda=0.5?

AGREED

A: Yes, the RMSE is based on the predicted and measured K(h) data. We did not set Lambda = 0.5, but fitted it for the dataset based on measured K(h) data. For the description of the hydraulic conductivity curve we predicted all of the following parameters:  $\theta_r$ : residual water content ( $\text{cm}^3 \text{cm}^{-3}$ ),  $\theta_s$ : saturated water content ( $\text{cm}^3 \text{cm}^{-3}$ ),  $\alpha$  ( $\text{cm}^{-1}$ ) and n (-): fitting parameters,  $K_0$ : the hydraulic conductivity acting as a matching point at saturation ( $\text{cm day}^{-1}$ ) and L: shape parameter related to pore tortuosity (-). Parameter m is provided based on  $m=1-1/n$  (van Genuchten, 1980). Thank you for highlighting it.

We added a paragraph entitled “Practical guidance on how to use the PTFs” on P13 L25 – P14 L19, in which we shortly summarize what parameters are predicted with euptfv2.

**RC3:**

L5, P 12. That’s interesting to show Comparison of point and parameter predictions, however, you should emphasize that this works only when water retention curve matters. Because one can use the n value of WRC and l=0.5 for K function.

AGREED

A: We will strengthen the description on why point and parameter predictions were compared. To overcome this confusion, we added to P6 L25-26:

*“The aim of this comparison was to analyse whether point or parametric prediction performs better when only THS and/or FC/FC\_2 and/or WP are needed.”*

and included the complementary information on P13 L5-6:

*“When moisture retention curve is not needed, but only THS and/or FC/FC\_2 and/or WP, we recommend to compute those with the point PTFs, more detailed explanation on it is included in Tóth et al. (2015).”*

**RC3:**

During some trials to run the package, I have faced with various errors such as

Error in source\_data

("https://github.com/TothSzaboBrigitta/euptfv2/blob/master/suggested\_PTFs/FC\_EUHYDI/FC\_PTF07.rdata?raw=True") : could not find function "source\_data"

please check the files again in the attached zip files. I could not also find neither euptf1nor 2 in CRAN repository.

AGREED

A: As mentioned above, the github repository includes the R scripts, which were used to develop the predictions and the derived pedotransfer functions. The dataset which we used for training and testing the algorithms cannot be shared according to the agreement between the data holders. We added the following to P15 L3-4 for clarification:



*“The training data set cannot be made publicly available due to legal restrictions of the EU-HYDI dataset, thus only a test sample is provided along with the model code.”*

euptfv1 is available from: <https://esdac.jrc.ec.europa.eu/themes/soil-hydraulic-properties> ,

[https://esdac.jrc.ec.europa.eu/public\\_path/shared\\_folder/themes/euptf.zip](https://esdac.jrc.ec.europa.eu/public_path/shared_folder/themes/euptf.zip) .

The PTFs of euptfv2 are available from the web interface which can be used without any coding skills. The R package is under construction. After finalizing the package it will be available from the European Soil Data Center site of the EC JRC (<https://esdac.jrc.ec.europa.eu/>). It will not be possible to have the package in the CRAN repository because it will have too large size for it – it will include several RF models.

## **II. LIST OF AUTHOR’S CHANGES IN MANUSCRIPT AND SUPPLEMENTARY MATERIAL**

Page and line numbering refer to that of the revised manuscript with track changes included under III. MARKED-UP MANUSCRIPT AND SUPPLEMENTARY MATERIAL VERSION section of this document.

**The following changes have been made in the manuscript:**

- P1 L17: matric potential is corrected to 15,000 cm,
- P1 L25: sentence is added on practical guidance,
- P1 L30 – P2 L1: sentence is rephrased,
- P2 L17-19: more information is added related to uncertainty assessment,
- P2 L21: matric potential is corrected to 15,000 cm,
- P3 L2-3: further information is added about the estimation of uncertainty,
- P4 L4: matric potential is corrected to 15,000 cm,
- P4 L13-16: description of model parameters is added,
- P6 L1-2: information on relative importance is added,
- P6 L15-19: computation of normalized root mean square error (NRMSE) is added,
- P6 L25-26: clarifying sentence – related to point and parametric prediction – is added,
- P7 L15: RMSE of KS is rounded to two digits,
- P7 L16-18: sentence on table of NRMSE is added,
- P7 L30: format of citation is corrected,
- P10 L11-13: range of KS in the training dataset and NRMSE is added,
- P10 L16: RMSE is adjusted,
- P10 L17-21: NRMSE of other published PTFs is added,
- P10 L21: RMSE is rechecked and corrected,
- P11 L11: “;” is changed to “,”,
- P11 L31: matric potential is corrected to 15,000 cm,
- P12 L1: matric potential is corrected to 15,000 cm,
- P12 L16-18: the sentence starting with “In parts, this is ...” is deleted,
- P12 L19-L27: information is added to interpret the results,
- P12 L29: language correction,
- P12 L30 – P13 L2 and P13 L4-6: information and discussion is added,
- P13 L9-13 and L14-21: discussion is added,
- P13 L14: language correction – were instead of was,
- P13 L25 – P14 L19: section on “Practical guidance on how to use the PTFs” is added,

- P14 L25: matric potential is corrected to 15,000 cm,
- P15 L3-4: information about the training dataset is added,
- P16 L25-27: reference is added,
- P18 L31-32: reference is added,
- P19 L8-10: reference is added,
- P19 L31-32: reference is added,
- P29: “a” is formatted to superscript in the 3<sup>rd</sup> row of WP,
- P31: Table 11 is added, which lists the recommended PTFs,
- P33-34: Figure 2 is reedited and replaced –  $R^2$  is added,
- P34 L7: description of count is added,
- P35: Figure 3 is reedited and replaced – relative importance is shown,
- P35 L3: figure’s caption is corrected,
- P37: Figure 5 is replaced –  $R^2$  is added,
- P37 L6-7: description of count is added,
- P38: Figure 6 is replaced –  $R^2$  is added,
- P38 L6-7: description of count is added,
- P39: Figure 7 is reedited and replaced – relative importance is shown,
- P39 L3-5: figure’s caption is corrected.

**The following changes have been made in the short summary:**

- complementing the following sentence with the text in blue: “The influence of predictor variables on predicted soil hydraulic properties is explored and practical guidance on how to use the derived PTFs is provided.”

**The following changes have been made in the supplementary material:**

- P1-2: Figure S1: figure is reedited and replaced –  $R^2$  is added, meaning of count is added in the caption (L7).
- P5: Table S3 on NRMSE values added,
- P6-17, 19: Figures S2, S4, S6, S8, S10, S12, S14, S16, S19 are reedited and replaced and “Squared error (SE)” is changed to “Root mean square error (RMSE)” in the figures’ captions,
- P21-22: Table S3 is moved to the manuscript.

### **III. MARKED-UP MANUSCRIPT AND SUPPLEMENTARY MATERIAL VERSION**

Please find revised marked-up manuscript and supplementary material on the following pages.

# Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (eupthv2)

Brigitta Szabó<sup>1</sup>, Melanie Weynants<sup>2</sup>, Tobias KD Weber<sup>3</sup>

<sup>1</sup>Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Herman Ottó út 15, 1022 Budapest, Hungary. ORCID: 0000-0003-1485-8908

<sup>2</sup>European Commission Joint Research Centre, via Enrico Fermi 2749, 21027 Ispra, Italy. ORCID: 0000-0002-1447-0105

<sup>3</sup>Institute of Soil Science and Land Evaluation, University of Hohenheim, Emil-Wolff-Straße 27, 70593 Stuttgart, Germany. ORCID: 0000-0002-3448-5208

Correspondence to: Brigitta Szabó ([toth.brigitta@agrar.mta.hu](mailto:toth.brigitta@agrar.mta.hu))

**Abstract.** Soil hydraulic properties are often derived indirectly, i.e. computed from easily available soil properties with pedotransfer functions (PTFs), when those are needed for catchment, regional or continental scale applications. When predicted soil hydraulic parameters are used for the modelling of the state and flux of water in soils, uncertainty of the computed values can provide more detailed information when drawing conclusions. The aim of this study was to update the previously published European PTFs (Tóth et al., 2015, eupthv1.4.0) by providing prediction uncertainty calculation built into the transfer functions. The new set of algorithms was derived for point predictions of soil water content at saturation (0 cm matric potential head), field capacity (both -100 and -330 cm matric potential head), wilting point (-15,000 cm matric potential head), plant available water, and saturated hydraulic conductivity, as well as the Mualem-van Genuchten model parameters of the moisture retention and hydraulic conductivity curve. The minimum set of input properties for the prediction is soil depth and sand, silt and clay content. The effect of including additional information like soil organic carbon content, bulk density, calcium carbonate content, pH and cation exchange capacity were extensively analysed. The PTFs were derived adopting the random forest method. The advantage of the new PTFs is that they i) provide information about prediction uncertainty, ii) are significantly more accurate than the eupthv1, iii) can be applied for more predictor variable combinations than the eupthv1, 32 instead of 5, and iv) are now also derived for the prediction of water content at -100 cm matric potential head and plant available water content. [A practical guidance on how to use the derived PTFs is provided.](#)

## 1 Introduction

Quantitative information on state and flux of water in the critical zone is important for a wide range of environmental process models and decision support systems related to land surface processes (Lin, 2010; Zhao et al., 2018). Performance of hydrologic, climate, crop and other models related to soil hydrological processes depends on the quality and resolution of soil hydraulic input parameters (Vereecken et al., 2015). Simulations of [flow through](#) variably saturated ~~moisture-flux~~[soil media](#)

~~in the vadose zone~~ either rely on simple modelling approaches which only require few directly measureable input variables such as porosity, field capacity, and wilting point, or on the Richards equation. While the former are simple and straightforward to obtain, the Richards equation requires knowledge about the soil hydraulic properties over the full moisture range. In practice, one of the most common approaches to describe the water retention and hydraulic conductivity curves required to solve the Richards equation is arguably (Weber et al., 2019) the Mualem-van Genuchten model (MVG) (van Genuchten, 1980; Mualem, 1976). Since soil hydraulic measurements in the laboratory or in the field are often time consuming, expensive and difficult, indirect methods for estimating soil hydraulic properties using widely available surrogate data have been developed (Schaap, 2006). To date, a large number of pedotransfer functions have become popular to predict soil hydraulic properties and MVG model parameters (Van Looy et al., 2017).

Information on the uncertainty of the predicted soil hydraulic properties is important for modelling the state and flux of water in soil. The source of prediction uncertainty can be threefold: it can stem from the i) predictor (e.g. measurement uncertainty, non-representativeness of a sample), ii) predicted variables (e.g. uncertainty in the estimated soil hydraulic model parameters), and the iii) algorithm which describes the relation between the two. Information on the uncertainty of the predictor variables is commonly not available in PTFs derived before the 2000s, but has become a more intensively studied topic in the last decade.

For example, Weynants *et al.* (2009) quantified uncertainty of derived PTFs related to experimental, model and fitting errors with the one-step inversion method. Deng et al. (2009) differentiated and quantified intrinsic and input uncertainty of PTFs. ~~Tranter et al.; (2010) developed an uncertainty estimation method using fuzzy k-means with extragrades classification that can be applied in any PTF prediction.~~ ~~Kotlar et al.; (2019) presented uncertainty assessment of PTFs through deriving PTFs on tens of resamples for train and test sets.~~ Román Dobarco *et al.* (2019) introduced prediction interval coverage probability

to assess prediction uncertainty in PTFs derived on French soils. McNeill et al. (2018) provided estimation of the distribution and confidence intervals of the predicted soil hydraulic property (i.e. water content at 100 cm and 15,000 cm matric potential head and total available water). In the field of soil mapping it is an even more extensively studied topic where different computational methods have been proposed to assess uncertainty of the mapped properties. Examples are estimation of the 90% prediction intervals based on a triangular distribution (Odgers *et al.* 2014), quantification of mapped soil properties uncertainties by quantile regression forest (Vaysse and Lagacherie, 2017), and a detailed comparison of uncertainties in mapped soil organic carbon content by different geostatistical and machine learning methods (Sztatmári and Pásztor, 2019).

Machine learning methods can be more robust to construct PTFs in comparison to previous approaches such as linear regression or simple decision trees if relationship between the predictors and response is highly non-linear (Araya and Ghezzehei, 2019). The random forest algorithm (Breiman, 2001) is able to outperform other machine learning methods (Olson et al., 2018), which was also shown for predicting soil properties (Hengl et al., 2018; Nussbaum et al., 2018). Improvements in computing power, statistical methods and statistical software provide the possibility to apply more easily even complex models on large datasets. Therefore, complexity of a prediction algorithm is no longer a barrier in selecting a suitable algorithm to develop and apply PTFs. Most of the recent machine learning algorithms have the built in possibility to compute the uncertainty in the predicted variable, e.g. by quantile regression forest (Meinshausen, 2006) or generalized boosted regression

(Ridgeway, 2017). If PTFs are derived with these algorithms, the uncertainty of the predicted soil property can be directly estimated when applying the PTF (Szabó et al., 2019a), although this could also be achieved by applying the above mentioned uncertainty assessment methods without using machine learning methods (e.g. (Kotlar et al., 2019; Tranter et al., 2010).

Despite the above mentioned developments, the euptfv1 (Tóth et al., 2015) and derived soil hydraulic property maps for Europe on a 1km and 250m grid (Tóth et al., 2017) do not include uncertainties in the prediction. Hence, the aim of our study was to update the euptfv1 by deriving a new set of soil hydraulic PTFs (euptfv2) providing uncertainty calculation built into the PTF model. For this, we rely heavily on the datasets used in the construction of the euptfv1. Methodologically, we constructed new soil hydraulic PTFs on the basis of the random forest method which facilitates a quantification of prediction uncertainties. The predicted variables of interest included soil water content at saturation, field capacity and wilting point, plant available water content, saturated hydraulic conductivity, MVG parameters of the moisture retention and hydraulic conductivity curves. The predictions are based on easily available soil properties. The predictor variables were similar to those of euptfv1, except the topsoil and subsoil distinction, which was replaced by mean soil depth of the sample, since it is typically known, anyway. Additionally, the improved performance of the euptfv2 was assessed against predictions using the earlier version. Moreover, we determined the minimum sufficient predictor variables for 32 input variables combinations.

## 2 Materials and Methods

The construction of a pedotransfer function requires three elements: predictor variables, predicted variables as the property of interest, and a transfer method between the former two. The predicted variables are in this case directly measured soil hydraulic properties on samples contained in a large pan-European dataset, ensuring a representativeness of the PTF for Europe. Additionally, Tóth et al. (2015) had fitted MVG model parameters for each sample dataset individually by inverse modelling, which we reused in this study.

### 2.1 Dataset

The European Hydropedological Data Inventory (EU-HYDI) (Weynants et al., 2013) provided the basis for the preparation of the prediction algorithms. The dataset partitions for training and testing the prediction algorithms were almost identical to the ones used in Tóth et al. (2015), except that the samples had to have information on soil depth as well. Depending on the soil hydraulic property of interest, 76-99% of the originally selected samples were used to derive the new PTFs. It enabled comparison of the performance between the EU-PTFs (Tóth et al., 2015) – built in the euptfv1 (Weynants and Tóth, 2014) – and their improved version (euptfv2). Table 1 shows the number of samples in the training and test sets.

### 2.2 Predicted soil hydraulic properties

Prediction algorithms were derived for each of the following soil hydraulic properties:

- water content at saturation (THS): water content at 0 cm matric potential head;

- water content at field capacity at
  - -100 cm matric potential head (FC<sub>2</sub>), and
  - -330 cm matric potential head and (FC);
- water content at wilting point (WP): water content at -15,000 cm matric potential head;
- plant available water content (AWC) based on the following equations:
  - $AWC = FC - WP$  (1)
  - $AWC_2 = FC_2 - WP$  (2)
- saturated hydraulic conductivity (KS): hydraulic conductivity at 0 cm matric potential head;
- Mualem-van Genuchten model parameters (VG; for the water retention model only, MVG; for the water retention and hydraulic conductivity model).

Transformation of predicted variables, and explanation on how (i) the water content at a certain matric potential head values were harmonized and (ii) the Mualem-van Genuchten model parameters were fitted is provided in great detail in (Tóth et al., 2015). Similarly to euptfv1, for the description of the moisture retention curve (MRC), we predicted the VG model parameters: the residual water content ( $\theta_r$ ), the saturated water content ( $\theta_s$ ), and shape parameters  $\alpha$  and  $n$ . For the hydraulic conductivity curve, two additional parameters: the hydraulic conductivity acting as a matching point at saturation  $K_0$  and a shape parameter related to pore tortuosity ( $L$ ) are estimated too. FC<sub>2</sub> was not predicted in euptfv1 and was determined in this study as follows. In the EU-HYDI, 8231 samples have at least one water content observation in the matric potential head range -110 to -95 cm. 86% of those have a measured water retention value exactly at -100 cm matric potential head. In 10% of the cases, FC<sub>2</sub> was set to the water content measured at the closest matric potential head in the range [-110, -95]. In the absence of a measured value at -100 cm, in 4 % of the cases, FC<sub>2</sub> was computed by linear interpolation between the two closest matric potential heads smaller and greater than -100 cm. In the case of AWC and AWC<sub>2</sub> direct and indirect predictions were analysed, i.e. AWC was once predicted directly from the predictor variables and once computed from the PTF predicted variables WP, and FC and FC<sub>2</sub>, respectively.

### 2.3 Predictor variables

As predictors we used the following easily available soil properties: the particle size densities (PSD) characterised by the mass-percentages of clay (<2  $\mu\text{m}$ ), silt (2–50  $\mu\text{m}$ ) and sand (50–2000  $\mu\text{m}$ ), organic carbon content (OC; mass-%), bulk density (DB;  $\text{g cm}^{-3}$ ), calcium carbonate content (CACO<sub>3</sub>; mass-%), pH in water (PH<sub>H2O</sub>; -), cation exchange capacity (CEC;  $\text{cmol (+)}$   $\text{kg}^{-1}$ ), and replaced the former topsoil and subsoil distinction in euptfv1 with mean soil depth (cm) (DEPTH). At minimum, the predictor variables, clay, silt and sand content, as well as mean soil depth were used regardless of predicted variable. In addition to that, we tested every possible combination of the other above mentioned soil properties (predictor variables) to determine which combination significantly improves the performance of the predictions. A total of 32 different combinations of predictor variables were studied in their respective ability to predict the nine different properties of interest; i.e. the set of soil hydraulic properties and model parameters.

Replacing the topsoil/subsoil distinction with depth for the new PTFs was supported by the fact that this information is commonly available, too, or can be based on expert knowledge. Introducing more accurate information on depth might improve the performance without using machine learning algorithms for the prediction. However, we did not test this hypothesis, because our aim was to provide uncertainty of the predictions related to predictor variables of the PTFs. Tested predictor variables are shown in Table 1 with number of samples used to derive the PTFs and compute their performance.

#### 2.4 The Random Forest algorithm to derive PTFs

We derived the PTFs adopting the random forest method (Breiman, 2001), implemented in the ‘ranger’ R package (Wright and Ziegler, 2017). We selected this method, because (i) it is among the best performing prediction algorithms if there is a complex interaction structure in the dataset (Boulesteix et al., 2012), (ii) it computes quantiles of the predicted values, (iii) parallel processing is supported which saves significant computation time, and (iv) the initially black-box type algorithm can be interpreted based on computing variable importance and analysing partial dependence plots implemented in the ‘pdp’ R package (Greenwell, 2017b).

In the case of a continuous response variable, a random forest is an ensemble of de-correlated regression trees (Breiman, 2001). The regression tree approach divides the predictor space into non-overlapping regions through minimizing the residual sum of squares. The aim of the method is to subset the data as homogeneously as possible at each split. The observations can be assigned to the defined regions in which the mean of the response variable is the predicted value. Single trees of the forest are noisy and limited in performance, but if many unbiased trees are derived and averaged with bagging, the variance is reduced and performance of the prediction improves (Hastie et al., 2009). Building of de-correlated trees is achieved by randomization at two levels. Firstly, each tree of the forest is grown on a randomly selected two thirds of the data with replacement, which is called bootstrap sample or in-bag fraction. Secondly, at each node of a single tree, randomly selected sets of predictors are analysed to split the data. This feature of randomization allows correlation between the response variables (Ziegler and König, 2014), which is an important advantage in the case of pedotransfer functions where predictors are often highly correlated.

Parameter tuning of the ranger was performed with the ‘caret’ R package (Kuhn et al., 2017, 2018). With the implemented train function, a fivefold cross-validation was repeated ten times to tune the number of randomly selected predictor variables at each split (*mtry*) and find the best performing splitting rule (*splitrule*) during training. We started the tuning by setting the number of randomly selected predictor variables to two, then added one by one until the number of all available predictors for each input variable combination was reached. All three built-in splitting rules in the ranger function were tuned, namely *variance*, *extratrees* and *maxstat*. The minimum node size was kept to 10. In addition to the tuning options included in the train function of the caret package, we optimized the number of trees in the forest. The above described tuning was performed by discretely altering the number of trees in the forest in separate tuning steps to 50, 100, 200, 500 and 1000, analysing the results and choosing the best number of trees for the random forest.

We analysed the relevance of predictors and their influence on the response variable. The relevance of predictors was determined by computing the variable importance based on the mean decrease in impurity (Hastie et al., 2009) in the ranger

function. The relative importance was assessed by dividing the variable importance of each predictor by the sum of the importance of all the predictors after (Kotlar et al., (2019)). The marginal effect of some selected predictors on the response – soil hydraulic parameters – was analysed with partial dependence plots (Greenwell, 2017a, 2017b).

The final prediction algorithm was built on the whole training set based on the result of the tuning. For the description of the uncertainty, quantile regression was performed. Quantiles of the predicted values were estimated as implemented in quantile regression forest (Meinshausen, 2006). We analysed the 90% prediction interval for all predictions, but the derived algorithms (PTFs) provide the possibility to compute the individual predictions of each tree.

## 2.5 Evaluation of derived PTFs

The performance of the PTFs was calculated using the median values predicted by the random forests. It was described with the root mean square error (RMSE) (Eq. 3.), and the coefficient of determination ( $R^2$ ) (Eq. 4.) computed for the training and test sets.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} = \sqrt{MSE} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (4)$$

where  $y_i$  is the measured and  $\hat{y}_i$  the predicted soil water content or log-transformed saturated or unsaturated hydraulic conductivity,  $\bar{y}$  is the average of  $y_i$ ,  $N$  is the number of  $y_i$  and  $\hat{y}_i$  data pairs, and MSE is the mean square error. The different data range of the dataset influences the performance of the PTFs when that is compared to the studies in the literature. Therefore, normalized RMSE (NRMSE) was computed (Eq. 5.), where  $y_{max}$  and  $y_{min}$  are the maximum and minimum value of variable.

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \quad (5)$$

For each predicted variable, there was an initial set of 32 predictor combinations (Table 1), whose individual performance for each of the predicted variables was assessed. Based on the test results, we derived recommendations which PTF should be used when certain sets of predictor variables are available. We compared the performance of PTFs to quantify if there are significant differences between the predictions as a consequence of adding certain soil properties to the predictor variables. We also compared the performance of point and parameter estimations for those input combinations, which reflect the most frequently available soil property combination from a practical point of view. The aim of this comparison was to analyse whether point or parametric prediction performs better when only THS and/or FC/FC<sub>2</sub> and/or WP are needed.

Additionally, the performance of the presented random forest based PTFs was compared to that of the euptfv1 (Tóth et al., 2015). For comparison, those PTFs from euptfv2 were selected which corresponded to the analysed input variable combination of the euptfv1.



The comparison of PTFs was done using a non-parametric Kruskal-Wallis test at the 5% significance level applied on the MSE values – computed on TEST\_BASIC and/or TEST\_CHEM+ sets (Table 1) – using the R package agricolae (De Mendiburu, 2017). Recommendation of PTFs for a given set of predictor variables was based on the performance of euptfv2 on the test sets. If there was no significant difference in performance, the PTF derived from the largest population was selected.

5 All statistical analysis was performed in R [version 3.6.0] (R Core Team, 2019).

### 3 Results and discussion

#### 3.1 General performance

In the process of tuning the random forest parameters, the number of trees was found to be sufficient when set to 200 in all cases. The number of candidate predictors was found to be higher than the recommended square root of the number of available  
10 predictor variables ( $p$ ) in most of the cases, especially when  $p$  was greater than 5 (Fig.1). When optimizing the splitting rules to build the trees in the forest, overall, the best performance was achieved by the *extratrees* rule in 54 %, by the *variance* rule in 28%, and by the *maxstat* rule in 18% of the cases (Fig. 1).

The RMSE values were between 0.020 and 0.068  $\text{cm}^3 \text{cm}^{-3}$  for THS (Table 2), 0.046 and 0.055  $\text{cm}^3 \text{cm}^{-3}$  for FC (Table 3), 0.040 and 0.060  $\text{cm}^3 \text{cm}^{-3}$  for FC\_2 (Table 4), 0.037 and 0.048  $\text{cm}^3 \text{cm}^{-3}$  for WP (Table 5), 0.043 and 0.053  $\text{cm}^3 \text{cm}^{-3}$  for AWC  
15 (Table S1), 0.045 and 0.060  $\text{cm}^3 \text{cm}^{-3}$  for AWC\_2 (Table S2), and 0.989-09 and 1.18  $\log_{10} (\text{cm day}^{-1})$  for KS (Table 6) in the case of including different predictor variables computed on the test sets. [Table S3 shows the NRMSE for the point predictions computed for the TEST\\_BASIC and TEST\\_CHEM+ sets to provide possibility for comparison with other PTFs available from the literature.](#) In the case of VG and MVG, RMSE for the entire matric potential head range was between 0.041 and 0.068  $\text{cm}^3 \text{cm}^{-3}$  for the moisture retention (Table 7) and 0.61 and 0.71  $\log_{10} (\text{cm day}^{-1})$  for the hydraulic conductivity (Table 8). These  
20 RMSE values are within the range of recently published PTFs (McNeill et al., 2018; Nguyen et al., 2017; Román Dobarco et al., 2019; Zhang and Schaap, 2017). In the case of the point estimations, Figures 2, S1 depict the scatterplots of measured and predicted soil hydraulic parameters with 90% prediction interval. Performance of the worst to best PTFs are shown. The addition of predictors that significantly improve the predictions also decreases the uncertainty. Figures S2, S4, S6, S8, S10, S12, S14, S16, S19 show the squared error of the derived PTFs computed on the TEST\_BASIC and TEST\_CHEM+ sets. The  
25 PTFs are ordered based on their performance. Density plots of measured and predicted soil hydraulic values are included in Figures S3, S5, S7, S9, S11, S13, S15, S17, S20. Plots show the PTFs that use the most frequently available predictors.

This study strengthens the importance of chemical soil properties in the prediction. CEC was found to be an important predictor by Pachepsky and Rawls (1999) for FC and WP, by Botula et al. (2013) for water retention at several matric potential head values, and by Hodnett and Tomasella (2002) for the VG parameters. Hodnett and Tomasella (2002) showed that pH influenced  
30 all four VG parameters. The role of  $\text{CaCO}_3$  was shown to be not significant in the study of (Khodaverdiloo et al., 2011). They highlight that a possible influence of  $\text{CaCO}_3$  might already have been indirectly included by bulk density. The role of PSD, BD and OC has been studied extensively by various authors, e.g. Nemes et al. (2003); Rawls et al. (2003); Vereecken et

al. (1989); Weynants et al. (2009); Wösten et al. (1999), which is in line with the general pattern of variable influence we see in this study.

Table S3 summarizes the recommended PTF for each combination of available predictor variables. The importance and influence of soil properties on the performance of hydraulic PTFs and results of partial dependence plots are reported below by predicted soil hydraulic properties.

### 3.2 Point estimations

The performance of the PTFs was computed for the training and test sets (Tables 2-8 and Tables S1-2) indicating the presence of significant differences. For each predictor variable, the recommended PTF number is indicated and its predictor variables are highlighted in bold font in the respective tables. For easier comparison with euptfv1, the corresponding PTF number used in Tóth et al. (2015) is additionally provided in each table. In the following, detailed results of the constructed PTFs for the individual predicted variables are presented and discussed.

#### Water content at saturation

Table 2, Figures S2 and S3 show the performance of the PTFs predicting THS. The best performing random forest is PTF03. It is also the one trained on the largest population. It uses PSD, DEPTH and BD as predictors. For the prediction of THS, the most important variable by far is BD (Fig. 3). When BD is not used for the computation of THS, values above  $0.60 \text{ cm}^3 \text{ cm}^{-3}$  are not well predicted (Fig. S3). The addition of OC or CACO3 or PH\_H2O to PSD and DEPTH improves significantly the performance of the PTF. The picture changes if BD is known: if PSD, DEPTH and BD were available, further addition of OC or CACO3 or PH\_H2O or CEC does not significantly improve the prediction, neither do their combinations. Figure 4 shows the dependence of THS on OC and BD, considering the average effect of the other predictor variables – i.e. PSD and DEPTH. When BD is lower than  $1.5 \text{ g cm}^{-3}$  changes in OC does not influence THS. If BD is larger than  $1.5 \text{ g cm}^{-3}$ , samples with higher OC have higher THS.

#### Water content at field capacity

The performance of the PTFs computed on training and test set are shown in Table 3, Figures S4 and S5 for FC\_2 and in Table 4, Figures S6 and S7 for FC. The best performing PTF derived from the largest population is the one using i) PSD, DEPTH, OC, BD and PH\_H2O (PTF18) in the case of FC\_2, and ii) PSD, DEPTH, OC and BD (PTF07) for FC.

For FC\_2, the two most important variables are USSAND and BD (Fig. 3). When BD and USSAND increase, FC\_2 decreases (Fig. 4). Adding OC or BD to PSD and DEPTH significantly improves the prediction of FC\_2. If either of CACO3, PH\_H2O or CEC is added as a further predictor to PSD and DEPTH, the performance of the PTF does not significantly improve. If PSD, DEPTH and BD are available, adding OC or CACO3 or PH\_H2O does not significantly improve the prediction. Including CEC as an additional predictor besides PSD, DEPTH and BD, significantly improves the estimation of FC\_2.

USSAND and USCLAY are the two most important variables for the prediction of FC (Fig. 3). Instead of analysing these two soil properties, both characterizing the soil texture, we include OC next to USSAND in the partial dependence plot analysis, because the amount of OC can be altered due to change in climate, land use, soil and water management, cropping systems, etc. (Wiesmeier et al., 2019). Within the range of OC in the dataset FC increases with increasing OC regardless of USSAND content by up to  $0.08 \text{ cm}^3 \text{ cm}^{-3}$  even when USSAND is greater than 60 % (Fig. 4). Adding OC or CEC to PSD and DEPTH significantly improves prediction of FC. The effect of CEC on the prediction of FC was also shown by Pachepsky and Rawls (1999). BD or CaCO<sub>3</sub> or PH<sub>2</sub>O do not significantly improve the predictions if PSD, DEPTH, or PSD, DEPTH and OC are available. Predictions significantly improve when both CaCO<sub>3</sub> and PH<sub>2</sub>O are added as predictors to PSD, DEPTH and OC.

#### Water content at wilting point

The performance of PTFs derived for WP prediction is shown in Table 5, Figures S8 and S9. Among the best performing PTFs, PTF09 is derived on the largest training set. It uses PSD, DEPTH, OC and PH<sub>2</sub>O as predictors. Even though the most important variables for WP prediction were USCLAY and USSAND (Fig. 3), we included OC on the partial dependence plot (Fig. 4) as in the FC analysis. USCLAY had the strongest influence on WP. The influence of OC on WP can be detected for soils with OC less than 4 % and USCLAY less than 50 %. Below 10 % USCLAY, the WP slightly increases with increasing OC. When USCLAY is between 10 and 50 % and OC is less than 4%, increasing OC generally decreases WP. OC significantly improves the prediction of WP if added to PSD and DEPTH. If BD or CaCO<sub>3</sub> or PH<sub>2</sub>O or CEC are added to PSD and DEPTH, the performance of the prediction does not improve significantly. Adding CaCO<sub>3</sub> and CEC to PSD, DEPTH and OC significantly improves the prediction.

#### Plant available water content

Table S1, S2 and Figures S1, S10-13 show the performance of AWC and AWC<sub>2</sub> predictions. PTF03 is the best performing algorithm with largest training set for both. It considers PSD, DEPTH and BD for the prediction. For both AWC and AWC<sub>2</sub>, BD is the most important predictor among the analysed variables (Fig. 3). The second most important variable is USCLAY in the case of AWC<sub>2</sub> and USSILT for AWC. Increasing BD and USCLAY decreases AWC<sub>2</sub>. In the case of AWC, increasing BD and decreasing USSILT decreases the water content (Fig. 4). OC and BD significantly improve the prediction of AWC<sub>2</sub> when added as input variables next to PSD and DEPTH. If either BD or OC is already included, adding the respective other, does not significantly improve the prediction. Neither PH<sub>2</sub>O, CaCO<sub>3</sub> nor CEC significantly improve the prediction. For the prediction of AWC, further addition of only BD or OC or CaCO<sub>3</sub> or PH<sub>2</sub>O or CEC to PSD and DEPTH does not significantly improve the prediction. If both OC and BD are included as predictors next to PSD and DEPTH, the prediction significantly improves.

There is no significant difference between direct and indirect predictions, neither for AWC nor for AWC<sub>2</sub>. However, the size of the test set used for the statistical analysis is limited. There were only 145 samples in the TEST\_BASIC set and 64 samples

in TEST\_CHEM+ set after merging datasets available for both direct and indirect predictions for analysing AWC, and 70 and 34 samples in the case of AWC\_2. Thus, if prediction of FC\_2/FC and WP are needed in addition to AWC\_2/AWC, we recommend to compute AWC from those to save on computing time. Variation in AWC could be explained less efficiently (Table S1, S2) than the other studied water retention values but the performance of the prediction is comparable with that of published in the literature (Li et al., 2016; Malone et al., 2009).

### Saturated hydraulic conductivity

The performance of KS prediction is shown in Table 6, Figure S14 and S15. The predictors of the best performing PTF derived on largest training set are PSD, DEPTH and OC (PTF02). The prediction of KS significantly improves if OC is included among the predictor variables next to PSD and DEPTH. No other predictors significantly improve the performance of the PTF. On the training dataset, when OC is greater than 2.5 %, the influence of clay content on KS is more dominant than that of OC (Fig. 4). In the case of KS prediction, the simplest best performing PTF – which was derived on a training dataset with KS ranging between  $-3.00$  and  $4.67 \log_{10}(\text{cm day}^{-1})$  – has an RMSE of  $0.94 \log_{10}(\text{cm day}^{-1})$  and NRMSE  $0.14 \log_{10}(\text{cm day}^{-1})$  (Table S3). PSD and CEC are the most important input variables for the prediction of KS when all nine variables are considered as predictors (Fig. 3). In that case, OC is the fifth and BD is only the eighth most important variable. The prediction performance is influenced by the heterogeneity of measurement methods of KS in the EU-HYDI dataset. When the methods are homogeneous, the RMSE value is usually around  $0.6$ – $0.87 \log_{10}(\text{cm day}^{-1})$  as reviewed by (Zhang and Schaap, 2019). ROSETTA3 PTF with PSD and BD predictors had an RMSE of  $0.68 \log_{10}(\text{cm day}^{-1})$  with an NRMSE of  $0.11 \log_{10}(\text{cm day}^{-1})$  (Zhang and Schaap, 2017). Araya and Ghezzehei (2019) report that the published PTF using PSD, BD and OC predictors with the highest accuracy in the literature has with an RMSE of  $0.3$ – $0.4 \log_{10}(\text{cm day}^{-1})$  and NRMSE of  $0.06 \log_{10}(\text{cm day}^{-1})$ . In Lilly et al. (2008), the performance of the KS predictions and findings were similar to this study. They report an RMSE between  $0.95$  and  $1.098 \log_{10}(\text{cm day}^{-1})$  – with an NRMSE between  $0.17$  and  $0.20 \log_{10}(\text{cm day}^{-1})$  – for the KS prediction when analysed with several input combinations. Even when information on soil structure and crack orientation was considered – next to topsoil and subsoil distinction, PSD, BD and OC – the RMSE was  $0.97 \log_{10}(\text{cm day}^{-1})$ . BD would be among the most important variables, but also in their analysis its influence was masked out. They derived the PTFs on the HYPRES dataset (Wösten et al., 1999), which also includes very diverse methods to determine the saturated hydraulic conductivity and part of which is also contained in the EU-HYDI. The uncertainty in the predictions (Fig. 2) could be decreased if the predictions would be differentiated according to the measurement methods, but that might decrease the applicability of the PTFs. On the contrary, this study indicates the necessity to include saturated hydraulic conductivity values determined from many different measurement techniques, otherwise the PTFs are expected to lose their generality.

### 3.3 Parameter estimations

The performance of parametric PTFs are shown in Tables 7 and 8 and Figures 5, 6, S16–S21. Figure 7 illustrates the importance of variables for the prediction of VG and MVG parameters. The best performing PTF derived on the largest training set is

PTF29 – with PSD, DEPTH, OC, BD, PH<sub>2</sub>O and CEC – for MRC and PTF27 – with PSD, DEPTH, OC, BD, CACO<sub>3</sub>, PH<sub>2</sub>O – for HCC.

For  $\theta_r$ , overall, BD is the most important predictor while all other predictors show similar variable importance (Fig. 7). Interpretation of this parameter is complex, but it was demonstrated that it is influenced by the soil specific surface area (Assouline and Or, 2013), and the measured data range (Weber et al. 2019). For  $\theta_r$ , the most important predictor is by far BD, similarly to THS. The importance of CEC has to be noted for the prediction of parameters  $\alpha$ ,  $n$  and  $L$ . For prediction of parameter  $n$  – which relates to the pore size distribution – USCLAY and USSAND are the most important variables.  $K_0$  is influenced by several soil properties besides those included in the dataset used here, e.g. pore connectivity, tortuosity, primary pore orientation, some of which are not direction. These properties cannot be directly inferred from other soil properties limiting the explanatory power of the available properties. The prediction of  $K_0$  remains complex and challenging. Variable importance of all studied predictors is greater than 70%. Moreover,  $K_0$  is influenced by the data quality, and, moreover, is correlated in parameter space, which is not treated, here.

Only a few studies have analysed the importance of CEC for MRC and HCC PTFs (Botula et al., 2013; Hodnett and Tomasella, 2002; Pachepsky and Rawls, 1999) which might be linked to the fact that CEC is rarely available in soil hydraulic datasets. It is noteworthy to highlight that all best performing MRC PTFs (PTF24, PTF28, PTF29, PTF30, PTF31) include CEC among the predictors (Table 7). In addition to that, Hodnett and Tomasella (2002) found that CEC was important for the prediction of  $\theta_r$  and  $\alpha$  parameters of the van Genuchten model. This is because CEC provides indirect information on soil mineralogy and reflects soil specific surface area, charge density and pore size which influence soil water retention (Lal and Shukla, 2004).

#### Moisture retention curve

If BD or OC or CACO<sub>3</sub> or CEC or PH<sub>2</sub>O are added as a predictor to information on PSD and DEPTH, the performance of the PTF significantly improves (Table 7., Fig. S16). Adding BD next to PSD and DEPTH improves the predictions more than adding OC (Table 7., Fig. S17). BD and OC together significantly improve the prediction compared to using PSD, DEPTH together with either BD or OC. Adding OC next to PSD, DEPTH, BD and chemical soil properties (CACO<sub>3</sub> and/or CEC and /or PH<sub>2</sub>O) does not significantly improve the prediction. If PSD, DEPTH, CACO<sub>3</sub> and CEC are available, further addition of PH<sub>2</sub>O does not improve the prediction. The best performing PTF includes USSAND, USSILT, USCLAY, DEPTH, BD, CACO<sub>3</sub>, CEC. Figure 5 shows a scatterplot of measured and predicted water content values, including the performance of the worst and the best performing PTF (PTF01 and PTF29). The importance of including chemical properties and most importantly bulk density among the predictors is visible when measured water contents are greater than 0.50 cm<sup>3</sup> cm<sup>-3</sup>. Those high water content values are characteristic when the soil is close to saturation, thus indirect information about the structure is needed for more accurate predictions of those water content values. Parametric PTFs underestimate water content near saturation and between -200 and -15,000 cm matric potential head (Fig. S18). Overestimation occurs between -10 and -50 cm matric potential head and above 16000 cm matric potential head. When chemical soil properties are included, the degree of underestimation

decreases between -200 and -15,000 cm matric potential head, but overestimation increases between -5 and -10 cm with around  $0.02 \text{ cm}^3 \text{ cm}^{-3}$ .

### Hydraulic conductivity curve

OC, CaCO<sub>3</sub>, PH<sub>2</sub>O and CEC significantly improve the prediction of HCC when added to PSD and DEPTH. Adding BD next to PSD and DEPTH does not improve the predictions (Table 8, Fig. S19, S20). If PSD, DEPTH and OC are used as predictors, further addition of BD or CaCO<sub>3</sub> or PH<sub>2</sub>O or CEC does not significantly improve the performance of the PTFs. However, adding CaCO<sub>3</sub> and CEC or PH<sub>2</sub>O significantly improve the prediction. The performance of the worst and the best performing PTF is shown on Figure 6. The PTF with only PSD and DEPTH underestimate hydraulic conductivity values smaller than  $0.01 \text{ cm day}^{-1}$ . When OC, BD, PH<sub>2</sub>O and CEC are included, the underestimation decreases. This could be explained by the fact that these predictors contain indirect information of soil particle surface area and surface characteristics, which are some of the governing properties of low hydraulic conductivities.

When soil chemical properties are not used as predictors, hydraulic conductivity is underestimated close to saturation and at matric potential heads smaller than -500 cm; overestimation occurs between -10 and -500 cm matric potential head (Fig. S21). If chemical properties are also considered, hydraulic conductivity is i) underestimated at matric potential head smaller than -

5000 cm, and ii) overestimated between -5 and -5000 cm. With added information on chemical properties, the degree of underprediction decreases close to saturation and at the very dry end of the hydraulic conductivity curve. ~~In parts, this is not an effect of the PTFs but the limitations inherent to MVG to describe the entire hydraulic conductivity curve (Weber et al., 2019).~~ Increase in prediction performance for values lower than  $0.1 \text{ cm day}^{-1}$  is visible also on Figure 6.

~~Samples with measurements of the HCC at pressure heads < -1000 cm are less frequent and are not as numerous within a dataset of a single sample, if it was measured. Since the dataset of estimated VG model parameters were identical in this study and in Tóth et al. (2015), differences between the two studies of the unsaturated HCC are related to the PTF methods involved. However, at pressure heads < -1000 cm, the HCC is dominated by non-capillary conductivity (Streck and Weber, 2020; Weber et al., 2019), which is not included in the MVG model. The considerable data mismatch observable for the dry range (Fig. 6) can only be overcome by a different soil hydraulic property model and by a different PTF, because of compensatory effects in the VG. With this we mean that better data descriptions in the dry end, will lead to a larger mismatch in the wet end, as a consequence of the rigid model structure in the MVG model, which only accounts for capillary storage and conductivity. For better data description at < -1000 cm other more comprehensive models need to be adopted (Weber et al., 2020).~~

### 3.4 Comparison of point and parameter predictions

We compared the performance of the best point prediction methods (Table 2-5) with the best parameter estimations (Table 7) on the test sets. In 5 out of 20 cases, point predictions are significantly more accurate ~~and for further 8 cases RMSE were smaller~~. In all other cases, we have no significant difference between point and parametric PTFs (Table 9). ~~The reason for higher RMSE in parameter estimation can be that the VG model does not always adequately describe the measured MRC data~~

(Weber et al., 2019). Therefore, when THS, FC, FC\_2 and WP are computed with parameter estimation those are not only affected by the uncertainty of the prediction of VG parameters but by the goodness of VG model fit as well. We found similar results in the case of euptfv1 (Tóth et al., 2015). Tomasella et al. (2003) and Børgesen and Schaap (2005) had comparable findings regarding the performance of point and parametric PTFs. For THS point estimation performed better than parameter estimation. When the moisture retention curve is not needed, but only THS and/or FC/FC\_2 and/or WP, we recommend to compute THS, FC, FC\_2 and WP those with the point PTFs, more detailed explanation on it is included in Tóth et al. (2015).

### 3.5 Comparison of euptfv1 and v2

In 14 out of 19 cases, the PTFs of euptfv2 perform significantly better predicting the test sets than the PTFs of euptfv1. In the remaining 5 cases there is no significant difference (Table 10). Predictions of FC and MRC improve in all cases. The most important reason for it can be that the interaction between the target variable and the predictors is more complex for the cases of predicting FC or VG parameters – to describe the MRC –, which can be untangled using random forest. This may provide a reason the random forest algorithm performed significantly better than the PTFs derived with linear regression or a simple regression tree. For THS, WP, KS, and MVG only those PTFs did not improve significantly, for which comparisons on the TEST\_CHEM+ set were possible – which includes reduced number of samples. The RMSE of THS prediction was somewhat lower for euptfv1 than for euptfv2, but the difference was not significant. It could be due to the close to linear relationship between THS and BD and high relative importance of BD in THS prediction (84 %). This way their interaction can be efficiently described with the linear regression which is capable to extrapolate as well. Extrapolation with the random forest algorithm is not possible outside the training data, which can limit its performance. The general improvement of the PTFs in euptfv2 is twofold/threefold, the better performance is due to i) using random forest instead of single regression tree or linear regression, and ii) including more detailed information on soil sampling depth, not only distinguishing topsoils and subsoils and iii) providing information on prediction uncertainty.

We recommend the use of euptfv2 instead of euptfv1 if continuous soil properties are available. If only texture classes – i.e. no particle size distribution – are available, class PTFs of euptfv1 can be used, that is PTF18 for modified FAO texture classes and PTF19 for USDA texture classes.

### 4 Practical guidance on how to use the PTFs

The minimum input requirements for all PTFs are sand, silt and clay content, and soil depth. Soil depth needs to be considered in regard to the depth of the other input properties and soil hydraulic data needs, e.g. if the soil hydraulic properties of the top 20 cm (0-20 cm) is needed, then depth needs to be set at 10 cm in the input data of the prediction.

If only soil texture information is available for the predictions, the class PTFs from euptfv1 could be applied (Tóth et al., 2015).

We emphasise that:

1. the units of input soil properties (predictors) have to be the same as indicated in the text and that the sand, silt, and clay are defined by the following particle diameters: clay < 2  $\mu\text{m}$ , silt between 2 and 50  $\mu\text{m}$ , and sand between 50 and 2000  $\mu\text{m}$ .
  2. when only specific water content values at saturation, field capacity or wilting point are required (ie. THS, FC, WP) it is recommended to use point PTFs. This is also true for the prediction of KS.
  3. for AWC, the most accurate way is by first predicting FC and WP with the point predictions and then compute AWC using Eq. (1), and similarly for AWC<sub>2</sub> using FC<sub>2</sub> and Eq. (2).
  4. it is recommended to do the VG prediction if only moisture retention curve parameters are needed, and
  5. the MVG prediction when both moisture retention and hydraulic conductivity parameters are required.
- The VG algorithms predict the following van Genuchten model parameters: the residual water content  $\theta_r$  ( $\text{cm}^3 \text{ cm}^{-3}$ ), the saturated water content  $\theta_s$  ( $\text{cm}^3 \text{ cm}^{-3}$ ), and shape parameters  $\alpha$  ( $\text{cm}^{-1}$ ) and  $n$  (-). Parameter  $m$  is provided based on  $m=1-1/n$  (van Genuchten, 1980), and for the hydraulic conductivity curve, the two additional parameters:  $K_0$  ( $\text{cm day}^{-1}$ ) the hydraulic conductivity acting as a matching point at saturation and  $L$  (-), a shape parameter related to pore tortuosity. Table 11 shows the recommended PTFs for each predicted soil hydraulic property and available predictor variables. The users need to check which basic soil properties are available for the predictions, then look in Table 11 which PTF is recommended to use.
- The algorithms have been implemented in a web interface to facilitate the use of the PTFs, where the PTFs' selection is automated based on soil properties available for the predictions and required soil hydraulic property. The Code and data availability section provides information on how to access this resource.

## 45 Conclusions

The updated EU-PTFs – euptfv2 – perform significantly better than euptfv1 and are applicable for 32 predictor variables combinations. Uncertainties of the predicted soil hydraulic properties and model parameters can be computed. These uncertainties are, without further discrimination, related to the considered input data, predictors and the applied algorithm. The euptfv2 includes transfer functions to compute soil water content at saturation (0 cm matric potential head), field capacity (both -100 and -330 cm matric potential head) and wilting point (-15,000 cm matric potential head), plant available water content computed with field capacity at -100 and -330 cm matric potential head, saturated hydraulic conductivity, and Mualem-van Genuchten parameters of the moisture retention and hydraulic conductivity curves. For analyses of the impact as well as the significance of the uncertainties on the predicted soil hydraulic properties and model parameters, further studies are required.

**Code and data availability.** The current version of euptfv2 is available from a user friendly web interface: <https://ptfinterface.rissac.hu> (Szabó et al., 2019b) under the Creative Commons Attribution-NonCommercial 3.0



Unported License. The exact version of the model used to produce the results used in this paper is archived on Zenodo (<https://doi.org/10.5281/zenodo.3759443>, Szabó et al., 2020), as are the R scripts to develop the predictions and the derived pedotransfer functions – in RData format – presented in this paper. The training data set cannot be made publicly available due to legal restrictions of the EU-HYDI dataset, thus only a test sample is provided along with the model code.

5 **Supplement.** The Supplement related to this article is available online.

**Author contribution.** BSZ, TKDW and MW conceptualized the study and designed the methodology. BSZ supervised the research. MW cured the EU-HYDI dataset. BSZ, TKDW and MW prepared scripts for the statistical analysis, BSZ carried out the formal analysis, visualization and coordinated building of the PTF web interface. BSZ and TKDW performed the validation. BSZ and TKDW wrote the paper with considerable input from MW.

10 **Competing interests.** The authors declare that they have no conflict of interest.

**Acknowledgements.** B. Szabó is supported by the Hungarian National Research, Development and Innovation Office (NRDI) under grant KH124765 and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. On behalf of János Bolyai Research Scholarship we thank for the usage of MTA Cloud (<https://cloud.mta.hu/>) that significantly helped us achieving the results published in this paper. We praise the individual scientists and their institutions in Europe who contributed in the establishment of the EU-HYDI database. We would like to thank Gergely Tóth and Luca Montanarella for their effort in coordinating the formation of the EU-HYDI database. T. K.D Weber's contribution was supported by the Collaborative Research Center 1253 CAMPOS (Project 7: Stochastic Modelling Framework), funded by the German Research Foundation (DFG, Grant Agreement SFB 1253/1 2017). We thank to Jeromos Rózsa for defining the computing infrastructure. We also thank to Dávid Gyurkó for building the web interface for the prediction algorithms.

## 20 **References**

- Araya, S. N. and Ghezzehei, T. A.: Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations, *Water Resour. Res.*, 55(7), 5715–5737, doi:10.1029/2018WR024357, 2019.
- Assouline, S. and Or, D.: Conceptual and Parametric Representation of Soil Hydraulic Properties: A Review, *Vadose Zo. J.*, 12(4), doi:10.2136/vzj2013.07.0121, 2013.
- 25 Børgesen, C. D. and Schaap, M. G.: Point and parameter pedotransfer functions for water retention predictions for Danish soils, *Geoderma*, 127(1–2), 154–167, doi:10.1016/j.geoderma.2004.11.025, 2005.
- 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), 1-17., doi:10.2136/vzj2012.0123, 2013.
- Boulesteix, A. L., Janitza, S., Kruppa, J. and König, I. R.: Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics, *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 2(6), 493–507, doi:10.1002/widm.1072, 2012.
- Breiman, L.: Random Forests, *Mach. Learn.*, 45, 5–32, 2001.
- Deng, H., Ye, M., Schaap, M. G. and Khaleel, R.: Quantification of uncertainty in pedotransfer function-based parameter estimation for unsaturated flow modeling, *Water Resour. Res.*, 45(4), doi:10.1029/2008WR007477, 2009.
- van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am. J.*, 44, 892–898, 1980.
- Greenwell, B.: Partial Dependence Plots, , 15 [online] Available from: <https://cran.r-project.org/web/packages/pdp/pdp.pdf%0Ahttps://github.com/bgreenwell/pdp>, 2017a.
- Greenwell, B. M.: pdp: An R Package for Constructing Partial Dependence Plots, *R J.*, 9(1), 421–436 [online] Available from: <https://journal.r-project.org/archive/2017/RJ-2017-016/RJ-2017-016.pdf>, 2017b.
- Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning. Data Mining, Inference, and Prediction., Second edi., Springer. [online] Available from: [https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII\\_print10.pdf](https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print10.pdf) (Accessed 19 November 2018), 2009.
- Hengl, T., Nussbaum, M., Wright, M. N. and Heuvelink, B. M.: Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-temporal Variables, *PeerJ Prepr.*, 6:e5518, doi:10.7287/peerj.preprints.26693v3, 2018.
- Hodnett, M. G. and Tomasella, J.: Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer functions developed for tropical soils, *Geoderma*, 108(3–4), 155–180, doi:10.1016/S0016-7061(02)00105-2, 2002.
- Khodaverdiloo, H., Homaei, M., van Genuchten, M. T. and Dashtaki, S. G.: Deriving and validating pedotransfer functions for some calcareous soils, *J. Hydrol.*, 399(1–2), 93–99, doi:10.1016/j.jhydrol.2010.12.040, 2011.
- Kotlar, A. M., Jong van Lier, Q., Barros, A. H. C., Iversen, B. V. and Vereecken, H.: Development and Uncertainty Assessment of Pedotransfer Functions for Predicting Water Contents at Specific Pressure Heads, *Vadose Zo. J.*, 18(1), 190063, doi:10.2136/vzj2019.06.0063, 2019.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C. and Hunt, T.: Package “caret” Classification and Regression Training Description Misc functions for training and plotting classification and regression models. [online] Available from: <https://cran.r-project.org/web/packages/caret/caret.pdf>, 2017.
- Kuhn, M., Wing, J., Weston, S., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, R. C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C. and Hunt, T.: caret: Classification and Regression Training. R package version 6.0-79. [online] Available from: <https://github.com/topepo/caret/>, 2018.

- Lal, R. and Shukla, M. K.: Principles of soil physics, Marcel Dekker, Inc., New York., 2004.
- Li, D., Gao, G. and Fu, B.: Predicting available water of soil from particle-size distribution and bulk density in an oasis – desert transect in northwestern China, *J. Hydrol.*, 538, 539–550, doi:10.1016/j.jhydrol.2016.04.046, 2016.
- Lilly, A., Nemes, A., Rawls, W. J. and Pachepsky, Y. a: Probabilistic approach to the identification of input variables to estimate hydraulic conductivity, *Soil Sci. Soc. Am. J.*, 72(1), 16–24, doi:10.2136/sssaj2006.0391, 2008.
- 5 Lin, H.: Earth ' s Critical Zone and hydropedology : concepts , characteristics , and advances, *Hydrol. Earth Syst. Sci.*, 14., 25–45, doi:10.5194/hess-14-25-2010, 2010.
- 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.
- 10 Malone, B. P., McBratney, A. B., Minasny, B. and Laslett, G. M.: Mapping continuous depth functions of soil carbon storage and available water capacity, *Geoderma*, 154(1–2), 138–152, doi:10.1016/j.geoderma.2009.10.007, 2009.
- McNeill, S. J., Lilburne, L. R., Carrick, S., Webb, T. H. and Cuthill, T.: Pedotransfer functions for the soil water characteristics of New Zealand soils using S-map information, *Geoderma*, 326(April), 96–110, doi:10.1016/j.geoderma.2018.04.011, 15 2018.
- Meinshausen, N.: Quantile Regression Forests, *J. Mach. Learn. Res.*, 7, 983–999, 2006.
- De Mendiburu, F.: *agricolae: Statistical Procedures for Agricultural Research*. R package version 1.2-8. [online] Available from: <https://cran.r-project.org/package=agricolae>, 2017.
- 20 Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media., *Water Resour. Res.*, 12(3), 513–522, 1976.
- Nemes, A., Schaap, M. and Wösten, J.: Functional evaluation of pedotransfer functions derived from different scales of data collection, *Soil Sci. Soc. Am.*, 67, 1093–1102 [online] Available from: <https://www.soils.org/publications/sssaj/abstracts/67/4/1093> (Accessed 29 April 2013), 2003.
- 25 Nguyen, P. M., Haghverdi, A., Pue, J. De, Botula, Y., Le, K. V, Waegeman, W. and Cornelis, W. M.: Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils, *Biosyst. Eng.*, 153, 12–27, doi:10.1016/j.biosystemseng.2016.10.013, 2017.
- Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M. E. and Papritz, A.: Evaluation of digital soil mapping approaches with large sets of environmental covariates, *SOIL*, 4, 1–22, doi:10.5194/soil-4-1-2018, 30 2018.
- Odgers, N. P., McBratney, A. B. and Minasny, B.: Digital soil property mapping and uncertainty estimation using soil class probability rasters, *Geoderma*, 237, 190–198, doi:10.1016/j.geoderma.2014.09.009, 2014.
- Olson, R. S., La Cava, W., Mustahsan, Z., Varik, A. and Moore, J. H.: Data-driven advice for applying machine learning to bioinformatics problems, in *Biocomputing 2018*, edited by R. B. Altman, A. K. Dunker, L. Hunter, M. D. Ritchie, T. A.

Murray, and T. E. Klein, pp. 192–203, World Scientific., 2018.

Pachepsky, Y. A. and Rawls, W. J.: Accuracy and Reliability of Pedotransfer Functions as Affected by Grouping Soils, *Soil Sci. Soc. Am. J.*, 63(6), 1748–1757, doi:10.2136/sssaj1999.6361748x, 1999.

R Core Team: R: A language and environment for statistical computing., [online] Available from: <https://www.r-project.org>,  
5 2019.

Rawls, W., Pachepsky, Y. and Ritchie, J.: Effect of soil organic carbon on soil water retention, *Geoderma*, 116(1–2), 61–76, doi:10.1016/S0016-7061(03)00094-6, 2003.

Ridgeway, G.: gbm: Generalized Boosted Regression Models. R package version 2.1.3., 2017.

Román Dobarco, M., Cousin, I., Le Bas, C. and Martin, M. P.: Pedotransfer functions for predicting available water capacity  
10 in French soils, their applicability domain and associated uncertainty, *Geoderma*, 336(April 2018), 81–95, doi:10.1016/J.GEODERMA.2018.08.022, 2019.

Schaap, M. G.: Models for Indirect Estimation of Soil Hydraulic Properties, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson and J. J. McDonnell, John Wiley & Sons, Ltd, Chichester, UK., 2006.

Streck, T. and Weber, T. K. D.: Analytical expressions for noncapillary soil water retention based on popular capillary retention  
15 models, *Vadose Zo. J.*, 19(1), 1–5, doi:10.1002/vzj2.20042, 2020.

Szabó, B., Szatmári, G., Takács, K., Laborezi, A., Makó, A., Rajkai, K. and Pásztor, L.: Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics, *Hydrol. Earth Syst. Sci.*, 23, 2615–2635, doi:10.5194/hess-23-2615-2019, 2019a.

Szabó, B., Gyurkó, D., Weynants, M. and Weber, T. K. D.: Web interface for European hydraulic pedotransfer functions  
20 (euptfv2), , doi:10.34977/euptfv2.01, 2019b.

Szabó, B., Weynants, M. and Weber, T. K.: Scripts of updated European hydraulic pedotransfer functions (euptfv2), , doi:10.5281/ZENODO.3759443, 2020.

Szatmári, G. and Pásztor, L.: Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms, *Geoderma*, 337, 1329–1340, doi:10.1016/j.geoderma.2018.09.008, 2019.

25 Tomasella, J., Crestana, S. and Rawls, W. J.: Comparison of Two Techniques to Develop Pedotransfer Functions for Water Retention, *Soil Sci. Soc. Am. J.*, 67, 1085–1092, 2003.

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.

Tóth, B., Weynants, M., Pásztor, L. and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, *Hydrol. Process.*,  
30 31(14), 2662–2666, doi:10.1002/hyp.11203, 2017.

Tranter, G., Minasny, B. and McBratney, A. B.: Estimating Pedotransfer Function Prediction Limits Using Fuzzy -Means with Extragrades, *Soil Sci. Soc. Am. J.*, 74(6), 1967–1975, doi:10.2136/sssaj2009.0106, 2010.

Vaysse, K. and Lagacherie, P.: Using quantile regression forest to estimate uncertainty of digital soil mapping products, *Geoderma*, 291, 55–64, doi:10.1016/j.geoderma.2016.12.017, 2017.

- Vereecken, H., Maes, J., Feyen, J. and Darius, P.: Estimating the Soil Moisture Retention Characteristic From Texture, Bulk Density, and Carbon Content, *Soil Sci.*, 148(6), 389–403, doi:10.1097/00010694-198912000-00001, 1989.
- Vereecken, H., Huisman, J. A., Franssen Hendricks, H. J., Brüggemann, N., Boga, H. R., Kollet, S., Javaux, M., van der Kruk, J. and Vanderborght, J.: Soil hydrology: Recent methodological advances, challenges, and perspectives, *Water Resour. Res.*, 51, 1–18, doi:10.1002/2014WR016852, Received, 2015.
- Weber, T. K. D., Durner, W., Streck, T. and Diamantopoulos, E.: A Modular Framework for Modeling Unsaturated Soil Hydraulic Properties Over the Full Moisture Range, *Water Resour. Res.*, 4994–5011, doi:10.1029/2018WR024584, 2019.
- Weber, T. K. D., Finkel, M., Conceição Gonçalves, M., Vereecken, H. and Diamantopoulos, E.: Pedotransfer function for the Brunswick soil hydraulic property model and comparison to the van Genuchten-Mualem model, *Water Resour. Res.*, doi:10.1029/2019WR026820, 2020.
- Weynants, M. and Tóth, B.: The euptf package., *Eur. Soil Portal.*, 5 [online] Available from: <https://esdac.jrc.ec.europa.eu> (Accessed 19 May 2019), 2014.
- Weynants, M., Vereecken, H. and Javaux, M.: Revisiting Vereecken Pedotransfer Functions: Introducing a Closed-Form Hydraulic Model, *Vadose Zo. J.*, 8(1), 86, doi:10.2136/vzj2008.0062, 2009.
- Weynants, M., Montanarella, L., Tóth, G., Arnoldussen, A., Anaya Romero, M., Bilas, G., Borresen, T., Cornelis, W., Daroussin, J., Gonçalves, M. D. C., Haugen, L.-E., Hennings, V., Houskova, B., Iovino, M., Javaux, M., Keay, C. A., Kätterer, T., Kvaerno, S., Laktinova, T., Lamorski, K., Lilly, A., Mako, A., Matula, S., Morari, F., Nemes, A., Patyka, N. V., Romano, N., Schindler, U., Shein, E., Slawinski, C., Strauss, P., Tóth, B. and Woesten, H.: European HYdropedological Data Inventory (EU-HYDI), *EUR – Scientific and Technical Research series – ISSN 1831-9424, Luxembourg.*, 2013.
- Wiesmeier, M., Urbanski, L., Hobbey, E., Lang, B., von Lützw, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H. J. and Kögel-Knabner, I.: Soil organic carbon storage as a key function of soils - A review of drivers and indicators at various scales, *Geoderma*, 333, 149–162, doi:10.1016/j.geoderma.2018.07.026, 2019.
- Wösten, J. H. M., Lilly, A., Nemes, A., Bas, C. Le and Le Bas, C.: Development and use of a database of hydraulic properties of European soils, *Geoderma*, 90(3–4), 169–185, 1999.
- Wright, M. N. and Ziegler, A.: ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R, *J. Stat. Softw.*, 77(1), doi:10.18637/jss.v077.i01, 2017.
- Zhang, Y. and Schaap, M. G.: Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3), *J. Hydrol.*, 547(January 2017), 39–53, doi:10.1016/j.jhydrol.2017.01.004, 2017.
- Zhang, Y. and Schaap, M. G.: Estimation of saturated hydraulic conductivity with pedotransfer functions: A review, *J. Hydrol.*, 575(May), 1011–1030, doi:10.1016/j.jhydrol.2019.05.058, 2019.
- Zhao, H., Zeng, Y., Lv, S. and Su, Z.: Analysis of soil hydraulic and thermal properties for land surface modeling over the Tibetan Plateau, *Earth Syst. Sci. Data*, 10(2), 1031–1061, doi:10.5194/essd-10-1031-2018, 2018.

Ziegler, A. and König, I. R.: Mining data with random forests: Current options for real-world applications, Wiley Interdiscip. Rev. Data Min. Knowl. Discov., 4(1), 55–63, doi:10.1002/widm.1114, 2014.

TABLES

**Table 1.** Number of samples by predictor variable combinations used to derive the new European PTFs (euptfv2). Rows in italic font indicate PTFs with the same predictor variables as were tested in euptfv1 (Tóth et al., 2015).

Name	Predictor variables <sup>1</sup>	Number of samples in TRAIN set <sup>2</sup>								
		THS	FC_2	FC	WP	KS	AWC_2	AWC	VG	MVG
<i>PTF01</i>	<i>PSD+DEPTH</i>	3354	5109	2196	5264	3157	3528	1863	4669	739
<i>PTF02</i>	<i>PSD+DEPTH+OC</i>	2966	4131	1716	4802	2620	3208	1650	3708	407
PTF03	PSD+DEPTH+BD	3305	5034	2176	5197	3146	3472	1849	4593	726
PTF04	PSD+DEPTH+CACO3	678	1670	1537	1816	639	1548	1531	1671	273
PTF05	PSD+DEPTH+PH_H2O	1203	2062	1278	2039	907	1849	1245	1897	230
PTF06	PSD+DEPTH+CEC	895	1649	1097	1703	567	1550	1092	1488	141
<i>PTF07</i>	<i>PSD+DEPTH+OC+BD</i>	2959	4117	1711	4786	2609	3197	1645	3695	404
PTF08	PSD+DEPTH+OC+CACO3	673	1586	1340	1599	613	1464	1336	1589	250
PTF09	PSD+DEPTH+OC+PH_H2O	1052	1808	1100	1678	862	1615	1074	1663	224
PTF10	PSD+DEPTH+OC+CEC	744	1437	1001	1459	525	1358	998	1293	138
PTF11	PSD+DEPTH+BD+CACO3	678	1666	1526	1806	639	1545	1522	1670	272
PTF12	PSD+DEPTH+BD+PH_H2O	1156	2008	1267	1979	898	1796	1236	1847	229
PTF13	PSD+DEPTH+BD+CEC	848	1596	1093	1648	558	1498	1088	1437	140
PTF14	PSD+DEPTH+CACO3+PH_H2O	678	1314	1235	1375	620	1195	1230	1264	223
PTF15	PSD+DEPTH+CACO3+CEC	373	770	793	831	405	726	791	758	136
PTF16	PSD+DEPTH+PH_H2O+CEC	894	1350	744	1349	567	1255	739	1188	141
PTF17	PSD+DEPTH+OC+BD+CACO3	673	1585	1338	1596	613	1464	1334	1588	249
<i>PTF18</i>	<i>PSD+DEPTH+OC+BD+PH_H2O</i>	1047	1799	1098	1667	853	1607	1072	1655	223
PTF19	PSD+DEPTH+OC+BD+CEC	739	1427	998	1447	516	1349	995	1284	137
PTF20	PSD+DEPTH+OC+CACO3+PH_H2O	673	1249	1062	1183	613	1130	1059	1201	219
PTF21	PSD+DEPTH+OC+CACO3+CEC	369	727	709	743	401	683	707	712	135
PTF22	PSD+DEPTH+OC+PH_H2O+CEC	744	1142	663	1121	525	1067	660	996	138
PTF23	PSD+DEPTH+BD+CACO3+PH_H2O	678	1310	1224	1365	620	1192	1221	1263	222
PTF24	PSD+DEPTH+BD+CACO3+CEC	373	768	790	827	405	725	788	757	135
PTF25	PSD+DEPTH+BD+PH_H2O+CEC	847	1298	741	1295	558	1204	736	1138	140
PTF26	PSD+DEPTH+CACO3+PH_H2O+CEC	373	727	734	772	405	684	732	717	136
PTF27	PSD+DEPTH+OC+BD+CACO3+PH_H2O	673	1248	1060	1180	613	1130	1057	1200	218
PTF28	PSD+DEPTH+OC+BD+CACO3+CEC	369	726	707	740	401	683	705	711	134
PTF29	PSD+DEPTH+OC+BD+PH_H2O+CEC	739	1133	661	1110	516	1059	658	988	137
<i>PTF30</i>	<i>PSD+DEPTH+OC+CACO3+PH_H2O+CEC</i>	369	684	655	689	401	641	653	671	135
PTF31	PSD+DEPTH+BD+CACO3+PH_H2O+CEC	373	725	731	768	405	683	729	716	135
<i>PTF32</i>	<i>PSD+DEPTH+OC+BD+CACO3+PH_H2O+CEC</i>	369	683	653	686	401	641	651	670	134
Number of samples in TEST_BASIC set		1247	1762	801	2088	1117	1372	705	1591	176
Number of samples in TEST_CHEM+ set		156	296	280	294	169	274	279	288	57

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

<sup>2</sup>THS: saturated water content (pF 0); FC\_2: water content at -100 cm matric potential head (pF 2.0); FC: water content at -330 cm matric potential head (pF 2.5); AWC\_2: plant available water content based on FC\_2; AWC: plant available water content based on FC; WP: water content at wilting point (pF 4.2); KS: saturated hydraulic conductivity; VG: parameters of the van Genuchten model; MVG: parameters of the Mualem – van Genuchten model; TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC.

**Table 2.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict water content at saturation (THS). N: number of samples, RMSE: root mean square error ( $\text{cm}^3 \text{cm}^{-3}$ ), and  $R^2$ : determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD,  $\text{CaCO}_3$ ,  $\text{pH}_2\text{O}$  and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in euptfv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from euptfv1
		N	RMSE	$R^2$	N	RMSE	$R^2$	TEST_BASIC set	TEST_CHEM+ set		
PTF01	PSD+DEPTH	3354	0.067	0.366	1274	0.068	0.344	a	a	PTF01	-
PTF02	PSD+DEPTH+OC	2966	0.053	0.577	1274	0.056	0.552	b	abc	PTF02	PTF04
PTF03	PSD+DEPTH+BD	3305	0.029	0.880	1274	0.031	0.862	c	d	PTF03	-
PTF04	PSD+DEPTH+ $\text{CaCO}_3$	678	0.046	0.187	156	0.057	0.053	-	bc	PTF04	-
PTF05	PSD+DEPTH+ $\text{pH}_2\text{O}$	1203	0.056	0.298	156	0.053	0.193	-	bc	PTF05	-
PTF06	PSD+DEPTH+CEC	895	0.055	0.401	156	0.057	0.048	-	ab	PTF01	-
PTF07	PSD+DEPTH+OC+BD	2959	0.027	0.888	1274	0.030	0.869	c	d	PTF03	PTF05
PTF08	PSD+DEPTH+OC+ $\text{CaCO}_3$	673	0.044	0.209	156	0.055	0.118	-	bc	PTF02	-
PTF09	PSD+DEPTH+OC+ $\text{pH}_2\text{O}$	1052	0.046	0.457	156	0.050	0.272	-	c	PTF02	-
PTF10	PSD+DEPTH+OC+CEC	744	0.046	0.519	156	0.051	0.233	-	abc	PTF02	-
PTF11	PSD+DEPTH+BD+ $\text{CaCO}_3$	678	0.023	0.791	156	0.022	0.863	-	d	PTF03	-
PTF12	PSD+DEPTH+BD+ $\text{pH}_2\text{O}$	1156	0.027	0.826	156	0.021	0.878	-	d	PTF03	-
PTF13	PSD+DEPTH+BD+CEC	848	0.027	0.848	156	0.021	0.873	-	d	PTF03	-
PTF14	PSD+DEPTH+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$	678	0.045	0.231	156	0.050	0.265	-	bc	PTF05	-
PTF15	PSD+DEPTH+ $\text{CaCO}_3$ +CEC	373	0.045	0.257	156	0.054	0.164	-	abc	PTF04	-
PTF16	PSD+DEPTH+ $\text{pH}_2\text{O}$ +CEC	894	0.052	0.459	156	0.055	0.132	-	bc	PTF05	-
PTF17	PSD+DEPTH+OC+BD+ $\text{CaCO}_3$	673	0.019	0.856	156	0.021	0.872	-	d	PTF03	-
PTF18	PSD+DEPTH+OC+BD+ $\text{pH}_2\text{O}$	1047	0.024	0.848	156	0.021	0.871	-	d	PTF03	PTF06
PTF19	PSD+DEPTH+OC+BD+CEC	739	0.027	0.837	156	0.021	0.874	-	d	PTF03	-
PTF20	PSD+DEPTH+OC+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$	673	0.043	0.251	156	0.050	0.285	-	c	PTF02	-
PTF21	PSD+DEPTH+OC+ $\text{CaCO}_3$ +CEC	369	0.043	0.309	156	0.051	0.242	-	bc	PTF02	-
PTF22	PSD+DEPTH+OC+ $\text{pH}_2\text{O}$ +CEC	744	0.046	0.531	156	0.050	0.280	-	bc	PTF02	-
PTF23	PSD+DEPTH+BD+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$	678	0.023	0.796	156	0.021	0.869	-	d	PTF03	-
PTF24	PSD+DEPTH+BD+ $\text{CaCO}_3$ +CEC	373	0.021	0.841	156	0.021	0.869	-	d	PTF03	-
PTF25	PSD+DEPTH+BD+ $\text{pH}_2\text{O}$ +CEC	847	0.027	0.850	156	0.020	0.883	-	d	PTF03	-
PTF26	PSD+DEPTH+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$ +CEC	373	0.044	0.305	156	0.049	0.308	-	abc	PTF05	-
PTF27	PSD+DEPTH+OC+BD+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$	673	0.019	0.858	156	0.022	0.865	-	d	PTF03	-
PTF28	PSD+DEPTH+OC+BD+ $\text{CaCO}_3$ +CEC	369	0.021	0.845	156	0.021	0.874	-	d	PTF03	-
PTF29	PSD+DEPTH+OC+BD+ $\text{pH}_2\text{O}$ +CEC	739	0.026	0.843	156	0.020	0.880	-	d	PTF03	-
PTF30	PSD+DEPTH+OC+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$ +CEC	369	0.042	0.356	156	0.049	0.319	-	bc	PTF02	PTF04
PTF31	PSD+DEPTH+BD+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$ +CEC	373	0.021	0.843	156	0.021	0.871	-	d	PTF03	-
PTF32	PSD+DEPTH+OC+BD+ $\text{CaCO}_3$ + $\text{pH}_2\text{O}$ +CEC	369	0.021	0.844	156	0.021	0.876	-	d	PTF03	PTF06

<sup>1</sup>PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ );  $\text{CaCO}_3$ : calcium carbonate content (mass %);  $\text{pH}_2\text{O}$ : pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.



5

**Table 3.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict water content at -100 cm matric potential head (FC\_2). N: number of samples, RMSE: root mean square error ( $\text{cm}^3 \text{ cm}^{-3}$ ), and  $R^2$ : determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC. Recommended PTFs are highlighted in bold. FC\_2 was not analysed in euptfv1.

Name of PTF in euptfv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from euptfv1
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST_BASIC set	TEST_CHEM+ set		
PTF01	PSD+DEPTH	5109	0.062	0.651	1762	0.060	0.669	a	a	PTF01	-
PTF02	PSD+DEPTH+OC	4131	0.057	0.711	1762	0.055	0.718	b	ab	PTF02	-
PTF03	PSD+DEPTH+BD	5034	0.053	0.750	1762	0.052	0.745	bc	bcdef	PTF03	-
PTF04	PSD+DEPTH+CACO3	1670	0.052	0.566	296	0.054	0.467	-	abcd	PTF01	-
PTF05	PSD+DEPTH+PH_H2O	2062	0.056	0.630	296	0.056	0.419	-	abc	PTF01	-
PTF06	PSD+DEPTH+CEC	1649	0.056	0.658	296	0.054	0.469	-	abcde	PTF01	-
PTF07	PSD+DEPTH+OC+BD	4117	0.051	0.769	1762	0.050	0.769	c	bcdefg	PTF03	-
PTF08	PSD+DEPTH+OC+CACO3	1586	0.050	0.589	296	0.049	0.565	-	bcdefgh	PTF02	-
PTF09	PSD+DEPTH+OC+PH_H2O	1808	0.050	0.679	296	0.048	0.581	-	bcdefg	PTF02	-
PTF10	PSD+DEPTH+OC+CEC	1437	0.051	0.688	296	0.049	0.554	-	cdefghij	PTF06	-
PTF11	PSD+DEPTH+BD+CACO3	1666	0.044	0.701	296	0.046	0.616	-	fghijklmn	PTF03	-
PTF12	PSD+DEPTH+BD+PH_H2O	2008	0.046	0.746	296	0.043	0.657	-	efghijkl	PTF03	-
PTF13	PSD+DEPTH+BD+CEC	1596	0.046	0.763	296	0.046	0.614	-	hijklmn	PTF13	-
PTF14	PSD+DEPTH+CACO3+PH_H2O	1314	0.051	0.600	296	0.051	0.528	-	bcdef	PTF05	-
PTF15	PSD+DEPTH+CACO3+CEC	770	0.052	0.605	296	0.051	0.520	-	cdefghij	PTF04	-
PTF16	PSD+DEPTH+PH_H2O+CEC	1350	0.053	0.699	296	0.049	0.556	-	cdefghi	PTF05	-
PTF17	PSD+DEPTH+OC+BD+CACO3	1585	0.043	0.689	296	0.045	0.634	-	ghijklmn	PTF07	-
PTF18	PSD+DEPTH+OC+BD+PH_H2O	1799	0.044	0.749	296	0.042	0.679	-	ghijklmn	PTF07	-
PTF19	PSD+DEPTH+OC+BD+CEC	1427	0.045	0.753	296	0.044	0.650	-	jklmn	PTF13	-
PTF20	PSD+DEPTH+OC+CACO3+PH_H2O	1249	0.049	0.613	296	0.053	0.483	-	bcdefgh	PTF02	-
PTF21	PSD+DEPTH+OC+CACO3+CEC	727	0.050	0.603	296	0.046	0.620	-	fghijklmn	PTF08	-
PTF22	PSD+DEPTH+OC+PH_H2O+CEC	1142	0.051	0.693	296	0.045	0.630	-	efghijklm	PTF09	-
PTF23	PSD+DEPTH+BD+CACO3+PH_H2O	1310	0.044	0.701	296	0.045	0.629	-	defghijkl	PTF03	-
PTF24	PSD+DEPTH+BD+CACO3+CEC	768	0.043	0.722	296	0.043	0.666	-	lmn	PTF11	-
PTF25	PSD+DEPTH+BD+PH_H2O+CEC	1298	0.046	0.773	296	0.043	0.668	-	jklmn	PTF12	-
PTF26	PSD+DEPTH+CACO3+PH_H2O+CEC	727	0.051	0.633	296	0.048	0.587	-	defghijk	PTF05	-
PTF27	PSD+DEPTH+OC+BD+CACO3+ PH_H2O	1248	0.043	0.693	296	0.044	0.653	-	efghijklm	PTF07	-
PTF28	PSD+DEPTH+OC+BD+CACO3+CEC	726	0.044	0.702	296	0.041	0.687	-	klmn	PTF11	-
PTF29	PSD+DEPTH+OC+BD+PH_H2O+CEC	1133	0.046	0.757	296	0.042	0.681	-	ijklmn	PTF12	-
PTF30	PSD+DEPTH+OC+CACO3+PH_H2O+ CEC	684	0.050	0.617	296	0.051	0.533	-	efghijklm	PTF09	-
PTF31	PSD+DEPTH+BD+CACO3+PH_H2O+ CEC	725	0.043	0.731	296	0.041	0.698	-	mn	PTF11	-
PTF32	PSD+DEPTH+OC+BD+CACO3+ PH_H2O+CEC	683	0.044	0.712	296	0.040	0.709	-	n	PTF18	-

<sup>1</sup>PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

**Table 4.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict water content at -330 cm matric potential head, field capacity (FC). N: number of samples, RMSE: root mean square error (cm<sup>3</sup> cm<sup>-3</sup>), and R<sup>2</sup>: determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in eupftv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from eupftv1
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST_BASIC set	TEST_CHEM+ set		
PTF01	<b>PSD+DEPTH</b>	2196	0.056	0.639	801	0.054	0.595	a	a	PTF01	-
PTF02	<b>PSD+DEPTH+OC</b>	1716	0.049	0.707	801	0.050	0.650	b	abc	PTF02	PTF09
PTF03	<b>PSD+DEPTH+BD</b>	2176	0.048	0.727	801	0.049	0.668	ab	abcd	PTF01	-
PTF04	<b>PSD+DEPTH+CACO3</b>	1537	0.047	0.650	280	0.055	0.591	-	abcde	PTF01	-
PTF05	PSD+DEPTH+PH_H2O	1278	0.048	0.653	280	0.055	0.586	-	ab	PTF01	-
PTF06	<b>PSD+DEPTH+CEC</b>	1097	0.046	0.711	280	0.052	0.630	-	bcdefghi	PTF06	-
PTF07	<b>PSD+DEPTH+OC+BD</b>	1711	0.046	0.736	801	0.048	0.677	b	bcdefg	PTF02	PTF09
PTF08	<b>PSD+DEPTH+OC+CACO3</b>	1340	0.043	0.678	280	0.053	0.616	-	abcdef	PTF02	-
PTF09	<b>PSD+DEPTH+OC+PH_H2O</b>	1100	0.044	0.687	280	0.052	0.631	-	abcde	PTF02	-
PTF10	PSD+DEPTH+OC+CEC	1001	0.044	0.720	280	0.052	0.628	-	bcdefghi	PTF02	-
PTF11	<b>PSD+DEPTH+BD+CACO3</b>	1526	0.044	0.696	280	0.051	0.649	-	bcdefgh	PTF03	-
PTF12	PSD+DEPTH+BD+PH_H2O	1267	0.045	0.698	280	0.050	0.658	-	bcdefgh	PTF03	-
PTF13	PSD+DEPTH+BD+CEC	1093	0.044	0.741	280	0.049	0.678	-	fghi	PTF06	-
PTF14	<b>PSD+DEPTH+CACO3+PH_H2O</b>	1235	0.048	0.667	280	0.053	0.623	-	bcdef	PTF04	-
PTF15	PSD+DEPTH+CACO3+CEC	793	0.047	0.720	280	0.052	0.639	-	efghi	PTF04	-
PTF16	PSD+DEPTH+PH_H2O+CEC	744	0.047	0.726	280	0.051	0.651	-	efghi	PTF06	-
PTF17	PSD+DEPTH+OC+BD+CACO3	1338	0.042	0.699	280	0.050	0.667	-	cdefghi	PTF02	-
PTF18	PSD+DEPTH+OC+BD+PH_H2O	1098	0.043	0.704	280	0.050	0.660	-	bcdefgh	PTF02	PTF09
PTF19	PSD+DEPTH+OC+BD+CEC	998	0.042	0.739	280	0.048	0.684	-	fghi	PTF07	-
PTF20	PSD+DEPTH+OC+CACO3+PH_H2O	1062	0.044	0.694	280	0.052	0.634	-	abcde	PTF02	-
PTF21	PSD+DEPTH+OC+CACO3+CEC	709	0.045	0.709	280	0.051	0.652	-	efghi	PTF04	-
PTF22	PSD+DEPTH+OC+PH_H2O+CEC	663	0.046	0.706	280	0.050	0.664	-	defghi	PTF09	-
PTF23	PSD+DEPTH+BD+CACO3+PH_H2O	1224	0.045	0.704	280	0.051	0.651	-	bcdefgh	PTF03	-
PTF24	PSD+DEPTH+BD+CACO3+CEC	790	0.044	0.744	280	0.048	0.688	-	hi	PTF11	-
PTF25	PSD+DEPTH+BD+PH_H2O+CEC	741	0.045	0.748	280	0.048	0.682	-	hi	PTF11	-
PTF26	PSD+DEPTH+CACO3+PH_H2O+CEC	734	0.046	0.742	280	0.050	0.658	-	fghi	PTF14	-
PTF27	PSD+DEPTH+OC+BD+CACO3+ PH_H2O	1060	0.042	0.712	280	0.049	0.676	-	bcdefghi	PTF02	-
PTF28	PSD+DEPTH+OC+BD+CACO3+CEC	707	0.043	0.731	280	0.048	0.693	-	ghi	PTF07	-
PTF29	PSD+DEPTH+OC+BD+PH_H2O+CEC	661	0.044	0.725	280	0.046	0.709	-	fghi	PTF07	-
PTF30	PSD+DEPTH+OC+CACO3+PH_H2O+ CEC	655	0.044	0.731	280	0.049	0.672	-	fghi	PTF08	PTF09
PTF31	PSD+DEPTH+BD+CACO3+PH_H2O+ CEC	731	0.043	0.763	280	0.047	0.700	-	i	PTF06	-
PTF32	PSD+DEPTH+OC+BD+CACO3+ PH_H2O+CEC	653	0.043	0.743	280	0.047	0.696	-	fghi	PTF07	PTF09

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

**Table 5.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict water content at wilting point (WP). N: number of samples, RMSE: root mean square error ( $\text{cm}^3 \text{ cm}^{-3}$ ), and  $R^2$ : determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in eupftv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from eupftv1
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST_BASIC set	TEST_CHEM+ set		
PTF01	<b>PSD+DEPTH</b>	5264	0.048	0.736	2088	0.048	0.728	a	a	PTF01	-
PTF02	<b>PSD+DEPTH+OC</b>	4802	0.047	0.755	2088	0.046	0.745	bc	abc	PTF02	PTF12
PTF03	PSD+DEPTH+BD	5197	0.046	0.757	2088	0.046	0.754	ab	ab	PTF01	-
PTF04	PSD+DEPTH+CACO3	1816	0.042	0.693	294	0.042	0.643	-	a	PTF01	-
PTF05	PSD+DEPTH+PH_H2O	2039	0.046	0.673	294	0.044	0.621	-	abc	PTF01	-
PTF06	PSD+DEPTH+CEC	1703	0.043	0.725	294	0.041	0.662	-	a	PTF01	-
PTF07	<b>PSD+DEPTH+OC+BD</b>	4786	0.045	0.769	2088	0.044	0.769	c	abc	PTF02	PTF12
PTF08	<b>PSD+DEPTH+OC+CACO3</b>	1599	0.041	0.695	294	0.041	0.671	-	abcd	PTF02	-
PTF09	<b>PSD+DEPTH+OC+PH_H2O</b>	1678	0.045	0.682	294	0.041	0.661	-	abcd	PTF02	-
PTF10	PSD+DEPTH+OC+CEC	1459	0.043	0.704	294	0.040	0.674	-	abcd	PTF02	-
PTF11	PSD+DEPTH+BD+CACO3	1806	0.041	0.706	294	0.040	0.682	-	abcd	PTF01	-
PTF12	PSD+DEPTH+BD+PH_H2O	1979	0.045	0.691	294	0.041	0.671	-	abcd	PTF01	-
PTF13	PSD+DEPTH+BD+CEC	1648	0.042	0.729	294	0.040	0.683	-	abcd	PTF01	-
PTF14	PSD+DEPTH+CACO3+PH_H2O	1375	0.043	0.689	294	0.042	0.649	-	abcd	PTF01	-
PTF15	PSD+DEPTH+CACO3+CEC	831	0.044	0.657	294	0.039	0.694	-	abcd	PTF01	-
PTF16	PSD+DEPTH+PH_H2O+CEC	1349	0.043	0.727	294	0.040	0.681	-	abc	PTF01	-
PTF17	<b>PSD+DEPTH+OC+BD+CACO3</b>	1596	0.041	0.705	294	0.039	0.702	-	abcd	PTF07	-
PTF18	PSD+DEPTH+OC+BD+PH_H2O	1667	0.045	0.687	294	0.040	0.674	-	abcd	PTF07	PTF12
PTF19	PSD+DEPTH+OC+BD+CEC	1447	0.042	0.714	294	0.039	0.691	-	abcd	PTF07	-
PTF20	PSD+DEPTH+OC+CACO3+PH_H2O	1183	0.042	0.691	294	0.040	0.686	-	abcd	PTF02	-
PTF21	PSD+DEPTH+OC+CACO3+CEC	743	0.044	0.638	294	0.037	0.722	-	d	PTF08	-
PTF22	PSD+DEPTH+OC+PH_H2O+CEC	1121	0.044	0.697	294	0.039	0.701	-	abcd	PTF07	-
PTF23	PSD+DEPTH+BD+CACO3+PH_H2O	1365	0.042	0.701	294	0.040	0.678	-	abcd	PTF01	-
PTF24	PSD+DEPTH+BD+CACO3+CEC	827	0.043	0.673	294	0.038	0.708	-	abcd	PTF01	-
PTF25	PSD+DEPTH+BD+PH_H2O+CEC	1295	0.043	0.726	294	0.039	0.698	-	abcd	PTF01	-
PTF26	PSD+DEPTH+CACO3+PH_H2O+CEC	772	0.043	0.680	294	0.039	0.702	-	cd	PTF05	-
PTF27	PSD+DEPTH+OC+BD+CACO3+ PH_H2O	1180	0.042	0.698	294	0.039	0.703	-	abcd	PTF07	-
PTF28	PSD+DEPTH+OC+BD+CACO3+CEC	740	0.043	0.648	294	0.037	0.732	-	bcd	PTF17	-
PTF29	PSD+DEPTH+OC+BD+PH_H2O+CEC	1110	0.043	0.699	294	0.038	0.712	-	abcd	PTF07	-
PTF30	PSD+DEPTH+OC+CACO3+PH_H2O+ CEC	689	0.044	0.645	294	0.038	0.719	-	abcd	PTF02	PTF12
PTF31	PSD+DEPTH+BD+CACO3+PH_H2O+ CEC	768	0.043	0.678	294	0.037	0.720	-	cd	PTF05	-
PTF32	PSD+DEPTH+OC+BD+CACO3+ PH_H2O+CEC	686	0.043	0.656	294	0.037	0.723	-	d	PTF09	PTF12

<sup>1</sup>PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

**Table 6.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict saturated hydraulic conductivity (KS). N: number of samples, RMSE: root mean square error ( $\log_{10}$  (cm day<sup>-1</sup>)), and R<sup>2</sup>: determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CaCO<sub>3</sub>, PH<sub>2</sub>O and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in eupftv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from eupftv1
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST_BASIC set	TEST_CHEM+ set		
<b>PTF01</b>	<b>PSD+DEPTH</b>	3157	1.200	0.434	1117	1.181	0.307	a	ab	PTF01	-
<b>PTF02</b>	<b>PSD+DEPTH+OC</b>	2620	0.957	0.566	1117	0.953	0.548	b	bc	PTF02	PTF16
PTF03	PSD+DEPTH+BD	3146	1.160	0.467	1117	1.170	0.320	a	a	PTF01	-
PTF04	PSD+DEPTH+CaCO3	639	0.861	0.241	169	0.959	0.123	-	abc	PTF01	-
<b>PTF05</b>	<b>PSD+DEPTH+PH<sub>2</sub>O</b>	907	0.875	0.213	169	0.944	0.151	-	bc	PTF01	-
PTF06	PSD+DEPTH+CEC	567	0.984	0.215	169	0.940	0.157	-	bc	PTF01	-
PTF07	PSD+DEPTH+OC+BD	2609	0.931	0.590	1117	0.939	0.562	b	bc	PTF02	PTF16
PTF08	PSD+DEPTH+OC+CaCO3	613	0.872	0.244	169	0.943	0.153	-	bc	PTF02	-
PTF09	PSD+DEPTH+OC+PH <sub>2</sub> O	862	0.847	0.257	169	0.938	0.162	-	bc	PTF02	-
PTF10	PSD+DEPTH+OC+CEC	525	0.977	0.223	169	0.938	0.162	-	bc	PTF02	-
PTF11	PSD+DEPTH+BD+CaCO3	639	0.851	0.259	169	0.952	0.136	-	bc	PTF01	-
PTF12	PSD+DEPTH+BD+PH <sub>2</sub> O	898	0.853	0.256	169	0.947	0.145	-	bc	PTF05	-
PTF13	PSD+DEPTH+BD+CEC	558	0.980	0.230	169	0.941	0.157	-	bc	PTF01	-
PTF14	PSD+DEPTH+CaCO3+PH <sub>2</sub> O	620	0.855	0.267	169	0.923	0.189	-	bc	PTF05	-
PTF15	PSD+DEPTH+CaCO3+CEC	405	0.937	0.263	169	0.941	0.156	-	abc	PTF01	-
PTF16	PSD+DEPTH+PH <sub>2</sub> O+CEC	567	0.942	0.282	169	0.940	0.158	-	bc	PTF01	-
PTF17	PSD+DEPTH+OC+BD+CaCO3	613	0.856	0.272	169	0.933	0.171	-	bc	PTF02	-
PTF18	PSD+DEPTH+OC+BD+PH <sub>2</sub> O	853	0.831	0.289	169	0.932	0.172	-	bc	PTF02	PTF16
PTF19	PSD+DEPTH+OC+BD+CEC	516	0.979	0.228	169	0.928	0.179	-	c	PTF02	-
PTF20	PSD+DEPTH+OC+CaCO3+PH <sub>2</sub> O	613	0.860	0.264	169	0.929	0.177	-	bc	PTF02	-
PTF21	PSD+DEPTH+OC+CaCO3+CEC	401	0.935	0.271	169	0.925	0.184	-	bc	PTF02	-
PTF22	PSD+DEPTH+OC+PH <sub>2</sub> O+CEC	525	0.931	0.295	169	0.933	0.170	-	c	PTF02	-
PTF23	PSD+DEPTH+BD+CaCO3+PH <sub>2</sub> O	620	0.844	0.286	169	0.889	0.247	-	c	PTF05	-
PTF24	PSD+DEPTH+BD+CaCO3+CEC	405	0.922	0.286	169	0.958	0.125	-	abc	PTF01	-
PTF25	PSD+DEPTH+BD+PH <sub>2</sub> O+CEC	558	0.944	0.286	169	0.950	0.140	-	bc	PTF05	-
PTF26	PSD+DEPTH+CaCO3+PH <sub>2</sub> O+CEC	405	0.922	0.286	169	0.922	0.190	-	bc	PTF05	-
PTF27	PSD+DEPTH+OC+BD+CaCO3+ PH <sub>2</sub> O	613	0.844	0.293	169	0.893	0.241	-	c	PTF02	-
PTF28	PSD+DEPTH+OC+BD+CaCO3+CEC	401	0.926	0.285	169	0.925	0.185	-	abc	PTF02	-
PTF29	PSD+DEPTH+OC+BD+PH <sub>2</sub> O+CEC	516	0.932	0.301	169	0.921	0.193	-	bc	PTF02	-
PTF30	PSD+DEPTH+OC+CaCO3+PH <sub>2</sub> O+ CEC	401	0.931	0.278	169	0.887	0.250	-	bc	PTF02	PTF17
PTF31	PSD+DEPTH+BD+CaCO3+PH <sub>2</sub> O+ CEC	405	0.914	0.298	169	0.912	0.207	-	bc	PTF05	-
PTF32	PSD+DEPTH+OC+BD+CaCO3+ PH <sub>2</sub> O+CEC	401	0.921	0.292	169	0.916	0.201	-	bc	PTF02	PTF17

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CaCO<sub>3</sub>: calcium carbonate content (mass %); PH<sub>2</sub>O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

5 PTFs are highlighted in bold.

**Table 7.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict parameters of the van Genuchten model to describe soil moisture retention curve (VG). N: number of samples, RMSE: root mean square error ( $\text{cm}^3 \text{cm}^{-3} (\theta_{\text{org}} - (\text{cm} \cdot \text{day}^{-1}))$ ), and  $R^2$ : determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD,  $\text{CaCO}_3$ ,  $\text{pH}_{\text{H}_2\text{O}}$  and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in euptfv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from euptfv1
		N	RMSE	$R^2$	N	RMSE	$R^2$	TEST_BASIC set	TEST_CHEM+ set		
PTF01	<b>PSD+DEPTH</b>	4669	0.055	0.846	1591	0.068	0.776	a	a	PTF01	-
PTF02	<b>PSD+DEPTH+OC</b>	3708	0.047	0.887	1591	0.060	0.826	b	c	PTF02	PTF19
PTF03	<b>PSD+DEPTH+BD</b>	4593	0.041	0.913	1591	0.056	0.846	c	hi	PTF03	-
PTF04	<b>PSD+DEPTH+CaCO3</b>	1671	0.039	0.911	288	0.052	0.852	-	d	PTF04	-
PTF05	<b>PSD+DEPTH+PH_H2O</b>	1897	0.045	0.894	288	0.055	0.834	-	b	PTF05	-
PTF06	<b>PSD+DEPTH+CEC</b>	1488	0.044	0.886	288	0.054	0.839	-	d	PTF06	-
PTF07	<b>PSD+DEPTH+OC+BD</b>	3695	0.037	0.933	1591	0.054	0.859	d	fg	PTF07	PTF21
PTF08	<b>PSD+DEPTH+OC+CaCO3</b>	1589	0.036	0.924	288	0.048	0.871	-	f	PTF08	-
PTF09	<b>PSD+DEPTH+OC+PH_H2O</b>	1663	0.039	0.922	288	0.050	0.865	-	gh	PTF09	-
PTF10	<b>PSD+DEPTH+OC+CEC</b>	1293	0.036	0.920	288	0.051	0.858	-	fg	PTF10	-
PTF11	<b>PSD+DEPTH+BD+CaCO3</b>	1670	0.034	0.934	288	0.043	0.900	-	mn	PTF11	-
PTF12	<b>PSD+DEPTH+BD+PH_H2O</b>	1847	0.038	0.926	288	0.044	0.892	-	l	PTF12	-
PTF13	<b>PSD+DEPTH+BD+CEC</b>	1437	0.039	0.908	288	0.044	0.892	-	lm	PTF13	-
PTF14	<b>PSD+DEPTH+CaCO3+PH_H2O</b>	1264	0.037	0.928	288	0.052	0.854	-	e	PTF14	-
PTF15	<b>PSD+DEPTH+CaCO3+CEC</b>	758	0.040	0.907	288	0.049	0.870	-	ij	PTF15	-
PTF16	<b>PSD+DEPTH+PH_H2O+CEC</b>	1188	0.042	0.905	288	0.051	0.858	-	f	PTF16	-
PTF17	<b>PSD+DEPTH+OC+BD+CaCO3</b>	1588	0.031	0.944	288	0.042	0.904	-	n	PTF17	-
PTF18	<b>PSD+DEPTH+OC+BD+PH_H2O</b>	1655	0.033	0.943	288	0.043	0.900	-	l	PTF18	PTF22
PTF19	<b>PSD+DEPTH+OC+BD+CEC</b>	1284	0.033	0.934	288	0.044	0.892	-	lm	PTF19	-
PTF20	<b>PSD+DEPTH+OC+CaCO3+PH_H2O</b>	1201	0.033	0.943	288	0.048	0.874	-	f	PTF20	-
PTF21	<b>PSD+DEPTH+OC+CaCO3+CEC</b>	712	0.035	0.932	288	0.047	0.881	-	l	PTF21	-
PTF22	<b>PSD+DEPTH+OC+PH_H2O+CEC</b>	996	0.033	0.939	288	0.049	0.869	-	i	PTF22	-
PTF23	<b>PSD+DEPTH+BD+CaCO3+PH_H2O</b>	1263	0.032	0.948	288	0.044	0.895	-	lm	PTF23	-
PTF24	<b>PSD+DEPTH+BD+CaCO3+CEC</b>	757	0.033	0.939	288	0.041	0.906	-	o	PTF24	-
PTF25	<b>PSD+DEPTH+BD+PH_H2O+CEC</b>	1138	0.038	0.922	288	0.042	0.902	-	n	PTF25	-
PTF26	<b>PSD+DEPTH+CaCO3+PH_H2O+CEC</b>	717	0.037	0.924	288	0.047	0.878	-	jk	PTF26	-
PTF27	<b>PSD+DEPTH+OC+BD+CaCO3+PH_H2O</b>	1200	0.030	0.953	288	0.043	0.897	-	lm	PTF27	-
PTF28	<b>PSD+DEPTH+OC+BD+CaCO3+CEC</b>	711	0.032	0.941	288	0.041	0.906	-	o	PTF28	-
PTF29	<b>PSD+DEPTH+OC+BD+PH_H2O+CEC</b>	988	0.032	0.945	288	0.041	0.906	-	o	PTF29	-
PTF30	<b>PSD+DEPTH+OC+CaCO3+PH_H2O+CEC</b>	671	0.031	0.946	288	0.047	0.880	-	k	PTF30	PTF20
PTF31	<b>PSD+DEPTH+BD+CaCO3+PH_H2O+CEC</b>	716	0.031	0.948	288	0.042	0.904	-	o	PTF31	-
PTF32	<b>PSD+DEPTH+OC+BD+CaCO3+PH_H2O+CEC</b>	670	0.031	0.948	288	0.042	0.903	-	o	PTF32	PTF22

<sup>1</sup>PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ );  $\text{CaCO}_3$ : calcium carbonate content (mass %);  $\text{pH}_{\text{H}_2\text{O}}$ : pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

5

**Table 8.** Performance of pedotransfer functions (PTFs) by input combination on training and test datasets to predict parameters of the Mualem-van Genuchten model to describe soil moisture retention and hydraulic conductivity curve (MVG). N: number of samples, RMSE: root mean square error ( $\log_{10}$  (cm day<sup>-1</sup>)), and R<sup>2</sup>: determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in euptfv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>		Recom- mended PTF	Pair from euptfv1
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST_BASIC set	TEST_CHEM+ set		
<b>PTF01</b>	<b>PSD+DEPTH</b>	739	0.604	0.804	176	0.708	0.796	a	b	PTF01	-
<b>PTF02</b>	<b>PSD+DEPTH+OC</b>	407	0.619	0.829	176	0.676	0.814	b	jkl	PTF02	PTF19
PTF03	PSD+DEPTH+BD	726	0.568	0.824	176	0.688	0.808	a	ab	PTF01	-
<b>PTF04</b>	<b>PSD+DEPTH+CACO3</b>	273	0.587	0.878	57	0.644	0.863	-	ijk	PTF04	-
<b>PTF05</b>	<b>PSD+DEPTH+PH_H2O</b>	230	0.578	0.889	57	0.663	0.855	-	def	PTF05	-
<b>PTF06</b>	<b>PSD+DEPTH+CEC</b>	141	0.672	0.858	57	0.662	0.856	-	fghij	PTF06	-
PTF07	PSD+DEPTH+OC+BD	404	0.529	0.873	176	0.659	0.824	b	a	PTF02	PTF19
PTF08	PSD+DEPTH+OC+CACO3	250	0.587	0.880	57	0.699	0.839	-	b	PTF02	-
PTF09	PSD+DEPTH+OC+PH_H2O	224	0.597	0.882	57	0.686	0.845	-	fghi	PTF02	-
PTF10	PSD+DEPTH+OC+CEC	138	0.699	0.846	57	0.702	0.837	-	cde	PTF02	-
PTF11	PSD+DEPTH+BD+CACO3	272	0.542	0.895	57	0.637	0.866	-	defg	PTF04	-
<b>PTF12</b>	<b>PSD+DEPTH+BD+PH_H2O</b>	229	0.520	0.909	57	0.620	0.873	-	jklm	PTF12	-
<b>PTF13</b>	<b>PSD+DEPTH+BD+CEC</b>	140	0.644	0.866	57	0.637	0.866	-	lm	PTF13	-
PTF14	PSD+DEPTH+CACO3+PH_H2O	223	0.539	0.904	57	0.691	0.842	-	c	PTF04	-
PTF15	PSD+DEPTH+CACO3+CEC	136	0.735	0.830	57	0.684	0.846	-	c	PTF04	-
PTF16	PSD+DEPTH+PH_H2O+CEC	141	0.666	0.860	57	0.666	0.854	-	hijk	PTF06	-
PTF17	PSD+DEPTH+OC+BD+CACO3	249	0.526	0.902	57	0.662	0.855	-	ab	PTF02	-
PTF18	PSD+DEPTH+OC+BD+PH_H2O	223	0.553	0.897	57	0.642	0.864	-	klm	PTF02	PTF19
PTF19	PSD+DEPTH+OC+BD+CEC	137	0.619	0.876	57	0.676	0.849	-	b	PTF02	-
<b>PTF20</b>	<b>PSD+DEPTH+OC+CACO3+PH_H2O</b>	219	0.573	0.891	57	0.661	0.856	-	n	PTF20	-
<b>PTF21</b>	<b>PSD+DEPTH+OC+CACO3+CEC</b>	135	0.730	0.831	57	0.653	0.860	-	m	PTF21	-
PTF22	PSD+DEPTH+OC+PH_H2O+CEC	138	0.699	0.846	57	0.664	0.855	-	lm	PTF02	-
<b>PTF23</b>	<b>PSD+DEPTH+BD+CACO3+PH_H2O</b>	222	0.515	0.911	57	0.639	0.865	-	lm	PTF23	-
PTF24	PSD+DEPTH+BD+CACO3+CEC	135	0.678	0.852	57	0.656	0.858	-	c	PTF04	-
PTF25	PSD+DEPTH+BD+PH_H2O+CEC	140	0.595	0.885	57	0.646	0.862	-	ghijk	PTF12	-
PTF26	PSD+DEPTH+CACO3+PH_H2O+CEC	136	0.712	0.841	57	0.669	0.852	-	cd	PTF04	-
<b>PTF27</b>	<b>PSD+DEPTH+OC+BD+CACO3+PH_H2O</b>	218	0.524	0.907	57	0.606	0.879	-	o	PTF27	-
<b>PTF28</b>	<b>PSD+DEPTH+OC+BD+CACO3+CEC</b>	134	0.656	0.860	57	0.639	0.865	-	n	PTF28	-
<b>PTF29</b>	<b>PSD+DEPTH+OC+BD+PH_H2O+CEC</b>	137	0.646	0.865	57	0.638	0.866	-	n	PTF29	-
PTF30	PSD+DEPTH+OC+CACO3+PH_H2O+CEC	135	0.726	0.833	57	0.680	0.847	-	fghi	PTF20	PTF19
PTF31	PSD+DEPTH+BD+CACO3+PH_H2O+CEC	135	0.679	0.851	57	0.668	0.853	-	c	PTF12	-
PTF32	PSD+DEPTH+OC+BD+CACO3+PH_H2O+CEC	134	0.645	0.864	57	0.678	0.848	-	efgh	PTF27	PTF19

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

**Table 9.** The results of comparing the performance of parametric and point pedotransfer functions (PTFs) on the test sets of EU-HYDI to predict saturated water content (THS), water content at -100 cm matric potential head (FC\_2), water content at -330 cm matric potential head (FC), water content at wilting point (WP). Rows in italic indicate cases where there was no significant difference between the two PTFs.

Predicted soil hydraulic property	Available predictor variables <sup>1</sup>	Performance of parameter estimation (MRC with VG) <sup>2</sup>		Performance of point estimation		Number of samples in test dataset
		Recommended PTF number	RMSE	Recommended PTF number	RMSE	
<b>THS</b> (cm <sup>3</sup> cm <sup>-3</sup> )	<i>PSD+DEPTH_M+OC</i>	<i>PTF02<sup>a</sup></i>	<i>0.065</i>	<i>PTF02<sup>a</sup></i>	<i>0.061</i>	<i>216</i>
	<i>PSD+DEPTH_M+OC+BD</i>	<i>PTF07<sup>a</sup></i>	<i>0.041</i>	<i>PTF03<sup>b</sup></i>	<i>0.032</i>	<i>216</i>
	<i>PSD+DEPTH_M+OC+BD+PH_H2O</i>	<i>PTF12<sup>a</sup></i>	<i>0.028</i>	<i>PTF03<sup>b</sup></i>	<i>0.022</i>	<i>63</i>
	<i>PSD+DEPTH_M+OC+CACO3+PH_H2O+CEC</i>	<i>PTF21<sup>a</sup></i>	<i>0.051</i>	<i>PTF02<sup>a</sup></i>	<i>0.060</i>	<i>63</i>
	<i>PSD+DEPTH_M+OC+BD+CACO3+PH_H2O+CEC</i>	<i>PTF29<sup>a</sup></i>	<i>0.028</i>	<i>PTF03<sup>a</sup></i>	<i>0.022</i>	<i>63</i>
<b>FC_2</b> (cm <sup>3</sup> cm <sup>-3</sup> )	<i>PSD+DEPTH_M+OC</i>	<i>PTF02<sup>a</sup></i>	<i>0.057</i>	<i>PTF02<sup>b</sup></i>	<i>0.054</i>	<i>424</i>
	<i>PSD+DEPTH_M+OC+BD</i>	<i>PTF07<sup>a</sup></i>	<i>0.051</i>	<i>PTF03<sup>a</sup></i>	<i>0.051</i>	<i>424</i>
	<i>PSD+DEPTH_M+OC+BD+PH_H2O</i>	<i>PTF12<sup>a</sup></i>	<i>0.043</i>	<i>PTF07<sup>a</sup></i>	<i>0.049</i>	<i>68</i>
	<i>PSD+DEPTH_M+OC+CACO3+PH_H2O+CEC</i>	<i>PTF21<sup>a</sup></i>	<i>0.043</i>	<i>PTF09<sup>a</sup></i>	<i>0.047</i>	<i>68</i>
	<i>PSD+DEPTH_M+OC+BD+CACO3+PH_H2O+CEC</i>	<i>PTF29<sup>a</sup></i>	<i>0.036</i>	<i>PTF18<sup>a</sup></i>	<i>0.043</i>	<i>68</i>
<b>FC</b> (cm <sup>3</sup> cm <sup>-3</sup> )	<i>PSD+DEPTH_M+OC</i>	<i>PTF02<sup>a</sup></i>	<i>0.057</i>	<i>PTF02<sup>a</sup></i>	<i>0.048</i>	<i>319</i>
	<i>PSD+DEPTH_M+OC+BD</i>	<i>PTF07<sup>a</sup></i>	<i>0.056</i>	<i>PTF02<sup>a</sup></i>	<i>0.048</i>	<i>319</i>
	<i>PSD+DEPTH_M+OC+BD+PH_H2O</i>	<i>PTF12<sup>a</sup></i>	<i>0.047</i>	<i>PTF02<sup>a</sup></i>	<i>0.047</i>	<i>129</i>
	<i>PSD+DEPTH_M+OC+CACO3+PH_H2O+CEC</i>	<i>PTF21<sup>a</sup></i>	<i>0.046</i>	<i>PTF08<sup>a</sup></i>	<i>0.045</i>	<i>129</i>
	<i>PSD+DEPTH_M+OC+BD+CACO3+PH_H2O+CEC</i>	<i>PTF29<sup>a</sup></i>	<i>0.041</i>	<i>PTF07<sup>a</sup></i>	<i>0.046</i>	<i>129</i>
<b>WP</b> (cm <sup>3</sup> cm <sup>-3</sup> )	<i>PSD+DEPTH_M+OC</i>	<i>PTF02<sup>a</sup></i>	<i>0.064</i>	<i>PTF02<sup>b</sup></i>	<i>0.047</i>	<i>429</i>
	<i>PSD+DEPTH_M+OC+BD</i>	<i>PTF07<sup>a</sup></i>	<i>0.061</i>	<i>PTF02<sup>b</sup></i>	<i>0.047</i>	<i>429</i>
	<i>PSD+DEPTH_M+OC+BD+PH_H2O</i>	<i>PTF12<sup>a</sup></i>	<i>0.053</i>	<i>PTF07<sup>a</sup></i>	<i>0.045</i>	<i>91</i>
	<i>PSD+DEPTH_M+OC+CACO3+PH_H2O+CEC</i>	<i>PTF21<sup>a</sup></i>	<i>0.051</i>	<i>PTF07<sup>a</sup></i>	<i>0.045</i>	<i>91</i>
	<i>PSD+DEPTH_M+OC+BD+CACO3+PH_H2O+CEC</i>	<i>PTF29<sup>a</sup></i>	<i>0.054</i>	<i>PTF09<sup>a</sup></i>	<i>0.039</i>	<i>91</i>

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).  
<sup>2</sup>MRC: moisture retention curve; VG: parameters of the van Genuchten model. Different letters in a row indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter b is significantly better than the one noted with letter a. RMSE: root mean squared error.

formázott: Felső index

**Table 10.** The results of comparing the performance of euptfv1 and euptfv2 on the test sets of EU-HYDI to predict soil hydraulic properties. Rows in italic indicate cases where there was no significant difference between the two PTFs.

Predicted soil hydraulic property <sup>1</sup>	Performance <sup>2</sup>					Number of samples in test datasets
	euptfv1		euptfv2		Name of test set	
	Name of PTF	RMSE	Name of PTF	RMSE		
<b>THS</b> (cm <sup>3</sup> cm <sup>-3</sup> )	PTF04 <sup>a</sup>	0.063	PTF02 <sup>b</sup>	0.056	TEST_BASIC	1274
	PTF05 <sup>a</sup>	0.034	PTF03 <sup>b</sup>	0.031	TEST_BASIC	1274
	<i>PTF06<sup>a</sup></i>	<i>0.020</i>	<i>PTF03<sup>a</sup></i>	<i>0.024</i>	<i>TEST_CHEM+</i>	<i>156</i>
<b>FC</b> (cm <sup>3</sup> cm <sup>-3</sup> )	PTF09 <sup>a</sup>	0.054	PTF02 <sup>b</sup>	0.050	TEST_BASIC	801
	PTF09 <sup>a</sup>	0.054	PTF07 <sup>b</sup>	0.048	TEST_BASIC	801
	PTF09 <sup>a</sup>	0.058	PTF08 <sup>b</sup>	0.053	TEST_CHEM+	280
<b>WP</b> (cm <sup>3</sup> cm <sup>-3</sup> )	PTF12 <sup>a</sup>	0.048	PTF02 <sup>b</sup>	0.046	TEST_BASIC	2088
	PTF12 <sup>a</sup>	0.048	PTF07 <sup>b</sup>	0.044	TEST_BASIC	2088
	<i>PTF12<sup>a</sup></i>	<i>0.043</i>	<i>PTF09<sup>a</sup></i>	<i>0.041</i>	<i>TEST_CHEM+</i>	<i>294</i>
<b>KS</b> (log <sub>10</sub> cm day <sup>-1</sup> )	PTF16 <sup>a</sup>	1.06	PTF02 <sup>b</sup>	0.95	TEST_BASIC	1117
	<i>PTF17<sup>a</sup></i>	<i>1.00</i>	<i>PTF02<sup>a</sup></i>	<i>0.91</i>	<i>TEST_CHEM+</i>	<i>169</i>
<b>VG</b> (cm <sup>3</sup> cm <sup>-3</sup> )	PTF19 <sup>a</sup>	0.068	PTF02 <sup>b</sup>	0.060	TEST_BASIC	1591
	PTF21 <sup>a</sup>	0.064	PTF07 <sup>b</sup>	0.054	TEST_BASIC	1591
	PTF22 <sup>a</sup>	0.046	PTF12 <sup>b</sup>	0.044	TEST_CHEM+	288
	PTF20 <sup>a</sup>	0.054	PTF21 <sup>b</sup>	0.047	TEST_CHEM+	288
	PTF22 <sup>a</sup>	0.046	PTF29 <sup>b</sup>	0.041	TEST_CHEM+	288
<b>MVG</b> (log <sub>10</sub> cm day <sup>-1</sup> )	PTF19 <sup>a</sup>	0.77	PTF02 <sup>b</sup>	0.68	TEST_BASIC	176
	<i>PTF19<sup>a</sup></i>	<i>0.66</i>	<i>PTF20<sup>a</sup></i>	<i>0.66</i>	<i>TEST_CHEM+</i>	<i>57</i>
	<i>PTF19<sup>a</sup></i>	<i>0.66</i>	<i>PTF27<sup>a</sup></i>	<i>0.61</i>	<i>TEST_CHEM+</i>	<i>57</i>

<sup>1</sup>THS: saturated water content (pF 0); FC\_2: water content at -100 cm matric potential head (pF 2.0); FC: water content at -330 cm matric potential head (pF 2.5); WP: water content at wilting point (pF 4.2); KS: saturated hydraulic conductivity; VG: parameters of the van Genuchten model; MVG: parameters of the Mualem – van Genuchten model.

<sup>2</sup>Different letters in a row indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error;for example performance indicated with the letter b is significantly better than the one noted with letter a. RMSE: root mean squared error; TEST\_BASIC: sa mples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CaCO3, PH\_H2O and CEC; N: number of samples.

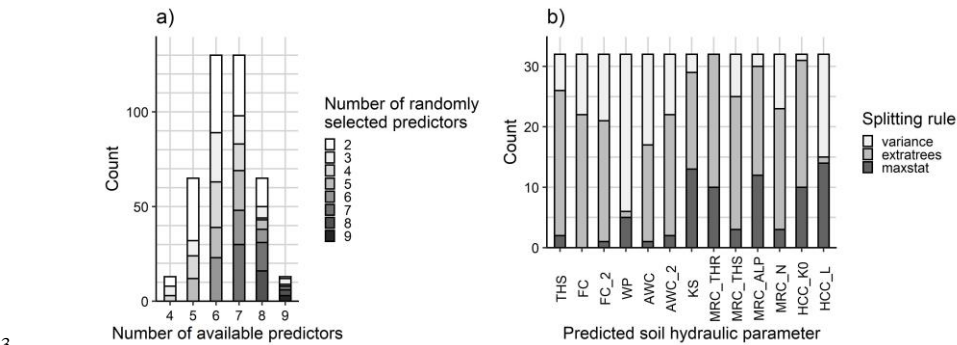


**Table 11.** List of recommended pedotransfer functions (PTFs) by predicted soil hydraulic property and available predictor variables.

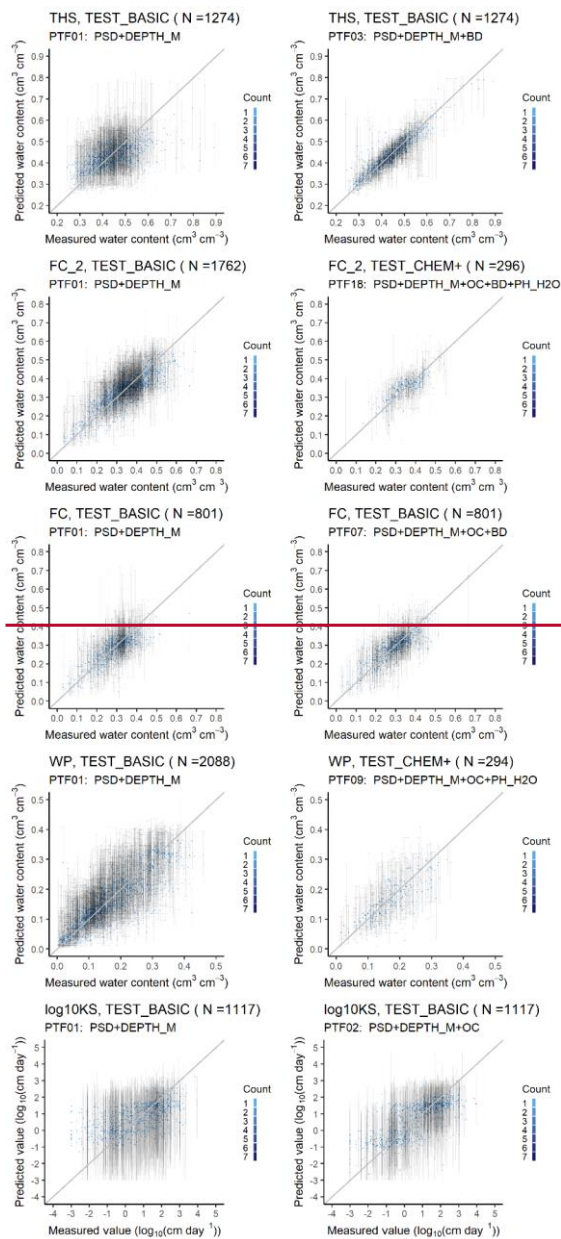
Predictor variables <sup>1</sup>	Recommended PTFs <sup>2</sup>								
	THS	FC 2	FC	WP	AWC	AWC	KS	VG	MVG
PSD+DEPTH_M	PTF01	PTF01	PTF01	PTF01	PTF01	PTF01	PTF01	PTF01	PTF01
PSD+DEPTH_M+OC	PTF02	PTF02	PTF02	PTF02	PTF02	PTF01	PTF02	PTF02	PTF02
PSD+DEPTH_M+BD	PTF03	PTF03	PTF01	PTF01	PTF03	PTF01	PTF01	PTF03	PTF01
PSD+DEPTH_M+CACO3	PTF04	PTF01	PTF01	PTF01	PTF01	PTF01	PTF01	PTF04	PTF04
PSD+DEPTH_M+PH_H2O	PTF05	PTF01	PTF01	PTF01	PTF01	PTF01	PTF01	PTF05	PTF05
PSD+DEPTH_M+CEC	PTF01	PTF01	PTF06	PTF01	PTF01	PTF01	PTF01	PTF06	PTF06
PSD+DEPTH_M+OC+BD	PTF03	PTF03	PTF02	PTF02	PTF03	PTF03	PTF02	PTF07	PTF02
PSD+DEPTH_M+OC+CACO3	PTF02	PTF02	PTF02	PTF02	PTF02	PTF01	PTF02	PTF08	PTF02
PSD+DEPTH_M+OC+PH_H2O	PTF02	PTF02	PTF02	PTF02	PTF02	PTF01	PTF02	PTF09	PTF02
PSD+DEPTH_M+OC+CEC	PTF02	PTF06	PTF02	PTF02	PTF02	PTF01	PTF02	PTF10	PTF02
PSD+DEPTH_M+BD+CACO3	PTF03	PTF03	PTF03	PTF01	PTF03	PTF01	PTF01	PTF11	PTF04
PSD+DEPTH_M+BD+PH_H2O	PTF03	PTF03	PTF03	PTF01	PTF03	PTF01	PTF05	PTF12	PTF12
PSD+DEPTH_M+BD+CEC	PTF03	PTF13	PTF06	PTF01	PTF03	PTF01	PTF01	PTF13	PTF13
PSD+DEPTH_M+CACO3+PH_H2O	PTF05	PTF05	PTF04	PTF01	PTF01	PTF01	PTF05	PTF14	PTF04
PSD+DEPTH_M+CACO3+CEC	PTF04	PTF04	PTF04	PTF01	PTF01	PTF01	PTF01	PTF15	PTF04
PSD+DEPTH_M+PH_H2O+CEC	PTF05	PTF05	PTF06	PTF01	PTF01	PTF01	PTF01	PTF16	PTF06
PSD+DEPTH_M+OC+BD+CACO3	PTF03	PTF07	PTF02	PTF07	PTF03	PTF03	PTF02	PTF11	PTF02
PSD+DEPTH_M+OC+BD+PH_H2O	PTF03	PTF07	PTF02	PTF07	PTF03	PTF03	PTF02	PTF12	PTF02
PSD+DEPTH_M+OC+BD+CEC	PTF03	PTF13	PTF07	PTF07	PTF03	PTF03	PTF02	PTF13	PTF02
PSD+DEPTH_M+OC+CACO3+PH_H2O	PTF02	PTF02	PTF02	PTF02	PTF02	PTF01	PTF02	PTF09	PTF20
PSD+DEPTH_M+OC+CACO3+CEC	PTF02	PTF08	PTF04	PTF08	PTF02	PTF01	PTF02	PTF21	PTF21
PSD+DEPTH_M+OC+PH_H2O+CEC	PTF02	PTF09	PTF09	PTF02	PTF02	PTF01	PTF02	PTF22	PTF02
PSD+DEPTH_M+BD+CACO3+PH_H2O	PTF03	PTF03	PTF03	PTF01	PTF03	PTF01	PTF05	PTF11	PTF23
PSD+DEPTH_M+BD+CACO3+CEC	PTF03	PTF11	PTF11	PTF01	PTF03	PTF01	PTF01	PTF24	PTF04
PSD+DEPTH_M+BD+PH_H2O+CEC	PTF03	PTF12	PTF11	PTF01	PTF03	PTF01	PTF05	PTF25	PTF12
PSD+DEPTH_M+CACO3+PH_H2O+CEC	PTF05	PTF05	PTF14	PTF05	PTF01	PTF01	PTF05	PTF15	PTF04
PSD+DEPTH_M+OC+BD+CACO3+PH_H2O	PTF03	PTF07	PTF02	PTF07	PTF03	PTF03	PTF02	PTF11	PTF27
PSD+DEPTH_M+OC+BD+CACO3+CEC	PTF03	PTF11	PTF07	PTF17	PTF03	PTF03	PTF02	PTF24	PTF28
PSD+DEPTH_M+OC+BD+PH_H2O+CEC	PTF03	PTF12	PTF07	PTF07	PTF03	PTF03	PTF02	PTF29	PTF29
PSD+DEPTH_M+OC+CACO3+PH_H2O+CE	PTF02	PTF09	PTF08	PTF02	PTF02	PTF01	PTF02	PTF21	PTF20
PSD+DEPTH_M+BD+CACO3+PH_H2O+CE	PTF03	PTF11	PTF06	PTF05	PTF03	PTF01	PTF05	PTF24	PTF12
PSD+DEPTH_M+OC+BD+CACO3+PH_H2O	PTF03	PTF18	PTF07	PTF09	PTF03	PTF03	PTF02	PTF29	PTF27

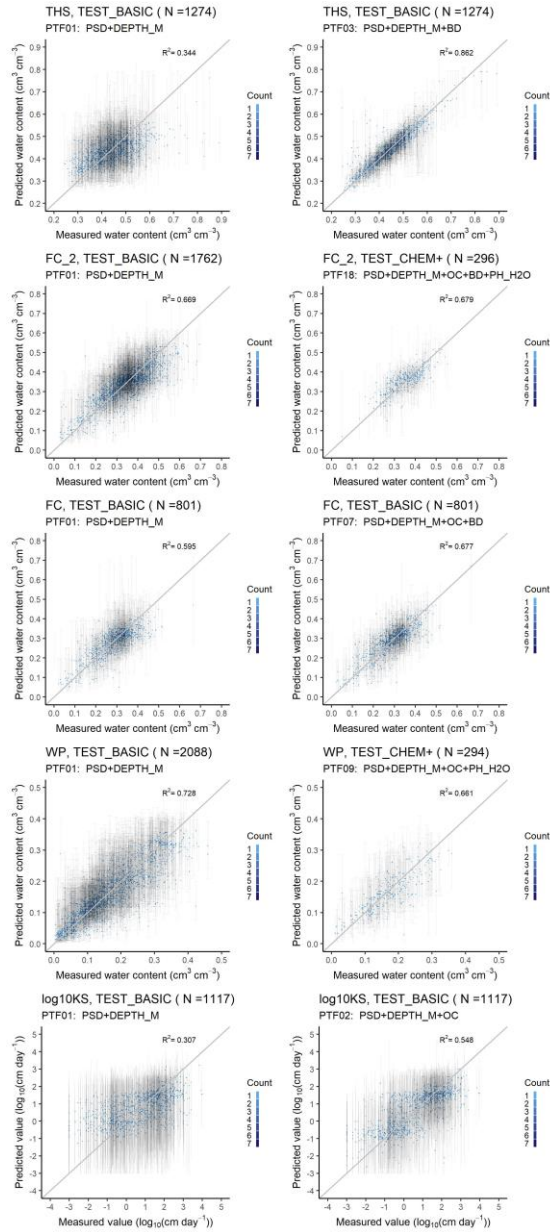
<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).  
<sup>2</sup>THS: saturated water content (pF 0); FC 2: water content at -100 cm matric potential head (pF 2.0); FC: water content at -330 cm matric potential head (pF 2.5); AWC 2: plant available water content based on FC 2; AWC: plant available water content based on FC; WP: water content at wilting point (pF 4.2); KS: saturated hydraulic conductivity; VG: parameters of the van Genuchten model; MVG: parameters of the Mualem – van Genuchten model; TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC.

1 **FIGURES**  
2

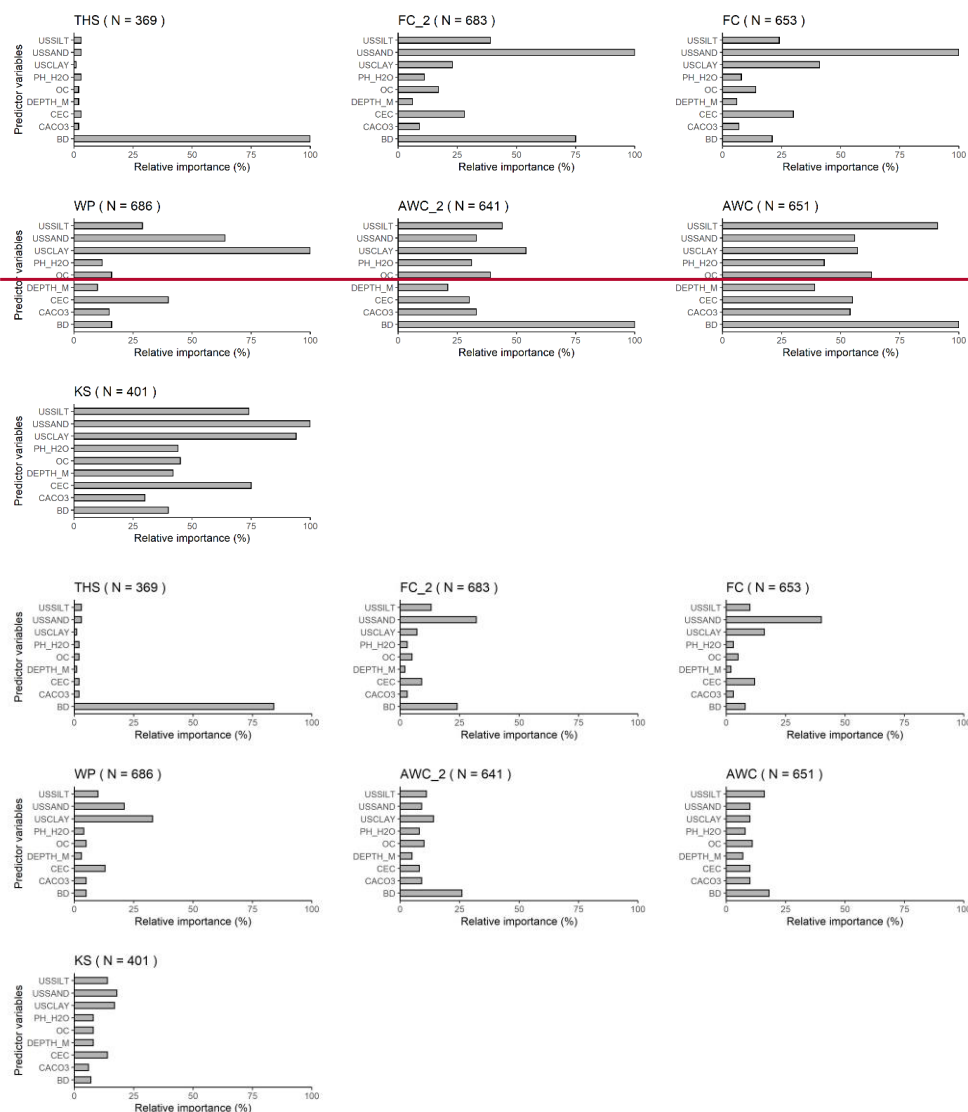


6 **Figure 1.** Results of parameter tuning of the random forest: optimization of a) the number of randomly selected predictors at  
7 each split by number of available predictors and b) splitting rule applied to build the trees in the random forest.

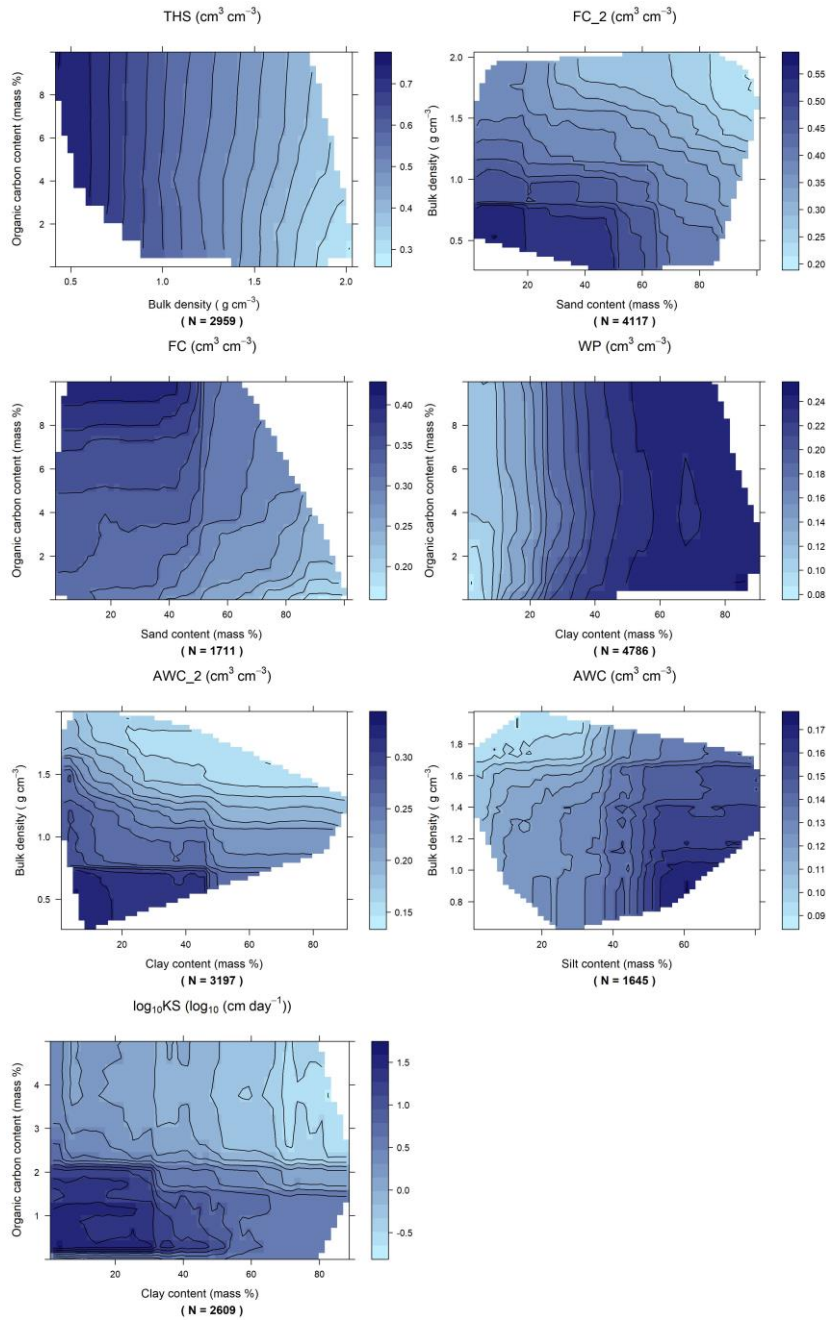




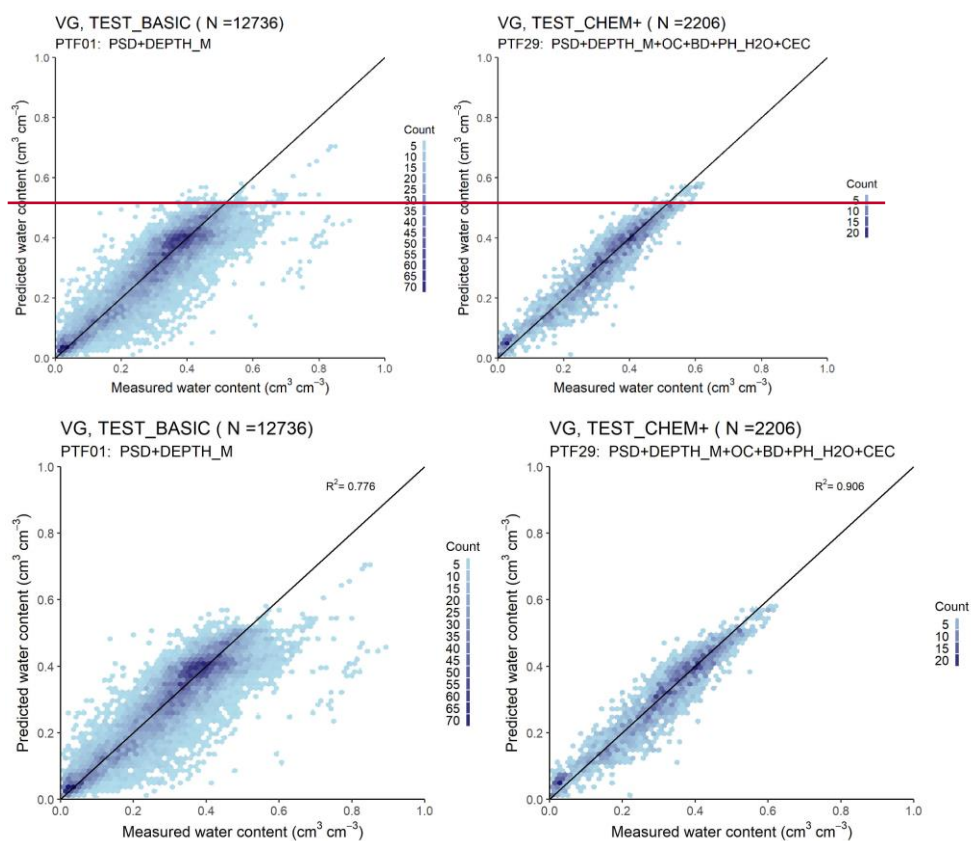
**Figure 2.** Scatter plot of the measured versus median predicted water retention values of the worst and best performing PTF with 90% prediction interval on test datasets. THS: saturated water content (PTF01 vs. PTF03); FC\_2: water content at -100 cm matric potential head (PTF01 vs. PTF18); FC: water content at -330 cm matric potential head (PTF01 vs. PTF07); WP: water content at wilting point (PTF01 vs. PTF09); log10KS: saturated hydraulic conductivity (PTF01 vs. PTF02); PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); PH\_H2O: pH in water (-); **Count:** the number of cases in each rectangle.



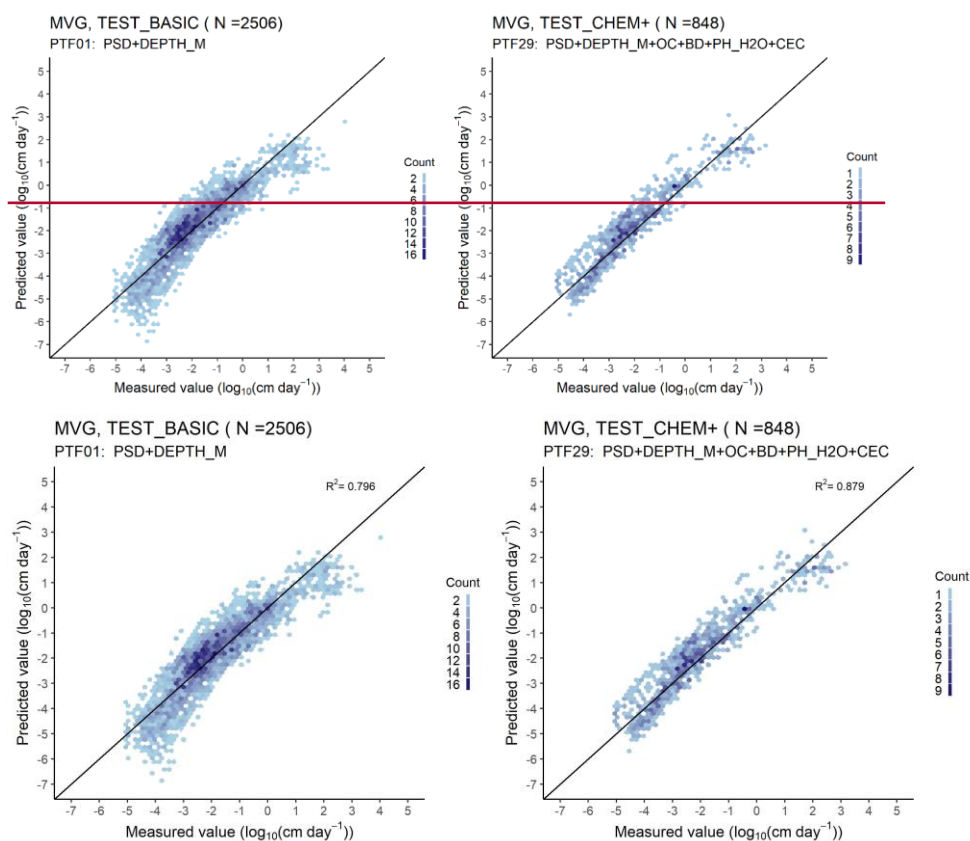
**Figure 3.** Relative variable importance computed with the random forest algorithm for the prediction of water content with PTF32 at saturation (THS), at field capacity; -100 (FC\_2) and -330 (FC) matric potential head, at wilting point (WP), of the plant available water content based on FC\_2 (AWC\_2) and FC (AWC), and the saturated hydraulic conductivity (KS). USSILT: silt content (2–50  $\mu\text{m}$  (mass %)); USSAND: sand content (50–2000  $\mu\text{m}$  (mass %)); USCLAY: clay content ( $<2$   $\mu\text{m}$  (mass %)); PH\_H2O: pH in water (-); OC: organic carbon content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); CEC: cation exchange capacity ( $\text{cmol (+)} \text{kg}^{-1}$ ); CACO3: calcium carbonate content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ).



**Figure 4.** Partial dependence plot computed based on the random forest algorithm (PTF07) for the prediction of water content at saturation (THS), field capacity at -100 (FC\_2) and -330 (FC) matric potential head, wilting point (WP), plant available water content computed with field capacity at -100 and -330 cm matric potential head (AWC\_2, AWC) and saturated hydraulic conductivity (KS) for selected predictors.

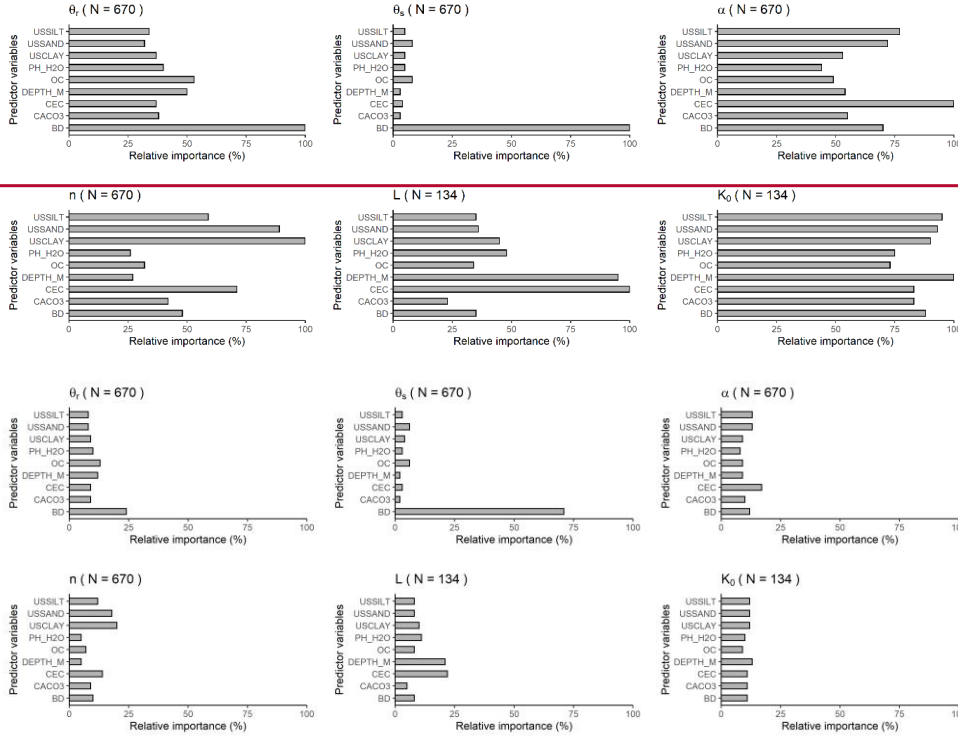


**Figure 5.** Scatter plot of the measured versus median predicted water retention values computed with the van Genuchten (VG) model (PTF01 vs. PTF29, i.e. the worst versus best performing PTF). PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ); Count: the number of cases in each hexagon.



**Figure 6.** Scatter plot of the measured versus median predicted hydraulic conductivity values computed with the Mualem-van Genuchten (MVG) model (PTF01 vs. PTF27, i.e. the worst versus best performing PTF). PSD: particle size distribution (sand, 50–2000  $\mu\text{m}$ ; silt, 2–50  $\mu\text{m}$ ; clay, <2  $\mu\text{m}$  (mass %)); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); Count: the number of cases in each hexagon.





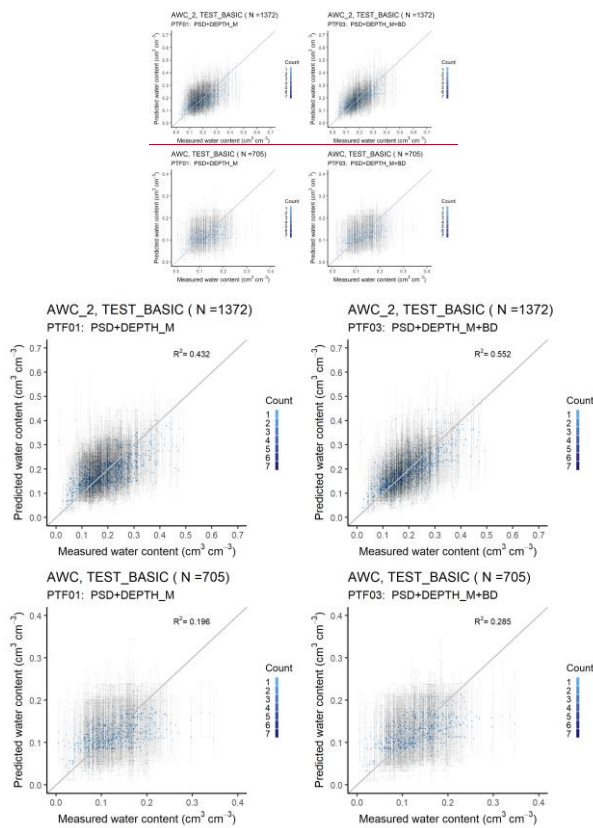
**Figure 7. Relative variable importance computed with the random forest algorithm for the prediction of parameters of the van Genuchten and Mualem-van Genuchten models based on PTF32.  $\theta_r$ : residual water content ( $\text{cm}^3 \text{cm}^{-3}$ );  $\theta_s$ : saturated water content ( $\text{cm}^3 \text{cm}^{-3}$ );  $\alpha$  ( $\text{cm}^{-1}$ ),  $n$  (-): fitting-shape parameters;  $K_0$ : the hydraulic conductivity acting as a matching point at saturation ( $\text{cm day}^{-1}$ );  $L$ : shape parameter related to pore tortuosity (-); USSILT: silt content (2–50  $\mu\text{m}$  (mass %)); USSAND: sand content (50–2000  $\mu\text{m}$  (mass %)); USCLAY: clay content, ( $<2 \mu\text{m}$  (mass %)); PH\_H2O: pH in water (-); OC: organic carbon content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ); CACO3: calcium carbonate content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ).**

*Supplement of*

## **Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (euptfv2)**

5 **Brigitta Szabó et al.**

\* Correspondence: [toth.brigitta@agrar.mta.hu](mailto:toth.brigitta@agrar.mta.hu) (B.Sz.)



**Figure S1.** The scatter plot of the measured versus predicted plant available water content values of the worst and best performing PTF with 90% prediction interval on test datasets. AWC<sub>2</sub>: plant available water content based on filed capacity at -100 cm matric potential head (PTF01 vs. PTF03); AWC: plant available water content based on filed capacity at -330 cm matric potential head (PTF01 vs. PTF03); PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH\_M: mean soil depth (cm); BD: bulk density (g cm<sup>-3</sup>); **Count**: the number of cases in each rectangle.

**Table S1.** Performance of pedotransfer functions (PTF) by input combination on training and test datasets to predict the plant available water content of the soil (AWC<sub>2</sub>) belonging to the -100 cm matric potential head. N: number of samples, RMSE: root mean square error (cm<sup>3</sup> cm<sup>-3</sup>), and R<sup>2</sup>: determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH<sub>2</sub>O and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in eupftv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>			Recommended PTF
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST- BASIC set	TEST- CHEM + set		
PTF01	PSD+DEPTH	3528	0.062	0.446	1372	0.060	0.432	a	ab	PTF01	
PTF02	PSD+DEPTH+OC	3208	0.055	0.540	1372	0.054	0.544	b	abcd	PTF02	
PTF03	PSD+DEPTH+BD	3472	0.054	0.581	1372	0.053	0.552	b	abcd	PTF03	
PTF04	PSD+DEPTH+CACO3	1548	0.050	0.326	274	0.055	0.219	-	abcd	PTF01	
PTF05	PSD+DEPTH+PH <sub>2</sub> O	1849	0.058	0.463	274	0.055	0.216	-	a	PTF01	
PTF06	PSD+DEPTH+CEC	1550	0.059	0.512	274	0.060	0.050	-	abcd	PTF01	
PTF07	PSD+DEPTH+OC+BD	3197	0.051	0.609	1372	0.051	0.588	b	abcd	PTF03	
PTF08	PSD+DEPTH+OC+CACO3	1464	0.048	0.353	274	0.053	0.257	-	abcd	PTF02	
PTF09	PSD+DEPTH+OC+PH <sub>2</sub> O	1615	0.055	0.490	274	0.053	0.270	-	abc	PTF02	
PTF10	PSD+DEPTH+OC+CEC	1358	0.054	0.563	274	0.053	0.278	-	abcd	PTF02	
PTF11	PSD+DEPTH+BD+CACO3	1545	0.044	0.470	274	0.048	0.396	-	d	PTF03	
PTF12	PSD+DEPTH+BD+PH <sub>2</sub> O	1796	0.052	0.565	274	0.048	0.406	-	abcd	PTF03	
PTF13	PSD+DEPTH+BD+CEC	1498	0.053	0.598	274	0.048	0.398	-	abcd	PTF03	
PTF14	PSD+DEPTH+CACO3+PH <sub>2</sub> O	1195	0.051	0.341	274	0.052	0.284	-	abcd	PTF01	
PTF15	PSD+DEPTH+CACO3+CEC	726	0.050	0.286	274	0.052	0.303	-	abcd	PTF01	
PTF16	PSD+DEPTH+PH <sub>2</sub> O+CEC	1255	0.058	0.539	274	0.051	0.331	-	abcd	PTF01	
PTF17	PSD+DEPTH+OC+BD+CACO3	1464	0.044	0.465	274	0.048	0.390	-	bcd	PTF03	
PTF18	PSD+DEPTH+OC+BD+PH <sub>2</sub> O	1607	0.051	0.556	274	0.048	0.407	-	abcd	PTF03	
PTF19	PSD+DEPTH+OC+BD+CEC	1349	0.052	0.593	274	0.046	0.441	-	abcd	PTF03	
PTF20	PSD+DEPTH+OC+CACO3+PH <sub>2</sub> O	1130	0.050	0.367	274	0.051	0.309	-	abcd	PTF02	
PTF21	PSD+DEPTH+OC+CACO3+CEC	683	0.049	0.305	274	0.050	0.359	-	abcd	PTF02	
PTF22	PSD+DEPTH+OC+PH <sub>2</sub> O+CEC	1067	0.054	0.561	274	0.049	0.367	-	abcd	PTF02	
PTF23	PSD+DEPTH+BD+CACO3+PH <sub>2</sub> O	1192	0.046	0.471	274	0.049	0.375	-	bcd	PTF03	
PTF24	PSD+DEPTH+BD+CACO3+CEC	725	0.045	0.420	274	0.046	0.444	-	d	PTF03	
PTF25	PSD+DEPTH+BD+PH <sub>2</sub> O+CEC	1204	0.052	0.621	274	0.046	0.456	-	abcd	PTF03	
PTF26	PSD+DEPTH+CACO3+PH <sub>2</sub> O+CEC	684	0.049	0.318	274	0.048	0.388	-	abcd	PTF01	
PTF27	PSD+DEPTH+OC+BD+CACO3+PH <sub>2</sub> O	1130	0.045	0.475	274	0.049	0.367	-	abcd	PTF03	
PTF28	PSD+DEPTH+OC+BD+CACO3+CEC	683	0.045	0.408	274	0.045	0.466	-	bcd	PTF03	
PTF29	PSD+DEPTH+OC+BD+PH <sub>2</sub> O+CEC	1059	0.052	0.603	274	0.045	0.473	-	bcd	PTF03	
PTF30	PSD+DEPTH+OC+CACO3+PH <sub>2</sub> O+CEC	641	0.049	0.330	274	0.048	0.393	-	abcd	PTF02	
PTF31	PSD+DEPTH+BD+CACO3+PH <sub>2</sub> O+CEC	683	0.044	0.450	274	0.045	0.480	-	cd	PTF03	
PTF32	PSD+DEPTH+OC+BD+CACO3+PH <sub>2</sub> O+CEC	641	0.045	0.425	274	0.045	0.471	-	cd	PTF03	

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH<sub>2</sub>O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

**Table S2.** Performance of pedotransfer functions (PTF) by input combination on training and test datasets to predict the plant available water content of the soil (AWC) belonging to the -330 cm matric potential head. N: number of samples, RMSE: root mean square error (cm<sup>3</sup> cm<sup>-3</sup>), and R<sup>2</sup>: determination coefficient, TEST\_BASIC: samples with measured PSD, DEPTH, OC and BD; TEST\_CHEM+: samples with measured PSD, DEPTH, OC, BD, CACO3, PH\_H2O and CEC. Recommended PTFs are highlighted in bold.

Name of PTF in eupftv2	Predictor variables <sup>1</sup>	Training set			Test set			Sign. difference <sup>2</sup>			Recommended PTF
		N	RMSE	R <sup>2</sup>	N	RMSE	R <sup>2</sup>	TEST- BASIC set	TEST- CHEM + set		
PTF01	PSD+DEPTH	1863	0.042	0.312	705	0.048	0.196	a	a	PTF01	
PTF02	PSD+DEPTH+OC	1650	0.041	0.337	705	0.045	0.288	ab	a	PTF01	
PTF03	PSD+DEPTH+BD	1849	0.040	0.374	705	0.045	0.285	ab	a	PTF01	
PTF04	PSD+DEPTH+CACO3	1531	0.040	0.366	279	0.050	0.199	-	a	PTF01	
PTF05	PSD+DEPTH+PH_H2O	1245	0.042	0.344	279	0.048	0.238	-	a	PTF01	
PTF06	PSD+DEPTH+CEC	1092	0.041	0.356	279	0.053	0.078	-	a	PTF01	
PTF07	PSD+DEPTH+OC+BD	1645	0.040	0.381	705	0.043	0.337	b	a	PTF03	
PTF08	PSD+DEPTH+OC+CACO3	1336	0.041	0.345	279	0.049	0.219	-	a	PTF01	
PTF09	PSD+DEPTH+OC+PH_H2O	1074	0.042	0.345	279	0.048	0.242	-	a	PTF01	
PTF10	PSD+DEPTH+OC+CEC	998	0.039	0.413	279	0.051	0.147	-	a	PTF01	
PTF11	PSD+DEPTH+BD+CACO3	1522	0.038	0.428	279	0.048	0.258	-	a	PTF01	
PTF12	PSD+DEPTH+BD+PH_H2O	1236	0.039	0.429	279	0.047	0.287	-	a	PTF01	
PTF13	PSD+DEPTH+BD+CEC	1088	0.038	0.429	279	0.049	0.231	-	a	PTF01	
PTF14	PSD+DEPTH+CACO3+PH_H2O	1230	0.041	0.376	279	0.047	0.263	-	a	PTF01	
PTF15	PSD+DEPTH+CACO3+CEC	791	0.041	0.366	279	0.049	0.214	-	a	PTF01	
PTF16	PSD+DEPTH+PH_H2O+CEC	739	0.042	0.321	279	0.048	0.237	-	a	PTF01	
PTF17	PSD+DEPTH+OC+BD+CACO3	1334	0.039	0.399	279	0.048	0.262	-	a	PTF03	
PTF18	PSD+DEPTH+OC+BD+PH_H2O	1072	0.040	0.393	279	0.047	0.293	-	a	PTF03	
PTF19	PSD+DEPTH+OC+BD+CEC	995	0.038	0.432	279	0.049	0.223	-	a	PTF03	
PTF20	PSD+DEPTH+OC+CACO3+PH_H2O	1059	0.042	0.362	279	0.047	0.289	-	a	PTF01	
PTF21	PSD+DEPTH+OC+CACO3+CEC	707	0.041	0.358	279	0.049	0.229	-	a	PTF01	
PTF22	PSD+DEPTH+OC+PH_H2O+CEC	660	0.041	0.339	279	0.048	0.253	-	a	PTF01	
PTF23	PSD+DEPTH+BD+CACO3+PH_H2O	1221	0.039	0.442	279	0.047	0.267	-	a	PTF01	
PTF24	PSD+DEPTH+BD+CACO3+CEC	788	0.039	0.405	279	0.047	0.269	-	a	PTF01	
PTF25	PSD+DEPTH+BD+PH_H2O+CEC	736	0.039	0.402	279	0.046	0.307	-	a	PTF01	
PTF26	PSD+DEPTH+CACO3+PH_H2O+CEC	732	0.040	0.405	279	0.048	0.254	-	a	PTF01	
PTF27	PSD+DEPTH+OC+BD+CACO3+PH_H2O	1057	0.040	0.415	279	0.046	0.312	-	a	PTF03	
PTF28	PSD+DEPTH+OC+BD+CACO3+CEC	705	0.040	0.383	279	0.047	0.277	-	a	PTF03	
PTF29	PSD+DEPTH+OC+BD+PH_H2O+CEC	658	0.040	0.385	279	0.046	0.315	-	a	PTF03	
PTF30	PSD+DEPTH+OC+CACO3+PH_H2O+CEC	653	0.040	0.395	279	0.047	0.274	-	a	PTF01	
PTF31	PSD+DEPTH+BD+CACO3+PH_H2O+CEC	729	0.039	0.431	279	0.047	0.290	-	a	PTF01	
PTF32	PSD+DEPTH+OC+BD+CACO3+PH_H2O+CEC	651	0.039	0.403	279	0.046	0.307	-	a	PTF03	

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

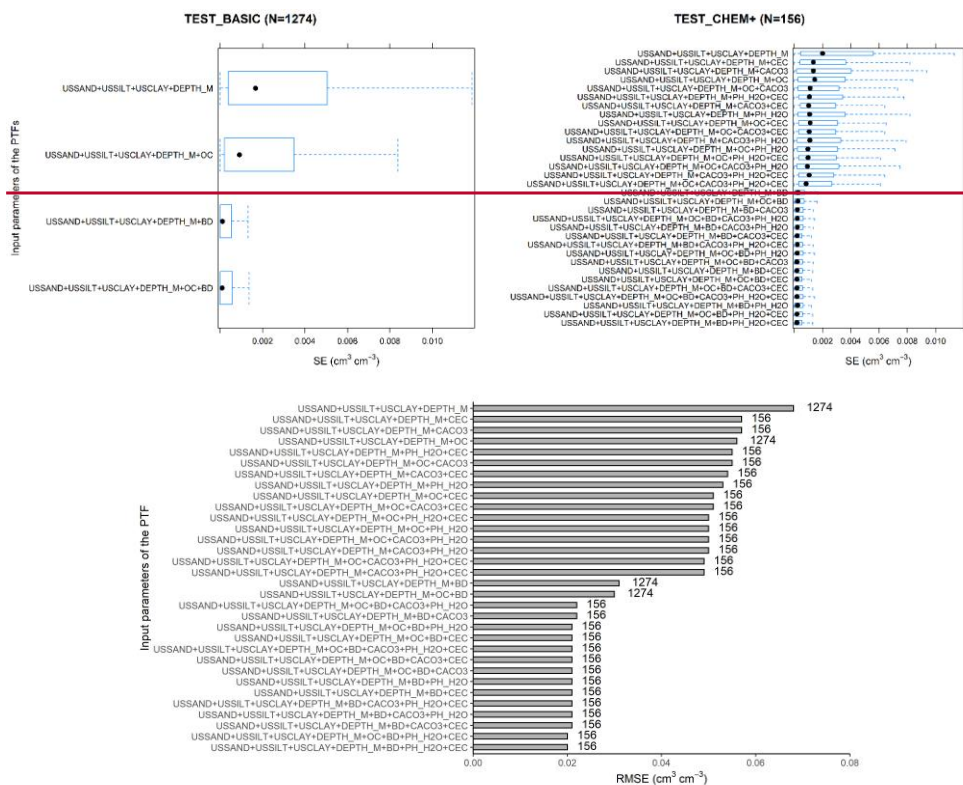
<sup>2</sup>Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error;for example performance indicated with the letter c is significantly better than the one noted with letters b and a.

**Table S3.** Normalized root mean square error (NRMSE) of the point predictions by soil hydraulic properties computed on the test datasets in cm<sup>3</sup> cm<sup>-3</sup> for water retention and log<sub>10</sub>(cm day<sup>-1</sup>) for saturated hydraulic conductivity. In case of PTF01, 02, 03 and 07 TEST\_BASIC set was used for the analysis, for the rest of the PTFs TEST\_CHEM+ set was considered.

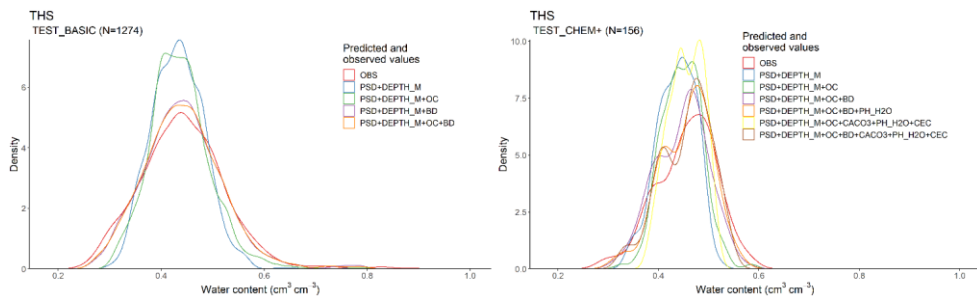
Name of PTF in eupfv2	Predictor variables <sup>1</sup>	NRMSE in test sets <sup>2</sup>						
		THS	FC 2	FC	WP	AWC 2	AWC	KS
PTF01	PSD+DEPTH_M	0.104	0.090	0.082	0.105	0.126	0.140	0.17
PTF02	PSD+DEPTH_M+OC	0.086	0.083	0.076	0.102	0.112	0.132	0.14
PTF03	PSD+DEPTH_M+BD	0.048	0.079	0.074	0.100	0.111	0.132	0.17
PTF04	PSD+DEPTH_M+CACO3	0.191	0.107	0.113	0.122	0.164	0.145	0.19
PTF05	PSD+DEPTH_M+PH_H2O	0.176	0.112	0.114	0.126	0.164	0.142	0.19
PTF06	PSD+DEPTH_M+CEC	0.191	0.107	0.107	0.118	0.181	0.156	0.19
PTF07	PSD+DEPTH_M+OC+BD	0.047	0.075	0.073	0.097	0.107	0.127	0.14
PTF08	PSD+DEPTH_M+OC+CACO3	0.184	0.097	0.109	0.117	0.160	0.143	0.19
PTF09	PSD+DEPTH_M+OC+PH_H2O	0.167	0.095	0.107	0.119	0.158	0.141	0.18
PTF10	PSD+DEPTH_M+OC+CEC	0.172	0.098	0.108	0.116	0.158	0.150	0.18
PTF11	PSD+DEPTH_M+BD+CACO3	0.072	0.091	0.105	0.115	0.144	0.140	0.19
PTF12	PSD+DEPTH_M+BD+PH_H2O	0.069	0.086	0.103	0.117	0.143	0.137	0.19
PTF13	PSD+DEPTH_M+BD+CEC	0.070	0.091	0.100	0.115	0.144	0.142	0.19
PTF14	PSD+DEPTH_M+CACO3+PH_H2O	0.168	0.101	0.109	0.121	0.157	0.139	0.18
PTF15	PSD+DEPTH_M+CACO3+CEC	0.179	0.102	0.106	0.113	0.155	0.144	0.19
PTF16	PSD+DEPTH_M+PH_H2O+CEC	0.183	0.098	0.104	0.115	0.152	0.142	0.19
PTF17	PSD+DEPTH_M+OC+BD+CACO3	0.070	0.089	0.102	0.111	0.145	0.139	0.18
PTF18	PSD+DEPTH_M+OC+BD+PH_H2O	0.070	0.083	0.103	0.116	0.143	0.136	0.18
PTF19	PSD+DEPTH_M+OC+BD+CEC	0.070	0.087	0.099	0.113	0.139	0.143	0.18
PTF20	PSD+DEPTH_M+OC+CACO3+PH_H2O	0.166	0.105	0.107	0.114	0.154	0.137	0.18
PTF21	PSD+DEPTH_M+OC+CACO3+CEC	0.171	0.090	0.104	0.108	0.149	0.142	0.18
PTF22	PSD+DEPTH_M+OC+PH_H2O+CEC	0.166	0.089	0.102	0.111	0.148	0.140	0.18
PTF23	PSD+DEPTH_M+BD+CACO3+PH_H2O	0.071	0.089	0.104	0.116	0.147	0.139	0.18
PTF24	PSD+DEPTH_M+BD+CACO3+CEC	0.071	0.085	0.099	0.110	0.138	0.139	0.19
PTF25	PSD+DEPTH_M+BD+PH_H2O+CEC	0.067	0.084	0.100	0.112	0.137	0.135	0.19
PTF26	PSD+DEPTH_M+CACO3+PH_H2O+CEC	0.163	0.094	0.103	0.111	0.145	0.140	0.18
PTF27	PSD+DEPTH_M+OC+BD+CACO3+PH_H2O	0.072	0.086	0.101	0.111	0.148	0.135	0.18
PTF28	PSD+DEPTH_M+OC+BD+CACO3+CEC	0.070	0.082	0.098	0.106	0.136	0.138	0.18
PTF29	PSD+DEPTH_M+OC+BD+PH_H2O+CEC	0.068	0.083	0.095	0.109	0.135	0.134	0.18
PTF30	PSD+DEPTH_M+OC+CACO3+PH_H2O+CEC	0.162	0.100	0.101	0.108	0.145	0.138	0.17
PTF31	PSD+DEPTH_M+BD+CACO3+PH_H2O+CEC	0.070	0.081	0.097	0.108	0.134	0.137	0.18
PTF32	PSD+DEPTH_M+OC+BD+CACO3+PH_H2O+CEC	0.069	0.079	0.097	0.107	0.135	0.135	0.18

<sup>1</sup>PSD: particle size distribution (sand, 50–2000 µm; silt, 2–50 µm; clay, <2 µm (mass %)); DEPTH: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm<sup>-3</sup>); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg<sup>-1</sup>).

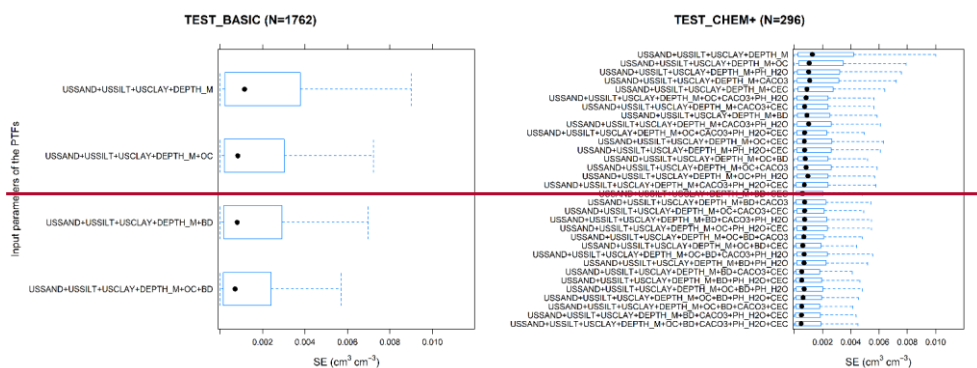
<sup>2</sup>THS: saturated water content (pF 0); FC 2: water content at -100 cm matric potential head (pF 2.0); FC: water content at -330 cm matric potential head (pF 2.5); AWC 2: plant available water content based on FC 2; AWC: plant available water content based on FC; WP: water content at wilting point (pF 4.2); KS: saturated hydraulic conductivity;



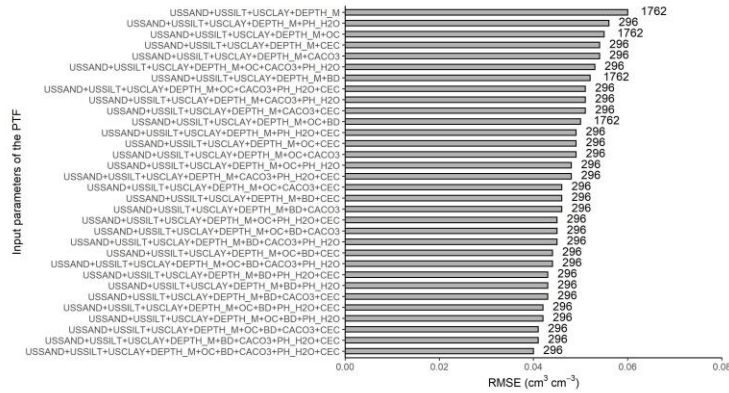
**Figure S2. Root mean squared error (RMSE) of the pedotransfer functions derived to predict water content at saturation (THS) computed on TEST\_BASIC and TEST\_CHEM+ set.** USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



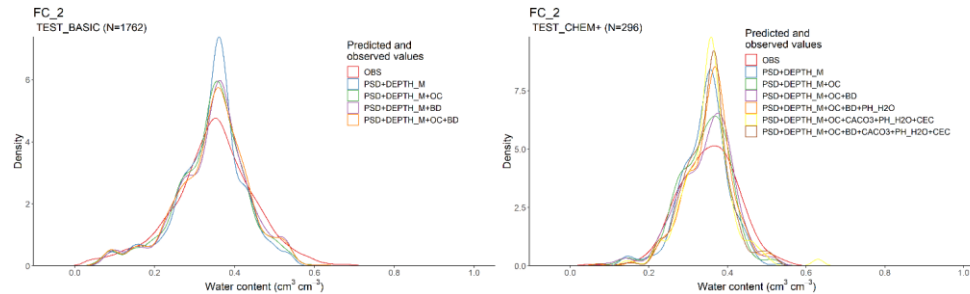
**Figure S3.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) water content at saturation (THS) for selected pedotransfer functions, computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



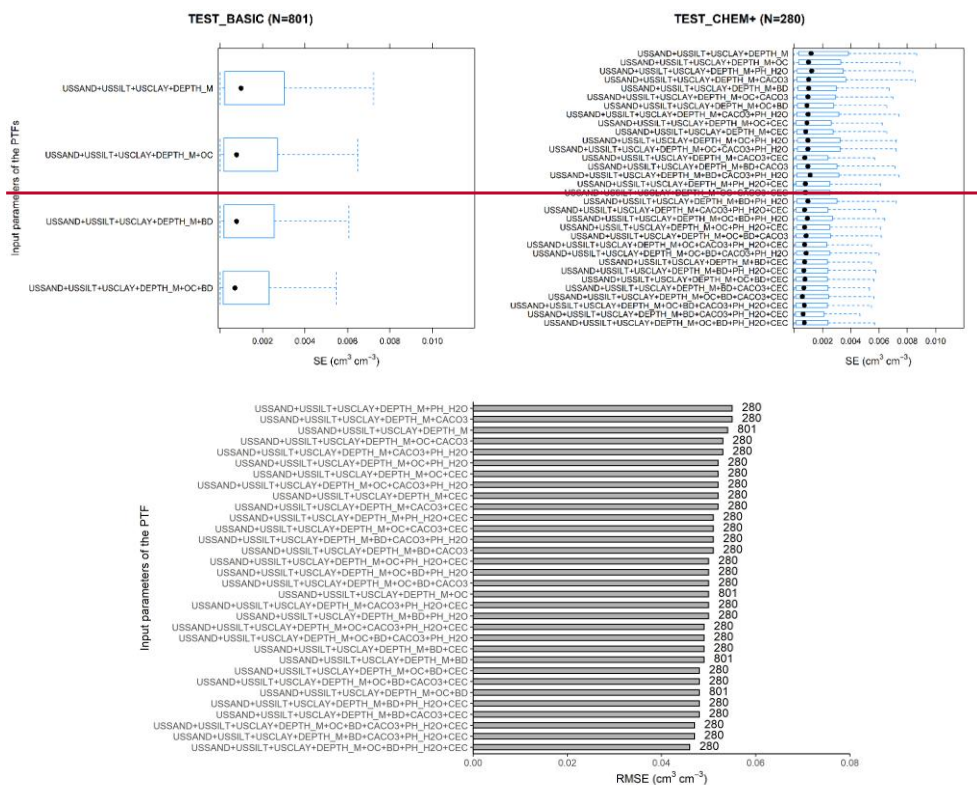




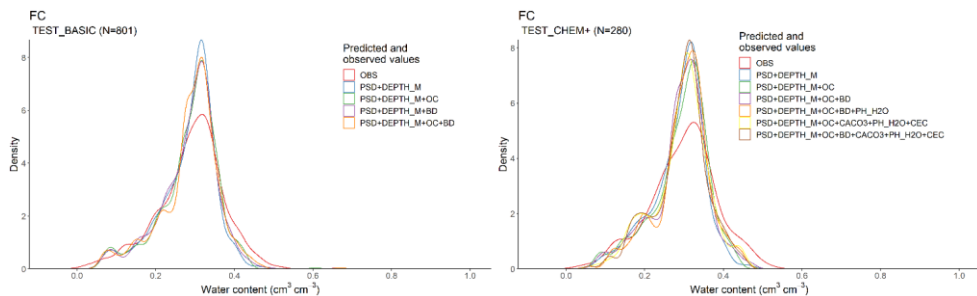
**Figure S4.** Root mean sQuared error (RMSE) of the pedotransfer functions derived to predict water content at -100 cm matric potential head (FC\_2) computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



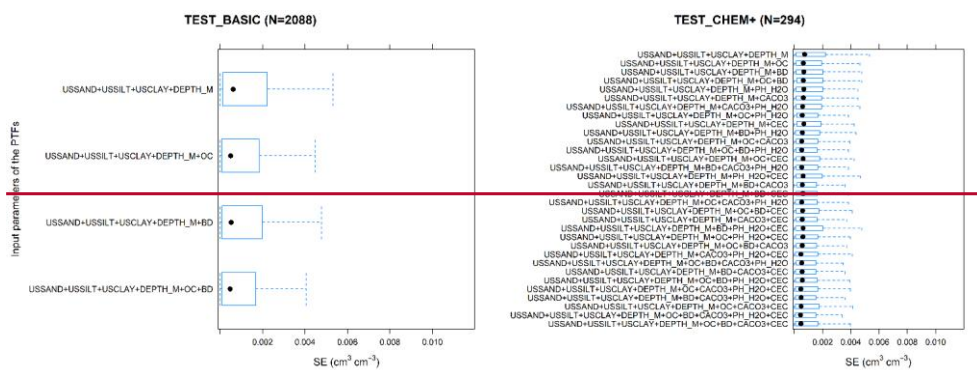
**Figure S5.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) water content at -100 cm matric potential head (FC\_2) for selected pedotransfer functions, computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

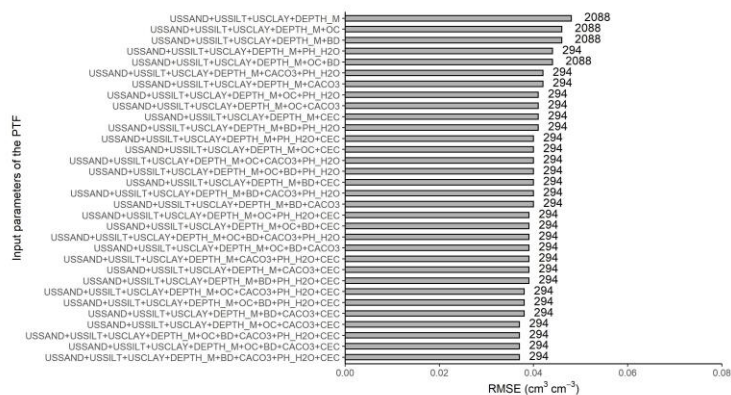


**Figure S6.** Root mean squared error (RMSE) of the pedotransfer functions derived to predict water content at -330 cm matric potential head (FC) computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

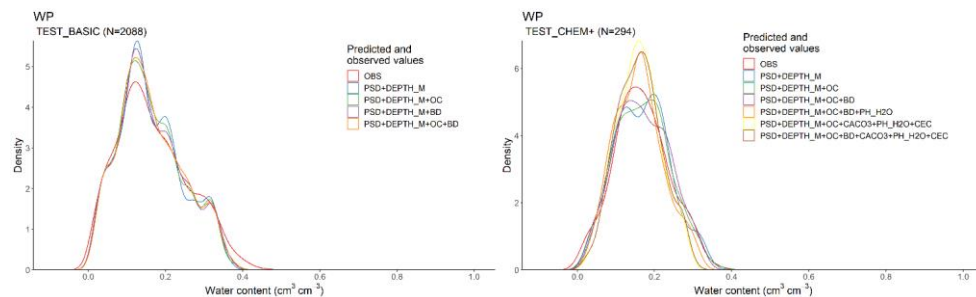


**Figure S7.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) water content at -330 cm matric potential head (FC) for selected pedotransfer functions, computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

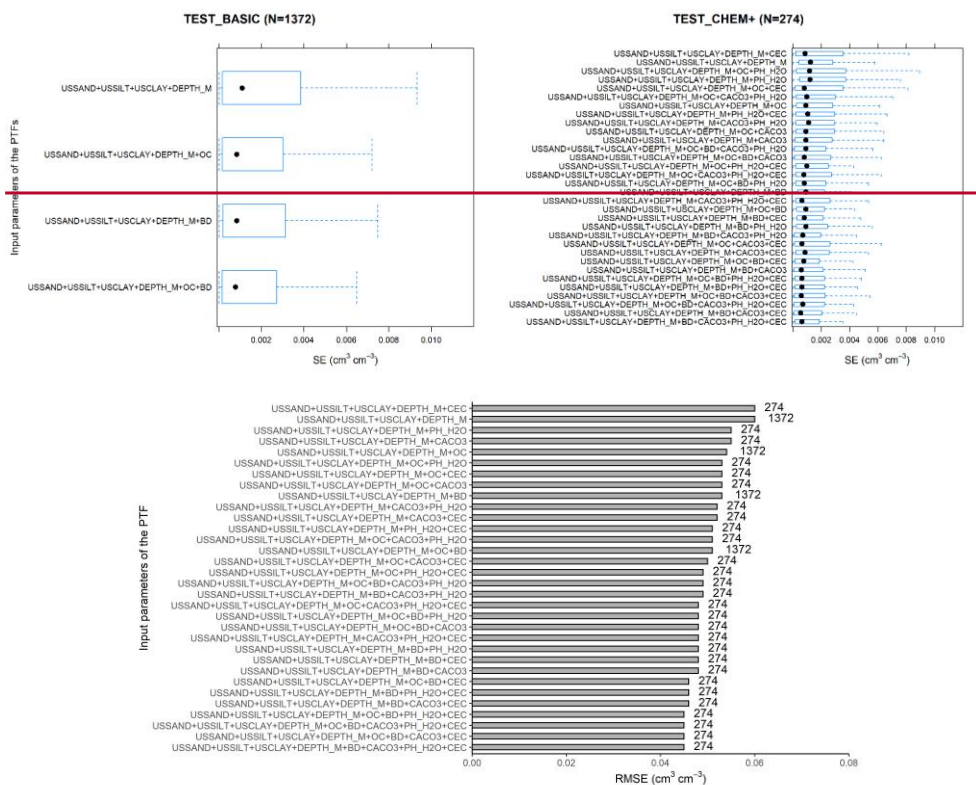




**Figure S8.** Root mean sSquared error (RMSE) of the pedotransfer functions derived to predict water content at wilting point (WP) computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

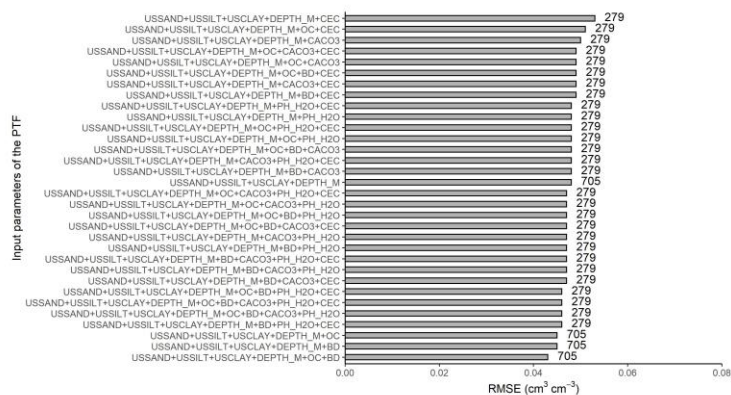


**Figure S9.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) water content at wilting point (WP) for selected pedotransfer functions, computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

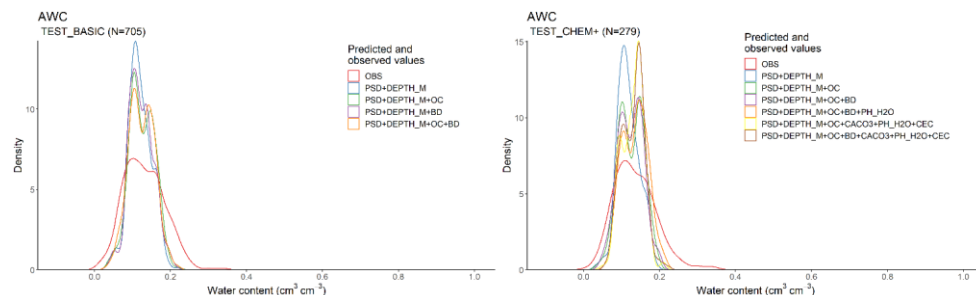


**Figure S10. Root mean squared error (RMSE) of the pedotransfer functions derived to predict plant available water content (AWC<sub>2</sub>) considering field capacity at -100 matric potential head (FC<sub>2</sub>), computed on TEST\_BASIC and TEST\_CHEM+ set.** USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



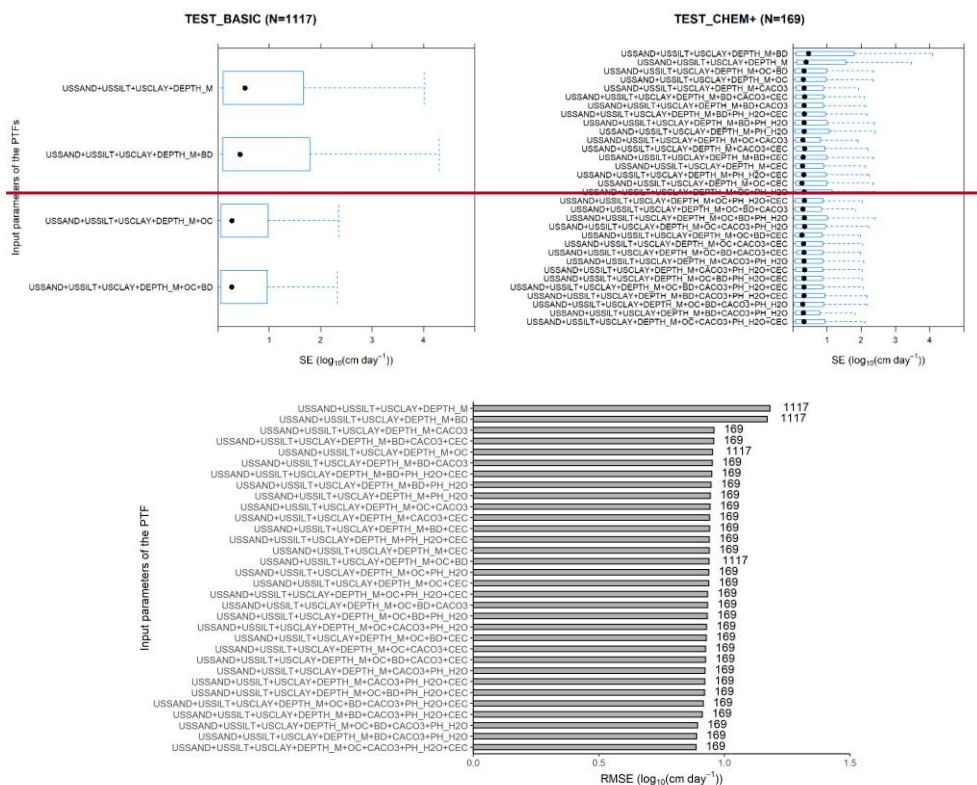


**Figure S12.** Root mean squared error (RMSE) of the pedotransfer functions derived to predict plant available water content (AWC) considering field capacity at -330 matric potential head (FC), computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



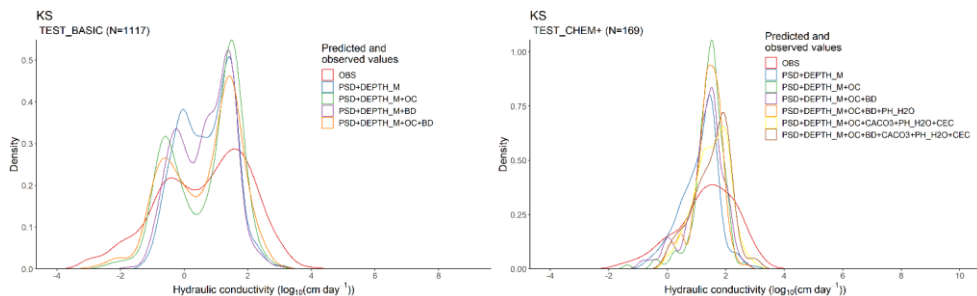
**Figure S13.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) plant available water content (AWC) considering field capacity at -330 matric potential head (FC) for selected pedotransfer functions, computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



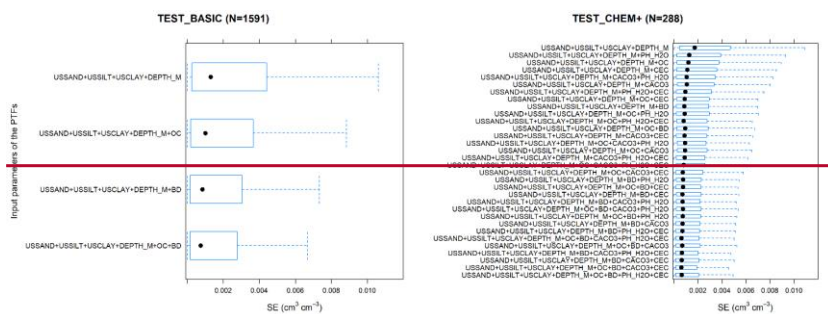


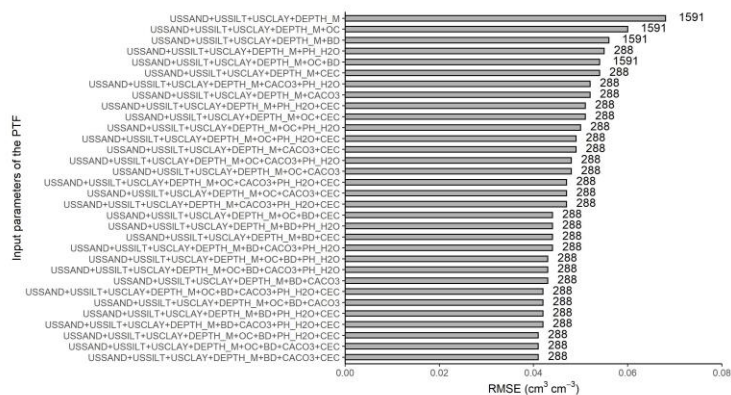
**Figure S14. Root mean squared error (RMSE) of the pedotransfer functions derived to predict saturated hydraulic conductivity (KS), computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).**



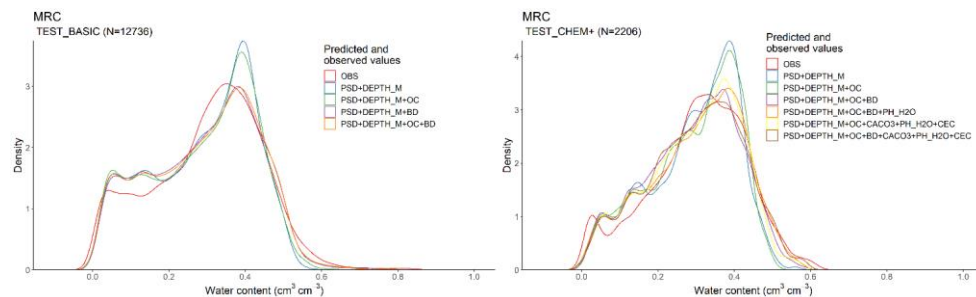


**Figure S15.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) saturated hydraulic conductivity (KS) for selected pedotransfer functions, computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).

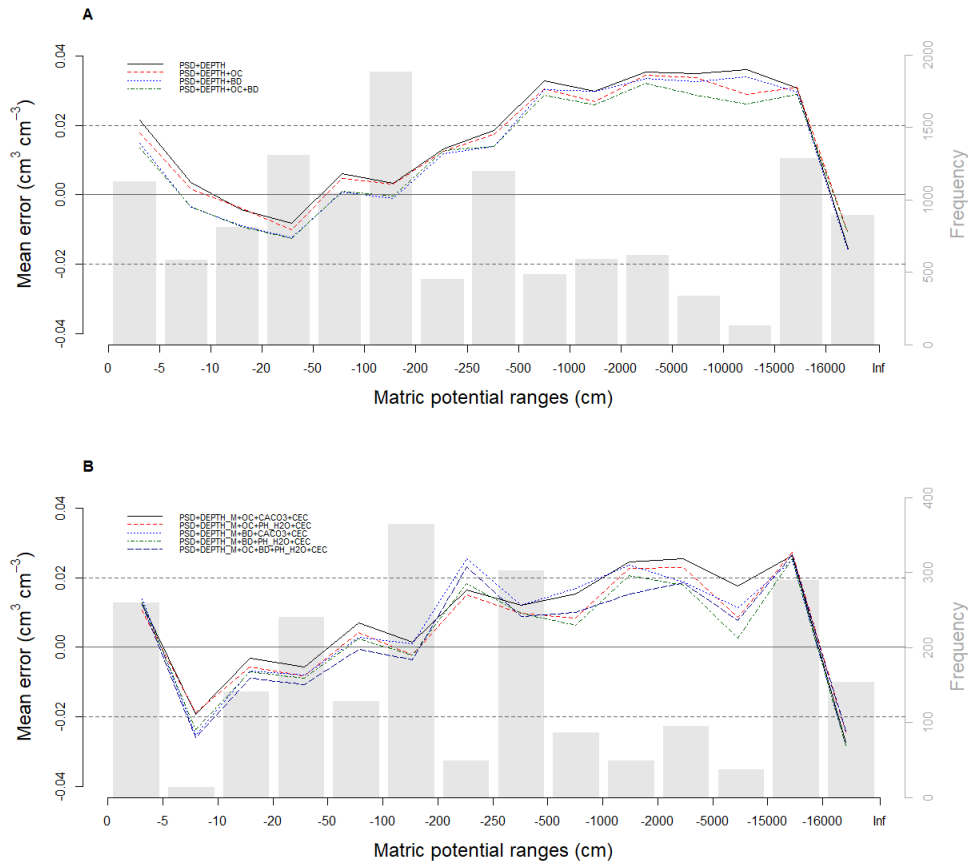




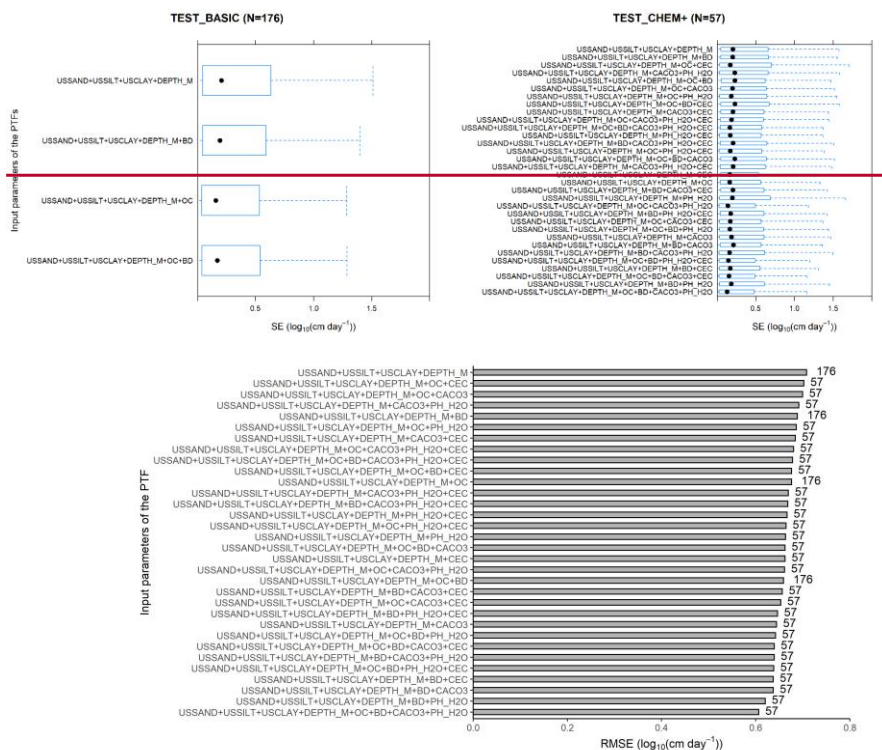
**Figure S16.** Root mean sQuared error (RMSE) of the pedotransfer functions derived to predict parameters of the van Genuchten model for the description of the moisture retention curve (MRC), computed on TEST\_BASIC and TEST\_CHEM+ set. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



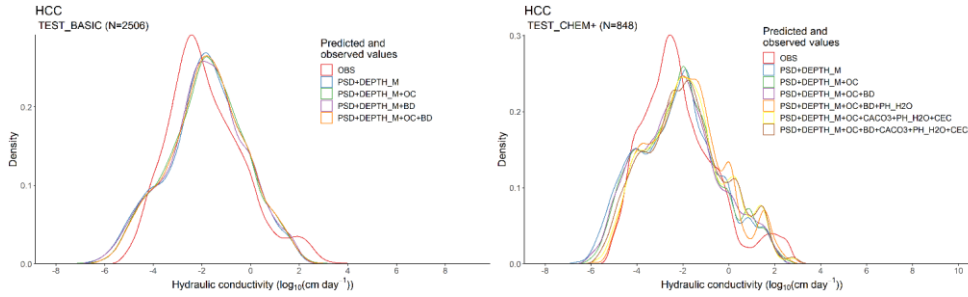
**Figure S17.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) water retention values (MRC) computed based on the parameters of the van Genuchten model, computed on TEST\_BASIC and TEST\_CHEM+ set. Predicted values of those PTFs are shown which use the most often available predictor variables. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



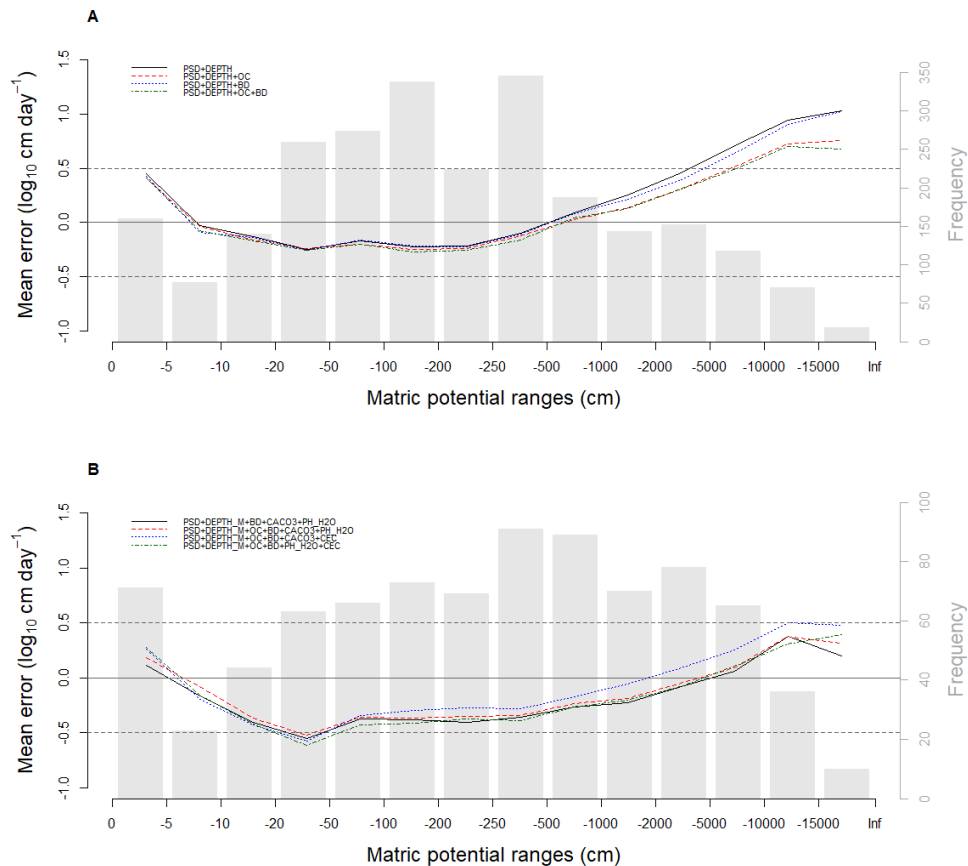
**Figure S18.** Mean error of the pedotransfer functions derived to predict parameters of the van Genuchten model for the description of the moisture retention curve, computed on TEST\_BASIC (N = 1591) (A) and TEST\_CHEM+ (N = 288) (B) sets by matric potential head values.



**Figure S19. Root mean squared error (RMSE) of the pedotransfer functions derived to predict parameters of the Mualem-van Genuchten model for the description of the hydraulic conductivity curve (HCC), computed on TEST\_BASIC and TEST\_CHEM+ set.** USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



**Figure S20.** Density plot of observed (OBS) and predicted median (PSD+DEPTH\_M+\*) hydraulic conductivity values (HCC) computed based on the parameters of the Mualem-van Genuchten model, computed on TEST\_BASIC and TEST\_CHEM+ set. Predicted values of those PTFs are shown which use the most often available predictor variables. USSAND: sand (50–2000  $\mu\text{m}$ ) content (mass %); USSILT: silt (2–50  $\mu\text{m}$ ) content (mass %), USCLAY: clay (<2  $\mu\text{m}$ ) content (mass %); DEPTH\_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density ( $\text{g cm}^{-3}$ ); CACO3: calcium carbonate content (mass %); PH\_H2O: pH in water (-); CEC: cation exchange capacity ( $\text{cmol (+) kg}^{-1}$ ).



**Figure S21.** Mean error of the pedotransfer functions derived to predict parameters of the Mualem-van Genuchten model for the description of the hydraulic conductivity curve, computed on TEST\_BASIC (N = 176) (A) and TEST\_CHEM+ (N = 57) (B) sets by matric potential head values. **Table S3.** List of recommended pedotransfer functions (PTFs) by predicted soil hydraulic property and available predictor variables.

Predictor variables <sup>1</sup>	Recommended PTFs <sup>2</sup>									
	THS	FC_2	FC	WP	AWC	AWC	KS	VG	MVG	
PSD+DEPTH_M	PTF0	PTF0	PTF0	PTF0	PTF0	PTF01	PTF0	PTF0	PTF0	
PSD+DEPTH_M+OC	PTF0	PTF0	PTF0	PTF0	PTF0	PTF01	PTF0	PTF0	PTF0	

formázott: Sorkizárt, Sorköz: 1,5 sor

formázott: Sorköz: 1,5 sor

formázott: Sorköz: 1,5 sor

