Response to Anonymous Referee #3

Thank you for the detailed review and suggestions for further improvements. In the following, we give a detailed presentation of how we will address all the questions and issues raised. Below we would like to answer and provide possible solutions for the comments and recommendations, 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 will be added to the manuscript are in blue italic font,
- the reference to the lines and pages relates to the discussion paper available from: https://gmd.copernicus.org/preprints/gmd-2020-36/gmd-2020-36.pdf .

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 eutptf 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 will correct it: "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 etal., 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 AmericaJournal, 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 will add this information to P2 L16:

"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 we will add the following to P2 L33:

"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 vs. TEST_BASIC set	Training vs. 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

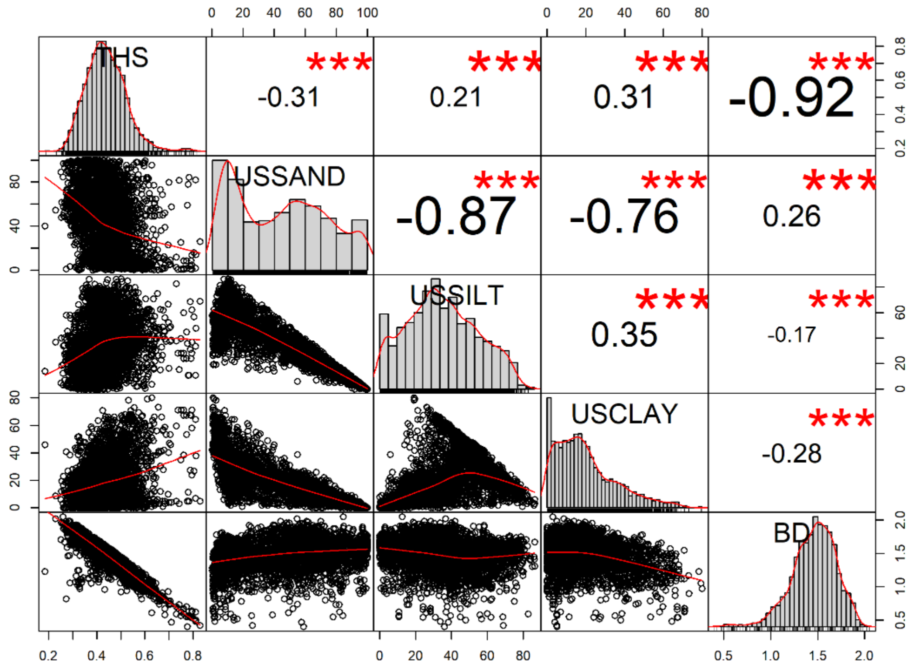
Train and test datasets in bootstraps are divided in the following way: in the random forest algorithm for each tree 63% of the data is selected with replacement to build the tree, i.e. number of selected data will be increased to reach the number of samples of the training set with the replacement, 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:

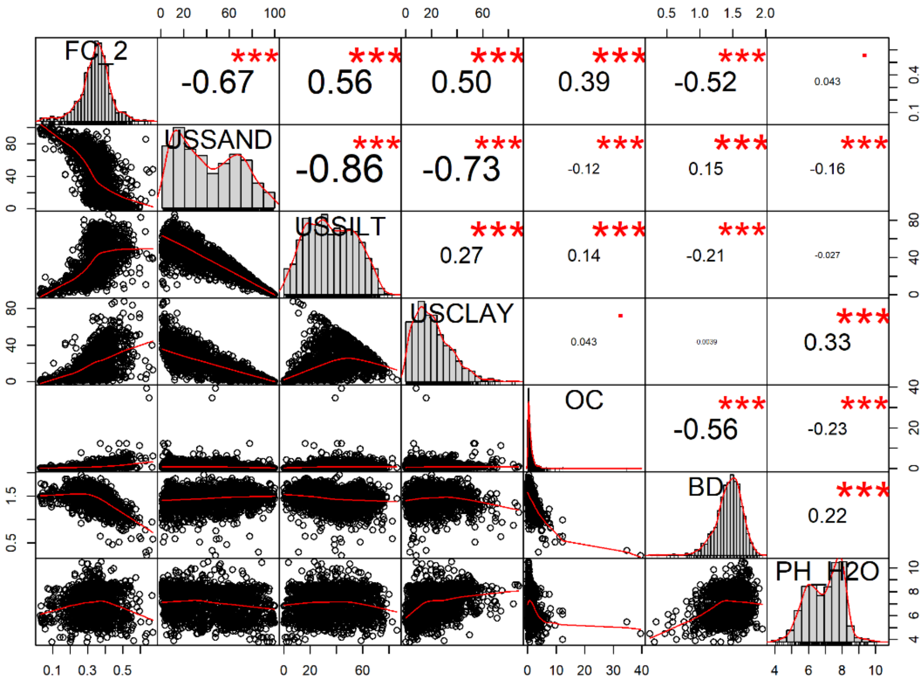
Table1. P.18. Numbers are not aligned exactly below the names.
Correlation matrix of observations would be useful information (in appendix) at least for the dataset used for the best PTFs.

AGREED

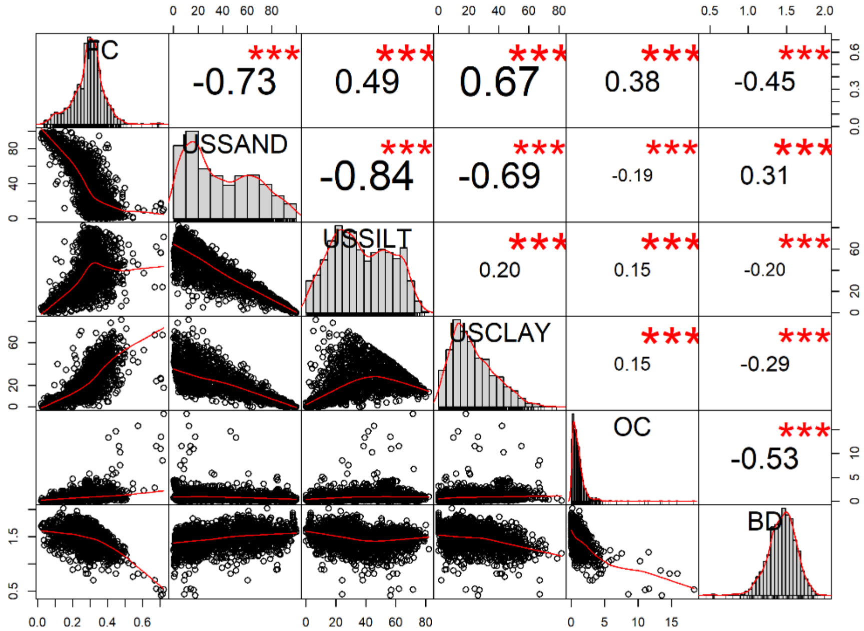
A: The columns' names will be aligned with the numbers below.
The correlation plots of the best PTFs are inserted below the answers (Fig_responses_1 – Fig_responses_7), however descriptive power of them are limited because the relationship between predicted parameters and predictors are not linear. This 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.



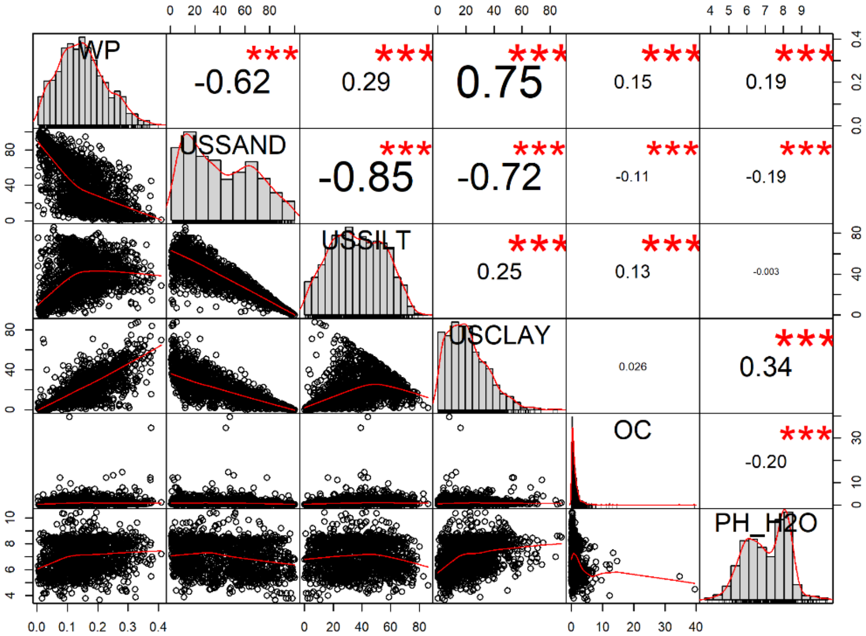
Fig_responses_1



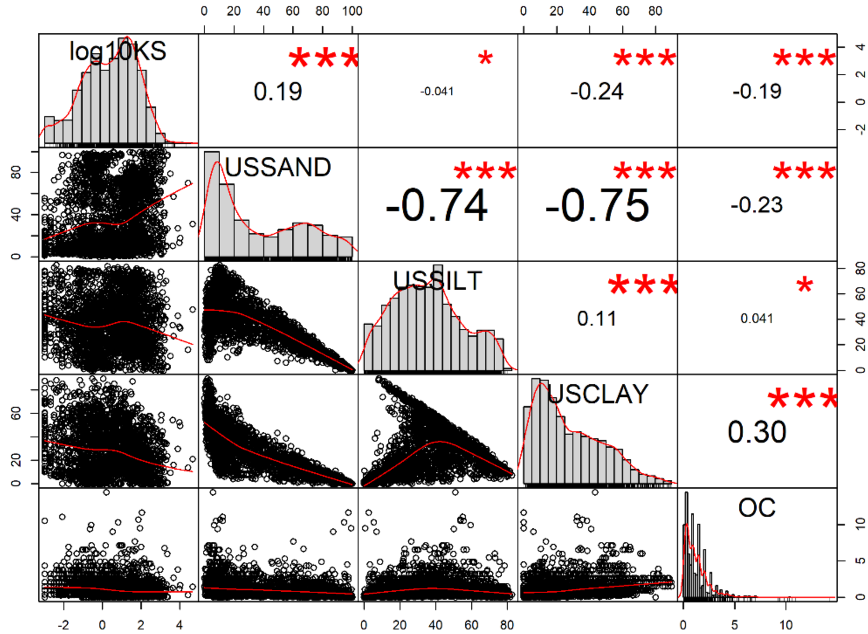
Fig_responses_2



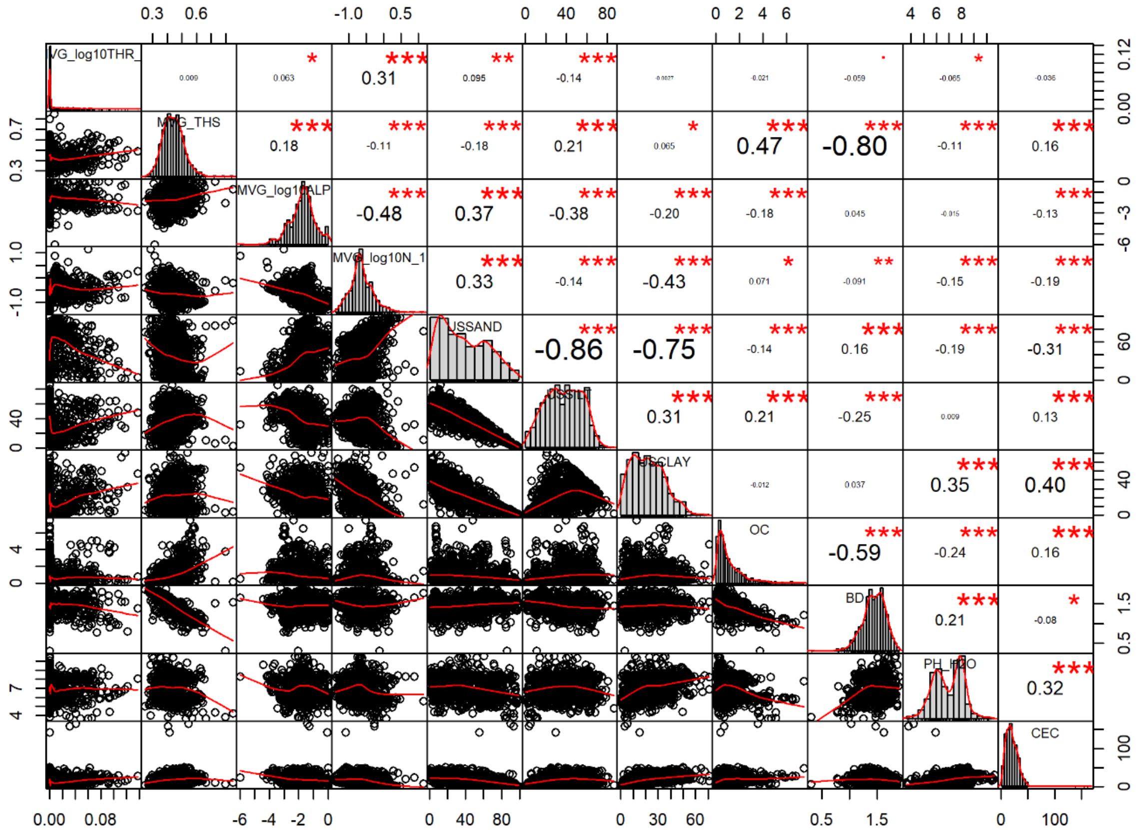
Fig_responses_3



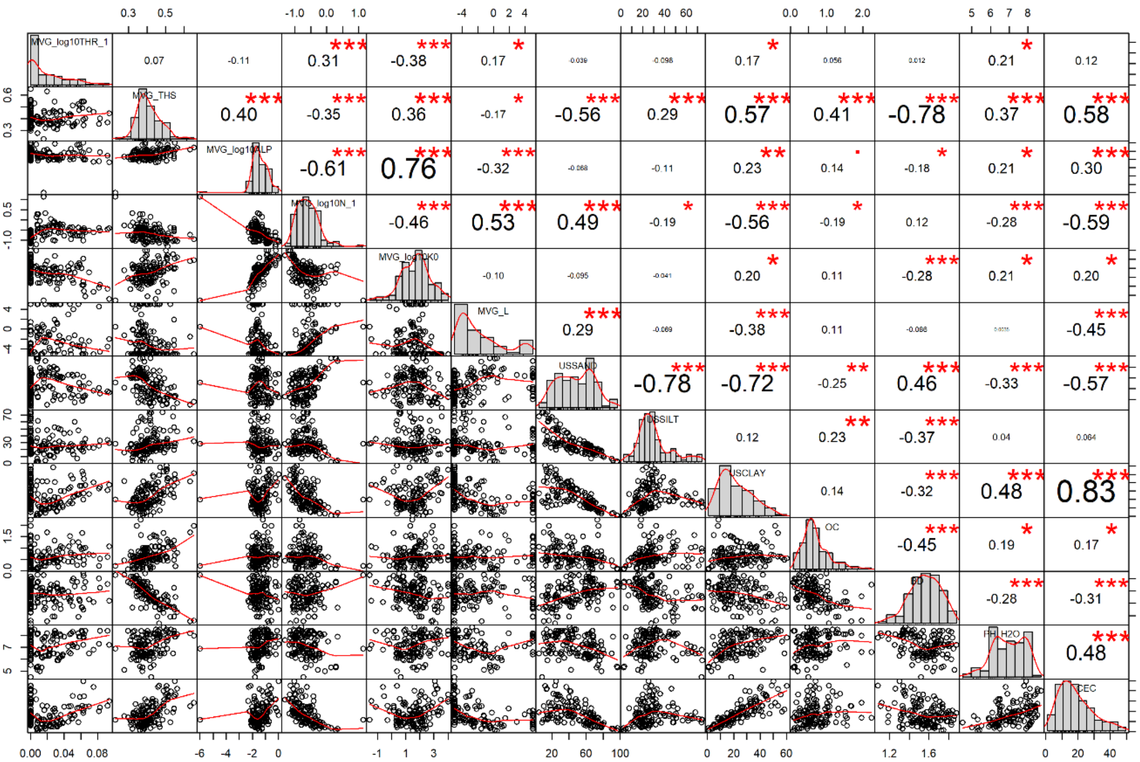
Fig_responses_4



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: Please find here the relative importance plots, with which we will replace Figure 3 and 7 and specify that relative variable importance is shown:

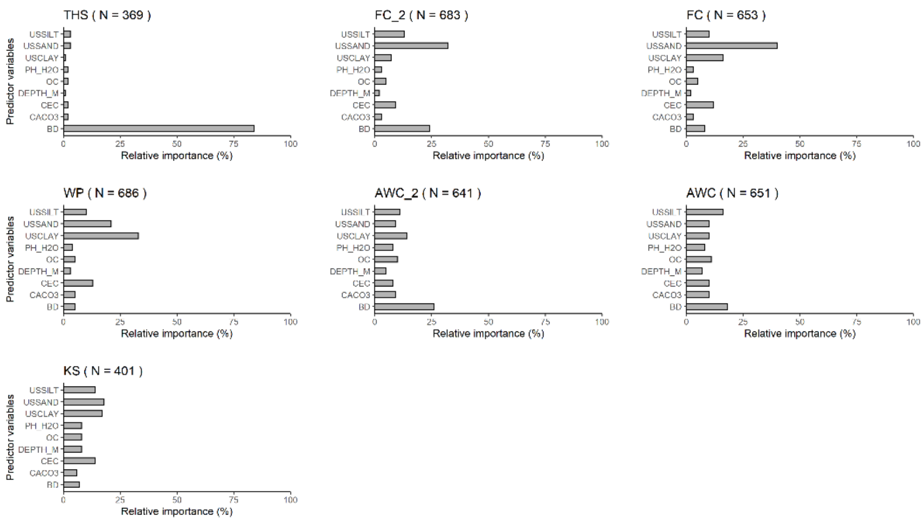


Figure 3. *Relative variable importance computed with the random forest algorithm for the prediction of water content with PTF32 at saturation (THS), at field capacity; -100 (FC_2) and -330 (FC) matric potential head, at wilting point (WP), of the plant available water content based on FC_2 (AWC_2) and FC (AWC), and the saturated hydraulic conductivity (KS). USSILT:*

silt content (2–50 µm (mass %)); USSAND: sand content (50–2000 µm (mass %)); USCLAY: clay content (<2 µm (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 (cmol (+) kg^{-1}); CACO3: calcium carbonate content (mass %); BD: bulk density (g cm^{-3}).

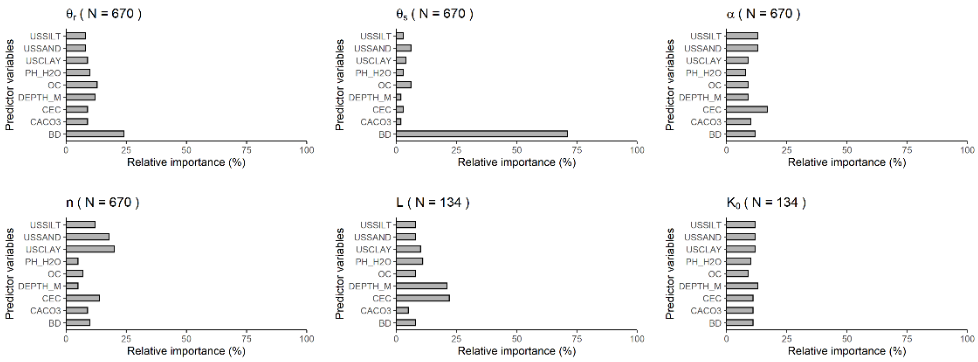


***Figure 7.** Relative variable importance computed with the random forest algorithm for the prediction of parameters of the van Genuchten and Mualem-van Genuchten models based on PTF32. ϑ_r: residual water content (cm^3 cm^{-3}); ϑ_s: saturated water content (cm^3 cm^{-3}); α (cm^{-1}), n (-): fitting parameters; K_0: the hydraulic conductivity acting as a matching point at saturation (cm day^{-1}); L: shape parameter related to pore tortuosity (-); USSILT: silt content (2–50 µm (mass %)); USSAND: sand content (50–2000 µm (mass %)); USCLAY: clay content,(<2 µm (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 (cmol (+) kg^{-1});> CACO3: calcium carbonate content (mass %); BD: bulk density (g cm^{-3}).*

We will add the following text in L29 P5:

"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 will add the normalized RMSE (RMSE/(ymax-ymin)), which was also suggested by the reviewer under "L1-8, P10".

The following texts will be added:

P6 L10:

"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 .

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

"

and in P7 L6 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 the following table will be included in the supplementary material as Table S3 (The original TableS3 of the supplementary material available from https://gmd.copernicus.org/preprints/gmd-2020-36/gmd-2020-36-supplement.pdf will be moved to the manuscript as Table 11).

Table S3. Normalized root mean squared error (NRMSE) of the point predictions by soil hydraulic properties computed on the test datasets in $cm^3\ cm^{-3}$ for water retention and $\log_{10}$ (cm day^{-1}) 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 euptfv2	Predictor variables[1]	NRMSE in test sets[2]						
		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

[1]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^{-3}); CACO3: calcium carbonate content (mass %); PH_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg^{-1}).

[2]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;

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 will add the following information in P12 L12:

"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 will replace Figure S2, S4, S6, S8, S10, S12, S14, S16, S19 showing SE with the one showing RMSE according to the following, e.g.:

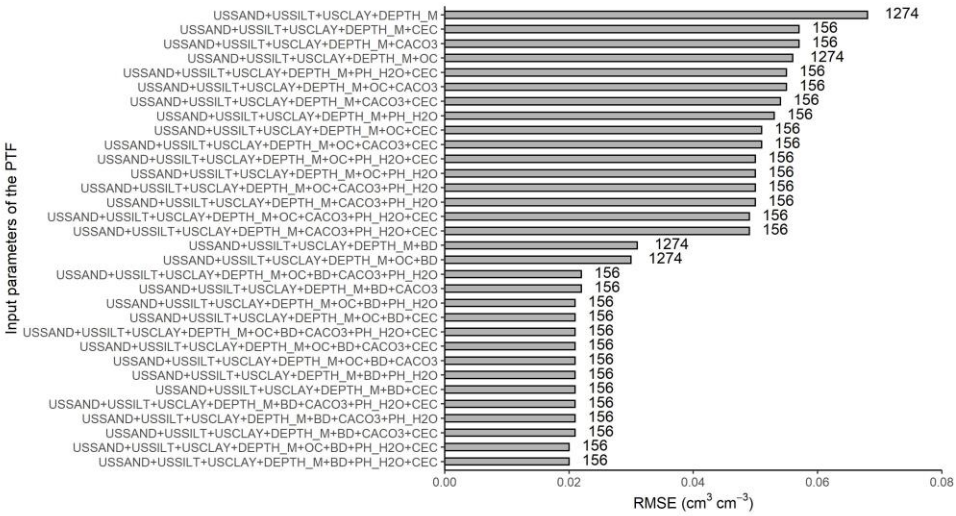


Figure S2. *Root mean squared error (RMSE) of the pedotransfer functions derived to predict water content at saturation (THS) computed on TEST_BASIC (N=1274) and TEST_CHEM+ set (N=156). USSAND: sand (50–2000 µm) content (mass %); USSILT: silt (2–50 µm) content (mass %), USCLAY: clay (<2 µm) content (mass %); DEPTH_M: mean soil depth (cm); OC: organic carbon content (mass %); BD: bulk density (g cm^{-3}); CACO3: calcium carbonate content (mass %); PH_H2O: pH in water (-); CEC: cation exchange capacity (cmol (+) kg^{-1}).*

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 and 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/cm3), 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^3 cm^{-3})**	**RMSE (cm^3 cm^{-3})**	**R^2**	**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 add it in that sentence:
"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}(cm day^{-1}) – has an RMSE of 0.94 log_{10}(cm day^{-1}) …"

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 will highlight this error measure in the case of KS and call it normalized RMSE (NRMSE). We also computed the NRMSE for
- the literature referred in the manuscript:
    - Zhang and Schaap (2017) (ROSETTA3): 0.11 (cm/day) (PSD+BD)
    - Lilly et al. (2018) (HYPRES) 0.18 log10 (cm/day) (topsoil/subsoil distinction+USDA soil texture class+PSD+BD+OC),
    - Araya and Ghezehei (2019) (USKSAT database) 0.06 log10 (cm/day) (PSD+BD+OC),
- Nemes et al. (2005) (HYPRES) 0.15 log10 (cm/day).

We will add the information on computing NRMSE to P6 L10 in the manuscript as mentioned above, and the following:

P9 L31:
"… has an RMSE of 0.94 log_{10}(cm day^{-1}) and NRMSE 0.14 log_{10}(cm day^{-1}) (Table S3)."

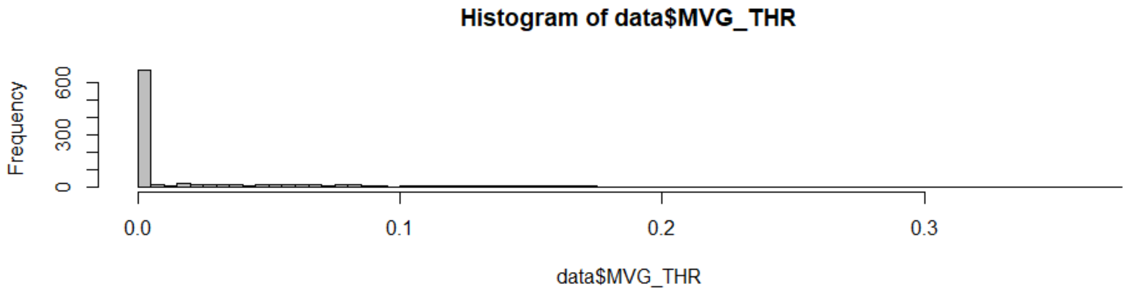
P10 L3-7:
"ROSETTA3 PTF with PSD and BD predictors had and RMSE of 0.68 log_{10} (cm day^{-1}) with an NRMSE of 0.11 log_{10} (cm day^{-1}) (Zhang and Schaap, 2017). Araya and Ghezzehei (2019) published PTF using PSD, BD and OC predictors with highest

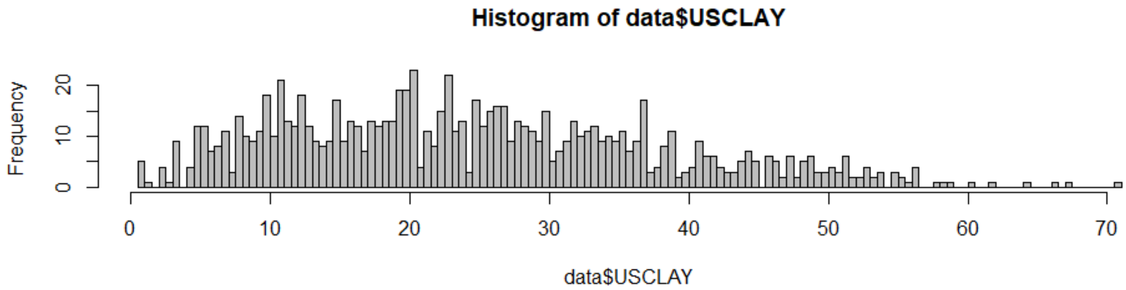
accuracy in the literature with an RMSE of 0.34 $\log_{10}$ (cm day^{-1}) and NRMSE of 0.06 $\log_{10}$ (cm day^{-1}). In Lilly et al. (2008), the performance of the KS predictions and findings were similar to this study. They report an RMSE between 0.95 and 1.08 $\log_{10}$(cm day^{-1}) – with an NRMSE between 0.17 and 0.20 $\log_{10}$(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 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 will be added to P4 L10:

Similarly to euptfv1, for the description of the moisture retention curve (MRC), we predicted the VG model parameters: the residual water content (ϑ_r), the saturated water content (ϑ_s), and shape parameters α 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 will delete the sentence starting with "In parts, this is …" (P11 L32- P12 L2) and add the following text to P12 L2:

"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 will be 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 will correct it to cm^3 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: θ_r: residual water content (cm^3 cm^{-3}), θ_s: saturated water content (cm^3 cm^{-3}), α (cm^{-1}) and n (-): fitting parameters, K_0: the hydraulic conductivity acting as a matching point at saturation (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 will add a paragraph entitled "Practical guidance on how to use the PTFs" on P12 L20, 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 will add to P6 L16:

"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 add the complementary information on P12 L8:

"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, that were used to develop the predictions and the derived pedotransfer functions. The dataset which we used for training and testing the algoritms can not be shared according to the agreement between the data holders.

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 .
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.