Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions

Sixtine Cueff, Yves Coquet, Jean-Noël Aubertot, Liliane Bel, Valérie Pot, Lionel Alletto

Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions

Sixtine Cueff[a,b,*], Yves Coquet[b], Jean-Noël Aubertot[a], Liliane Bel[c], Valérie Pot[b], Lionel Alletto[a,*]

[a] Université de Toulouse, INRAE, UMR AGIR, F-31326, Castanet-Tolosan, France

[b] Université Paris-Saclay, INRAE, AgroParisTech, UMR ECOSYS, 78850, Thiverval-Grignon, France

[c] Université Paris-Saclay, INRAE, AgroParisTech, UMR MIA-Paris, 75005, Paris, France

*corresponding authors. Lionel Alletto, lionel.alletto@inrae.fr; Sixtine Cueff, sixtine.cueff@gmail.com

Abstract

Conservation agriculture has been developed as a means to improve the sustainability of agricultural systems and reduce drawbacks of conventional agricultural practices. Cropping practices can have a large influence on soil properties such as water retention. Proper tools are needed to assess and model effects of conservation agriculture on soil properties. As measuring soil water retention is expensive and time consuming, pedotransfer functions (PTFs) are now commonly used to predict them. The objectives of this study were to (i) present a new dataset of conservation agriculture data, (ii) assess performances of existing PTFs in predicting soil water retention of soils under conservation agriculture and (iii) develop new specific PTFs to predict water retention in conservation agriculture more accurately. We used data collected only in fields under conservation agriculture in France to evaluate several published PTFs with three evaluation criteria (RMSE, prediction bias (ME) and Nash-Sutcliffe Efficiency (EF)). We then developed new PTFs using three methods – multiple linear regression, regression tree and random forest – to predict soil water content at matric heads of -100 (θ_{100}, field capacity for sandy soils), -330 (θ_{330}, field capacity for other soils) and -15 000 cm ($\theta_{15\,000}$, wilting point). Soil tillage, presence of a cover crop, rotation length and previous reduced/no tillage were used as predictors in addition to basic soil properties for regression trees and random forests. The quality of prediction (RMSE, ME and EF) was calculated for each new PTF using a cross-validation procedure. Generally, predictions of wilting point had lower absolute error than those of sandy-soil field capacity (RMSE = 0.044 and 0.066 cm^3/cm^3, respectively). EF was usually negative for all water contents. The cross-validation performance of the new PTFs was similar for multiple linear regression (RMSE: 0.028, ME: 0.000, EF: 0.34 for θ_{100}) and random forest (RMSE: 0.027, ME: 0.000, EF: 0.36 for θ_{100}), and generally worse for regression tree (especially EF). Multiple linear regression that did not consider cropping practices performed as well as random forest and thus did not identify any major influence of agricultural management on predicted water content. Future research on developing PTFs should focus on identifying more relevant predictors.

Keywords: soil water content, pedotransfer functions, available water capacity, soil tillage, linear regressions, regression trees, random forests

1. Introduction

Conservation agriculture was developed to enhance the sustainability of agricultural systems and reduce drawbacks of conventional agriculture, especially soil degradation due to erosion (Hobbs et al., 2008). Conservation agriculture combines three main interrelated soil conservation techniques: (i) little or no soil disturbance, (ii) permanent soil cover by crop residues and/or living cover crops and (iii) diversification of plant species (FAO, 2016). Interactions among these three techniques lead to complex and interrelated modifications in soil physical, chemical and biological properties. Considering these changes is crucial to assess performances of such agricultural systems properly. However, studies of impacts of conservation agriculture on soil properties show many inconsistencies, especially for soil hydraulic processes (Green et al., 2003; Strudley et al., 2008; Verhulst et al., 2010).

Effects of soil cultivation practices on soil properties has received much research attention in recent decades, but clear trends have not been established due to differences in location, soils and agricultural practices (Green et al., 2003; Strudley et al., 2008). Tillage tends to decrease bulk density and increase macroporosity, thus increasing the saturated and near-saturated hydraulic conductivity of the tilled layer. These effects are, however, strongly time-dependent and usually disappear rapidly after tillage (Mapa et al., 1986), due to natural soil reconsolidation caused by wetting and drying cycles (Ahuja et al., 1998). Simultaneously, tillage interrupts macropore connectivity between the soil surface and the untilled deeper soil, thus decreasing water movement throughout the entire soil profile (Cameira et al., 2003). Conversely, untilled soils have higher bulk density and greater pore connectivity (Gozubuyuk et al., 2014). Cover crops may (partially) counterbalance negative effects of no tillage on bulk density by, for example, creating stable biopores through their root development during the growing season (Williams and Weil, 2004; Abdollahi and Munkholm, 2014). Moreover, after cover crop destruction, the dead residues form a mulch that physically protects the soil surface from crusting (Baumhardt and Lascano, 1999). Maintaining crop residues on the soil surface also leads to accumulation of soil organic matter in topsoil layers (Kay and VandenBygaart, 2002) and improves aggregate stability (Devine et al., 2014). In parallel, increased macrofauna activity (especially of earthworms) in conservation agriculture systems forms biomacropores that improve water infiltration (Shipitalo et al., 2000). Finally, soils under conservation agriculture also tend to have a larger proportion of finer pores (micropores) (Hill et al., 1985). These changes in pore-size distribution could improve the storage of plant-available water (Bescansa et al., 2006).

The variety and complexity of the counteracting effects of conservation agriculture on soil properties call for developing new tools to properly assess and model these effects. Development of water- and solute-transport models has received much research attention in recent decades. The lack of accurate data on soil hydraulic properties, especially for soils under conservation agriculture, however, hinders the use of models, as they require water-retention and hydraulic conductivity data as inputs (Wösten et al., 1999). Despite significant improvements in measuring techniques, researchers agree that directly measuring water-retention curves remain expensive, time consuming and impossible at a large scale (Wösten et al., 2001; Vereecken et al., 2010; Román Dobarco et al., 2019).

Predicting hydraulic properties may be accurate enough to be used in water- and solute-transport models (Wösten et al., 2001). One promising solution to managing the scarcity of hydraulic data is to use pedotransfer functions (PTFs), which relate easily available soil properties to properties that are more difficult to measure, such as hydraulic ones (Al Majou et al., 2008b). Many PTFs have been developed, and two main groups of water-retention PTFs can be distinguished: “point” PTFs, which predict volumetric water content at a given matric head, and “parametric” PTFs, which predict parameters of the water-retention curve as described by van Genuchten (1980). In addition, depending on the type of input data used, PTFs can be further divided into “class-PTFs” and “continuous-PTFs”. Class-PTFs predict mean volumetric water content at a given matric head or mean water-retention curve parameters using information such as textural class, type of horizon and bulk density class (Al Majou et al., 2008b; Bruand et al., 2004). Continuous-PTFs are regression equations that predict volumetric water content at a given matric head or water-retention curve parameters using continuous input variables such as granulometric fractions, bulk density and soil organic carbon content (Al Majou et al., 2008a; Rawls et al., 1982). More recently, novel machine-learning methods have been used to develop PTFs based on regression trees (i.e. “tree-PTFs”) (Toth et al. 2015).

Although PTFs have significantly facilitated widespread application of water- and solute-transport models at the field scale and larger scales (Vereecken et al., 2010), some of their limits have been identified. Several authors suggested that using information in addition to the commonly used sand, silt and clay contents, bulk density and organic matter could improve prediction accuracy (Vereecken et al., 2010). Water contents at selected matric heads (Rawls et al., 1983; Al Majou et al., 2008a) or terrain attributes (Obi et al., 2014) have been proposed as additional information. Land cover (Nemes et al., 2003) or soil management (Tóth et al., 2015) have also been proposed, but they may create PTFs that are less applicable than those that use only soil properties as parameters. Whether the available PTFs apply equally to soils under conservation or conventional agriculture has not yet been explored. The type of agriculture under which the soils used to develop a particular PTF is rarely specified, but most PTFs seem to have been developed from soils under conventional agriculture. To our knowledge, no one has attempted to develop specific tools to predict water content in conservation agriculture systems. Chen et al. (1998) did observe that the relevant properties for describing hydraulic conductivity differed between tilled and untilled soil, which highlights the importance of soil management and supports the need for additional data and specific tools to predict water dynamics in soils under conservation agriculture.

The aims of this study were to (i) present a dataset of water retention data from soils under conservation agriculture (ii) assess performances of existing PTFs in predicting soil water retention of these soils and (iii) develop new PTFs using several statistical techniques to improve representation of the hydraulic properties of soils under conservation agriculture.

2. Materials and methods

2.1 Description of the dataset on conservation agriculture

Information on farming operations and soil chemical and physical characteristics were collected from 2009-2011 in 47 fields under conservation agriculture in the central basin of the Occitanie region in south-west France. Soil types there are mainly hypereutric cambisols, luvisols and calcaric cambisols (IUSS Working Group WRB, 2015). All fields had been cultivated using conservation practices since 1987-2003. Four types of tillage were used: deep tillage (DT), with a working depth >15 cm (n=7 fields); reduced tillage (RT), with a working depth of 5-15 cm (n=18); strip-till (ST), with tillage restricted to the future row (n=3); and no tillage (NT) (n=19). In addition to tillage, cover crops were used on 35 of the fields. Four classes of crop rotation were defined: rotation length > 4 years (n=24); > 2 years to ≤ 4 years (n=15); ≤ 2 years (n=2); and not fixed (n=6).

In each field, soil samples were collected from the topsoil (0-30 cm) and then divided into three layers: 0-10 cm (47 samples), 10-20 cm (47 samples) and 20-30 cm (46 samples). Several physicochemical properties were determined using international and French norms (NF) published by the French national organization for standardization (AFNOR) from one bulk sample per layer. The granulometric distribution of five decarbonated fractions (clay (<2 µm), fine silt (2-20 µm), coarse silt (20-50 µm), fine sand (50-200 µm), coarse sand (200-2000 µm)) was determined using NF X31-107. Soil samples from the fields were concentrated in the silty and clayey zones of the texture triangle (Fig. 1). NF ISO 10694 was used to determine carbon content and estimate organic matter content. NF ISO 10390 was used to determine pH (in water). NF ISO 11263 was used to determine phosphorus content (P_2O_5) using the Olsen method. NF ISO 10693 was used to determine total calcium carbonate content. Cation exchange capacity (CEC) and exchangeable CaO, Na_2O, K_2O and MgO were determined using NF ISO 23470. When CaO content was found to be saturated (i.e. not quantifiable by this method), it was calculated as CEC minus the sum of Na_2O, K_2O and MgO. The Kjeldahl method was used to determine nitrogen content.

In addition, for each topsoil layer, soil bulk density was determined from undisturbed soil samples collected with 250 cm^3 cylinders (8 cm in diameter, 5 cm high), and the soil water-retention curve was determined from undisturbed soil samples collected with 50 cm^3 cylinders (5 cm in diameter, 2.5 cm high). Bulk density was measured in triplicate for each layer. Soil water retention was usually measured in duplicate or triplicate (rarely, only one sample was available) and recorded in the dataset as a mean value. Volumetric water content (θ, cm^3/cm^3) was measured successively at 0 (θ_0), -100 (θ_{100}), -330 (θ_{330}), -3300 (θ_{3300}) and -15 000 ($\theta_{15\,000}$) cm of matric head. θ_0 was measured after the cylinders were saturated for two days on a tray filled with glass beads (diameter ≈ 0.45 µm). The other water contents were measured using pressure plates. The resulting data were used to fit water-retention curve parameters using the RETC program (van Genuchten et al., 1991) based on the van Genuchten (1980) equation (eq. 1):

$$\theta_h = \theta_r + \frac{\theta_s - \theta_r}{[1+(\alpha h)^n]^m} \qquad m = 1 - \frac{1}{n} \qquad (1)$$

where θ_r [cm^3/cm^3] and θ_s [cm^3/cm^3] are the residual and saturated volumetric water content (θ) respectively, h is matric head [cm], and α [cm^{-1}], n [-], and m [-] are shape parameters of the curve. The fit of the curves to the data had a mean R^2 (± 1 SD) of 0.98 ± 0.02.

Finally, plant available water capacity (AWC, in mm) was calculated as follows:

$$AWC = (\theta_{FC} - \theta_{WP}) \times H \qquad (2)$$

where θ_{FC} and θ_{WP} are volumetric water content at field capacity and permanent wilting point (cm^3/cm^3), respectively, and H is the depth of each of the three layers (here, 100 mm).

According to the literature, θ_{FC} can equal either θ_{100} (for sandy soils) or θ_{330} (for other soils), and θ_{WP} equals $\theta_{15\,000}$ (Hillel, 1971). Both AWC_{100} and AWC_{330} were considered for the two definitions of θ_{FC}. However, PTFs are usually used to predict volumetric water content at several matric heads rather than AWC. The rest of the study thus focused only on the relation between θ_{100}, θ_{330}, $\theta_{15\,000}$ and basic soil properties and/or cropping practices.

2.2 Analysis of the dataset

Principal component analysis (PCA) was performed to explore relations among the explanatory variables, using the "FactoMineR" package of R software (version 3.6.1) (R Core Team, 2019) using only soil properties. Soil water contents were used only as supplementary variables. Spearman correlations were calculated between explanatory variables and soil water content at different matric heads, using the "psych" R package. Unbalanced Type II analysis of variance (ANOVA) was performed to investigate effects of soil tillage, cover-crop presence, rotation length and soil depth on soil water contents, using the "car" R package.

2.3 Published pedotransfer functions

Twenty nine existing PTFs that predict θ_{100} or θ_{330}, and/or $\theta_{15\,000}$ (Table 1) and eight PTFs that predict three parameters (n, α, and θ_s) of the van Genuchten (1980) water-retention curve (Eq. 1, Table 2) were taken from the literature and applied to data for the 140 soils in this study. The study used class-PTFs (Cl), continuous-PTFs (Co) and tree-PTFs (Tr). PTFs were calibrated using several published databases (Table 1). Of the 26 PTFs that predict θ_{100}, 13 were Cl and 13 were Co. Of the 28 PTFs that predict θ_{330}, 13, 13 and 2 were Cl, Co and Tr, respectively. Of the 27 PTFs that predict $\theta_{15\,000}$, 13, 12 and 2 were Cl, Co and Tr, respectively. These published PTFs use different variables as predictors, such as texture/granulometric fractions, bulk density and organic carbon content. Two PTFs (M2_Co and M3_Co) also use θ_{FC} and/or θ_{WP} as predictors. However, as a water content cannot be used to predict itself, M2_Co and M3_Co were not used to predict θ_{330} or $\theta_{15\,000}$. Most publications identified in the literature (Table 1) also had PTFs for subsoil horizons (> 30 cm). We used only the published PTFs developed for the topsoil as the dataset contained only topsoil data. All PTFs were applied to soil data in our dataset to predict θ_{100}, θ_{330}, $\theta_{15\,000}$, n, α and θ_s.

2.4 Development of new pedotransfer functions

Three types of PTFs, which predicted θ_{100}, θ_{330} or $\theta_{15\,000}$, were developed. Redundant properties (calculated from another property), such as organic matter content and the C:N ratio, were removed from the input data. Table 3 provides summary statistics of the variables that were used for each of the following methods.

2.4.1 Multiple linear regression

We developed multiple linear regressions using stepwise regression with forward selection, which could include all soil properties as predictors. In this procedure, the Akaike information criterion (AIC) (Akaike et al., 1998) was used to determine which set of predictors predicted water content best. AIC is calculated at each step of the stepwise regression to determine the improvement brought by adding the new predictor. The "best" model is the one that helps decrease AIC the most. The procedure stops when no more improvement can be made by a new predictor or when all predictors are included.

2.4.2 Regression tree

Regression tree methods consist of recursive binary partitions of a dataset. At each node, observations are split according to a decision rule based on only one predictor. Splitting continues until all of the subsets (i.e. "terminal nodes" of the tree) are as homogeneous as possible with reference to the response variable (Hastie et al., 2009; Prasad et al., 2006). Splitting stops when the subset reaches a minimum size of 5 data points or when no more improvement can be made. The criterion used to decide which predictor splits the data best is based on ANOVA. First, a maximum tree is grown that likely overfits the training data. To reduce the size of the tree and avoid overfitting, the tree is then pruned using cost-complexity pruning (10 cross validation). Briefly, for each pair of terminal leaves with a common parent node, the error in classifying the testing dataset is calculated to see whether the sum of squares would be smaller by turning the parent nodes into a terminal leaf. The procedure is repeated until the pruning does not decrease the error in the testing data. The resulting pruned tree is usually smaller than the initial maximum tree, but in theory, pruned trees can range from the maximum size to minimum size (no partitions, no tree). The size of the pruned tree can depend on the cross-validation method used. The pruned tree to be used as a model for each water content was then randomly selected. The response variable was volumetric water content at a given matric head, and the terminal nodes of the tree represented mean water content in the partitions. The "rpart" R package (Therneau and Atkinson, 2019) was used to build the trees.

2.4.3 Random forest

Like for regression tree, random forest is also based on recursive partitions of the data. The difference is that a forest of multiple decorrelated trees is grown by using a randomly bootstrapped subset of data and a random subset of predictors (Hastie et al., 2009; Ließ et al., 2012). The "randomForest" R package (Liaw and Wiener, 2002) was used to build forests. The forest consisted of 500 trees, and six of 18 variables were randomly selected to grow each tree. Like for regression tree, the minimum size of a terminal node was 5 data points. Unlike for regression tree, however, a single tree cannot be extracted from the forest, but the relative importance of the predictors can be determined and used to help interpret the results. The relative importance of predictors was estimated according to how much worse the prediction would be if the data for that predictor were permuted randomly (Prasad et al., 2006).

2.5 Evaluation of pedotransfer functions

PTFs were evaluated by comparing predicted values to observed values in the dataset according to three criteria: root mean squared error (RMSE), mean error (ME) (also called "bias") (Bruand et al., 2003) and Nash-Sutcliffe efficiency (EF; Nash and Sutcliffe, 1970). They are calculated as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}[f(x_k) - y_k]^2} \quad (3)$$

$$ME = \frac{1}{N}\sum_{k=1}^{N}[f(x_k) - y_k] \quad (4)$$

$$EF = 1 - \frac{\sum_{k=1}^{N}[f(x_k) - y_k]^2}{\sum_{k=1}^{N}[y_k - \bar{y}]^2} \quad (5)$$

where $f(x_k)$ are the values predicted by the PTF, y_k are the observed values in the conservation agriculture dataset, x_k are the input data (basic soil properties) needed by PTF f, $\bar{y}$ is the mean of observed values and N is the number of data points.

RMSE = 0 indicates perfect prediction of the observed data, while the ME indicates whether the PTF overpredicts (positive ME) or underpredicts (negative ME) the observed data. The closer ME is to 0, the lower the bias is. EF=1 indicates perfect prediction of the observed data, while EF<0 indicates prediction worse than the that using the mean of observed values (for which EF=0). These criteria have no thresholds that can be used to conclude whether a prediction is good or not; nevertheless, to help interpret the results, we arbitrarily defined ranges to indicate satisfactory prediction of AWC: less than 0.020 cm^3/cm^3 for RMSE and ME, and greater than 0.50 for EF.

The three criteria were used to assess the performance of the published and new PTFs. For published PTFs, $f(x_k)$ corresponded to predictions using basic soil properties in the conservation agriculture dataset as input data, assessed with the criteria $RMSE_P$, ME_P and EF_P. For the new PTFs, two groups of criteria were used to evaluate their performance. One group of three criteria ($RMSE_A$, ME_A, EF_A) evaluated the quality of adjustment to the data. In this case, $f(x_k)$ corresponded to predictions by the new PTF using basic soil properties in the same dataset from which they had been developed. The second group of criteria ($RMSE_{CV}$, ME_{CV}, EF_{CV}) evaluated the cross-validation quality of prediction. As the dataset contained too few soils (N=140) to split out an independent validation dataset, leave-one-out cross validation (Hastie et al., 2009) was performed instead. In it, the dataset was split 140 times into two datasets of 139 and 1 soils, respectively. The 140 datasets of 139 soils were used to calibrate 140 new PTFs. The 140 predictions were then compared to their corresponding value in the dataset of observed values.

3. Results

3.1 Preliminary analysis of the dataset

AWC_{100} and AWC_{330} (in the 0-10, 10-20 and 20-30 cm soil layers) ranged from 10.4-28.6 mm and 4.2-22.9 mm, respectively, depending on the soil layer. Both varied little as a function of depth, tillage or cover-crop presence (Fig. 2). However, differences were larger as a function of rotation length (Fig. 2d, h). Mean AWC_{100} was ca. 20, 18 and 16 mm when the rotation length was variable, medium/long and short, respectively. Despite small differences, statistical analysis demonstrated a significant effect of the three cropping practices (i.e. tillage, cover-crop presence and rotation length) (except for tillage for AWC_{330}) and of depth for both AWCs. Both AWC_{100} and AWC_{330} were highest (by a small degree) in the 0-10 cm layer (Fig. 2a, e).

The plane defined by the first two axes of the PCA of basic soil properties explained 57% of the variance of the dataset (Fig. 3a). Of the 14 basic soil properties, only 8 contributed significantly (i.e. more than if each one had contributed equally (i.e., 7%)) to the first two axes. Strong correlations were found between CEC, CaO content and clay content ($r_{CEC/CaO} = 1$, $r_{Clay/CaO} = r_{CEC/CaO} = 0.9$), which contributed the most to the first two axes due to their large contributions to the first axis (17%, 16% and 16%, respectively). Nitrogen, organic carbon and phosphorus contents contributed the most to the second axis (22%, 21% and 19%, respectively). Strong to very strong correlations were found between organic carbon, nitrogen and K_2O contents ($r_{OC/N} = 0.94$, $r_{K_2O/OC} = r_{K_2O/N} = 0.7$). Thus, soil layers above the second axis of the PCA had higher organic carbon, nitrogen and phosphorus contents, which was related to their depth, as most soil layers above the second axis were 0-10 cm deep (Fig. 3b). This is consistent with the low mechanical disturbance of the soil surface under conservation agriculture, which results in a thin horizon 5-10 cm deep that can exhibit different soil properties, especially organic matter. When projected as supplementary variables on the plane, water contents were poorly represented (Fig. 3a), which suggested that none of the basic soil properties were strongly related to them, as confirmed by correlation coefficients. The strongest significant correlations for θ_{100} were with clay content ($r = 0.5$), bulk density ($r = -0.4$), sand content ($r = -0.4$) and CEC ($r = 0.4$). Correlations for θ_{330} were weaker, not exceeding 0.3 with clay content or -0.3 with bulk density. Correlations for $\theta_{15\,000}$ were the strongest among those for the three water contents: 0.6 with clay content, CEC and CaO content.

We plotted θ_{100}, θ_{330} and $\theta_{15\,000}$ vs. cropping practices, rotation length, soil tillage and cover-crop presence to identify the influence of conservation agriculture on water contents. We also investigated the influence of depth, as the PCA indicated a difference between the 0-10 cm layer and the other two layers. There were no clear differences between θ_{100} or θ_{330} as a function of agricultural practices, except for rotation length, with water content lower with variable rotations and higher with short rotations, compared to long or medium rotations (Fig. 4a, b). ANOVA confirmed a significant effect of rotation length on θ_{100} ($P < 0.001$) and θ_{330} ($P < 0.01$). For $\theta_{15\,000}$, water content was lower under strip-till than under the other types of tillage and had a trend similar to those of θ_{100} and θ_{330} for rotation length (Fig. 4c, d). All three cropping practices had a significant effect on $\theta_{15\,000}$ ($P < 0.01$ for soil tillage and $P < 0.001$ for cover-crop presence and rotation length). Unlike for AWC, depth had no significant effect on any of the water contents.

3.2 Evaluation of the performance of published pedotransfer functions

<u>3.2.1 Prediction of volumetric water content at selected matric heads</u>

For prediction of θ_{100}, $RMSE_P$ varied from 0.034 cm^3/cm^3 (M3_Co) to 0.262 cm^3/cm^3 (M2_Co) (Table 4). These extreme values were exceptions, however; mean (± 1 SD) $RMSE_P$ for most of the PTFs (22 of 26) was 0.055 ± 0.009 cm^3/cm^3. Of the 26 PTFs, 24 underpredicted θ_{100}, with ME_P ranging from -0.112 to -0.007 cm^3/cm^3. The same four PTFs that had extreme values of $RMSE_P$ (M1_Co, M2_Co, M3_Co and M10_Co) had extremely high or low ME_P. For EF_P, negative or near-zero values showed that none of the PTFs tested predicted θ_{100} well. According to the three criteria, M3_Co, despite having been developed from samples from many locations in the USA, predicted θ_{100} the best, but used both θ_{330} and $\theta_{15\,000}$ as predictors. However, the other two PTFs developed from the same data (M1_Co, M2_Co) predicted θ_{100} the worst. Among the remaining PTFs, which used only basic soil properties, eight French Cl PTFs (M7_Cl, M8_Cl, M12_Cl, M13_Cl, M14_Cl, M19_Cl, M20_Cl, M21_Cl) had better $RMSE_P$ (0.046 ± 0.004 cm^3/cm^3) and ME_P (-0.028 ± 0.004 cm^3/cm^3) than the others. However, ME remained unsatisfactory. All eight PTFs were Cl that used FAO texture or FAO texture and bulk density as classes.

For prediction of θ_{330}, $RMSE_P$ ranged from 0.037 cm^3/cm^3 (M4_Cl) to 0.080 cm^3/cm^3 (M10_Co) and were thus lower overall than those for θ_{100}. Of the 28 PTFs, 16 overpredicted θ_{330} (ME_P=0.017 ± 0.015 cm^3/cm^3). The worst ME_P (-0.069 cm^3/cm^3) was an underprediction by M10_Co (Table 4). Four PTFs (M1_Co, M2_Co, M10_Co and M16_Tr) performed worse than the others for all three criteria, especially M10_Co, a PTF for topsoil layers developed by Al Majou et al. (2007). Although the $RMSE_P$ and ME_P of the other 24 PTFs were lower, their EF_P never reached satisfactory values (≥ 0.5), so their potential use remains limited.

For prediction of $\theta_{15\,000}$, $RMSE_P$ varied from 0.034 cm^3/cm^3 (M22_Co) to 0.057 cm^3/cm^3 (M10_Co) and were thus lower overall than those of the other water contents (Table 4). Of the 27 PTFs, 18 overpredicted $\theta_{15\,000}$ (ME_P=0.008 ± 0.007 cm^3/cm^3), but there was no systematic bias. Overall, two groups of performance were identified. The first, with lower $RMSE_P$, low ME_P and positive EF_P, were the eight Co of Román Dobarco et al. (2019) and the Co of Tóth et al. (2015). This group of PTFs could probably be used with lower risk of poor prediction. Nevertheless, even though their EF_P were positive and much higher than those of the other two water contents, they still had difficulty reaching the satisfactory threshold.

3.2.2 Prediction of water-retention curve parameters

For predicting θ_s, $RMSE_P$ ranged from 0.035-0.439 cm^3/cm^3, while ME_P ranged from -0.438 to 0.010 cm^3/cm^3 (Table 5). P2_Co and P4_Co had large errors due to physically impossible values of θ_s (close to 0 or even negative). For the other PTFs that predicted θ_s, $RMSE_P$ and ME_P had satisfactory performances, with the best performance by P3_Cl, P7_Co, P8_Co and P9_Co ($RMSE_P$= 0.037 ± 0.001 cm^3/cm^3; ME_P= 0.023 ± 0.010 cm^3/cm^3; EF_P= -0.43 ± 0.12). Negative EF_P values, however, indicated that none of the PTFs performed better than the mean of observed values.

For predicting α, the French PTF P3_Cl had particularly poor performance according to all criteria, and P4_Cl predicted physically impossible values. Thus, the best predictions were obtained only with PTFs developed at the European scale, all of which performed similarly. For predicting n, $RMSE_P$ varied from 0.305-0.366. The nine PTFs always underpredicted n (negative ME_P), except for P8_Co, which had the only satisfactory ME_P (0.003) and the best $RMSE_P$. The two PTFs developed from soil samples from France performed slightly worse according to all criteria.

3.3 Development of new pedotransfer functions

3.3.1 Multiple linear regression

All regressions developed from our dataset (N=140) included clay content and bulk density as predictors (Table 6). The sign of the coefficients associated with these two variables was similar in each regression and indicated that water content increased as clay content increased and bulk density decreased. Regressions for θ_{100} and $\theta_{15\,000}$ also included silt content as predictor, with a positive effect. Other predictors were included only once in the regressions. Of the 14 potential predictors, only five, four and four were kept in the θ_{100}, θ_{330} and $\theta_{15\,000}$ regressions, respectively. The qualities of adjustment and cross-validation did not differ greatly, except for slightly better EF_A than EF_{CV} (Table 7, Fig. 5a). Predictions of θ_{330} had worse EF_A (and EF_{CV}) than the other water contents did.

3.3.2 Regression tree

The maximum tree grown for θ_{100}, θ_{330} and $\theta_{15\,000}$ had 11, 9 and 8 partitions, respectively, despite the inclusion of 18 potential predictors. After pruning, the θ_{330} tree was reduced to the minimum size (no partitions); thus, the mean of θ_{330} was the best compromise between a suitable tree size and low error in predicting the testing data. Consequently, only the trees that predicted θ_{100} and $\theta_{15\,000}$ were evaluated (Fig. 6). The pruned θ_{100} and $\theta_{15\,000}$ trees were split 7 and 4 times, respectively, and had three predictors in common: rotation length, clay content and bulk density. Both trees were first split according to rotation length, which split variable length from the other lengths. No other cropping practices appeared in the pruned trees. According to the criteria, all trees had satisfactory quality of adjustment to observed data, with ME_A=0 and EF_A≥0.5 (Table 7). All criteria except ME_A were slightly higher for $\theta_{15\,000}$ than for θ_{100}. The criteria for cross-validation quality of prediction had similar trends, with low ME_{CV}, but the trees did not predict well according to EF_{CV} (<0.21) (Table 7). Prediction performance thus decreased between adjustment and cross validation (Fig. 5b).

3.3.3 Random forest

Clay content was one of the two most important predictors in the θ_{100}, θ_{330} and $\theta_{15\,000}$ random forests (importance of 11%, 10% and 21%, respectively) (Fig. 7). Bulk density was the most important predictor for the θ_{100} and θ_{330} random forests (importance of 14% and 11%, respectively) but not for the $\theta_{15\,000}$ random forest (only 5% importance). Sand content also had significant importance in each random forest, while organic carbon was significant only in the θ_{100} random forest. Rotation length was one of the most important variables in the $\theta_{15\,000}$ random forest (importance of 12%), but the other cropping practices had low importance. All random forests fit well to the data, with low $RMSE_A$, ME_A=0 and EF_A>0.83 (Table 7). The cross-validation quality of prediction showed satisfactory $RMSE_{CV}$ and ME_{CV}, but EF_{CV} remained less than 0.5, which indicated limited performance of the models. Prediction performance thus decreased strongly between adjustment and cross validation (Fig. 5c).

4. Discussion

4.1 Evaluation of the performance of published pedotransfer functions

Most PTFs (24 of 26) underpredicted soil volumetric water content at -100 cm of matric head, while no clear trend (overestimation or underestimation) was observed at -330 and -15 000 cm. The $RMSE_P$ for predicting volumetric water content were largest for the -100 and -330 cm matric heads. Although EF_P was higher for several PTFs at -15 000 cm, it was never satisfactory (≥ 0.5). Overall, none of the 29 published PTFs provided satisfactory prediction of the volumetric water content at the selected matric heads (-100, -330 and -15 000 cm) according to any of the criteria, which limits their use in soil transport models under conservation agriculture.

The published PTFs may have had low-quality predictions for several reasons. First, differences in the sampling or measurement protocol between the databases used to develop the PTFs and the dataset that we used may be a source of uncertainty (Román Dobarco et al., 2019). For example, Al Majou et al. (2008b) measured water content using undisturbed aggregates (10-15 cm^3), whereas we used undisturbed soil cylinders (50 cm^3). Several studies have also highlighted the influence of sample size on soil water retention and the quality of PTFs developed (Ghanbarian et al., 2015; Silva et al., 2018). Furthermore, some of these PTFs were developed from large databases collated in the USA or Europe and covered a wide range of sand, silt, clay and organic matter contents and bulk densities (Rawls et al., 1982; Tóth et al., 2015). Like Cornelis et al. (2001), we calculated the ranges of the soil properties of our samples and found that all lay within those in the databases from the USA and Europe; nevertheless, the predictions were unsatisfactory according to the criteria. Nemes et al. (2003) suggested that using a small set of relevant data rather than a larger, more general dataset can produce more accurate PTFs. Indeed, for predicting Hungarian soils, they found that PTFs that had been developed by neural networks from data from throughout the USA and Europe performed worse than PTFs that had been developed from a smaller dataset that considered the pedoclimatic context (e.g. the subset of Hungarian soils). Testing published PTFs developed from large and general datasets with our dataset may explain the poor prediction in our study. However, most of the PTFs tested were developed from French databases (Bruand et al., 2004; Al Majou et al., 2007, 2008a, 2008b; Román Dobarco et al., 2019) and should have been more appropriate for predicting water content of the soils in our dataset. These French PTFs, however, did not necessarily perform better than those developed by Tóth et al. (2015) at the European scale. They did, however, perform better than those of Rawls et al. (1982), which were developed from soil samples from the USA, which appeared to be unsuitable (criteria among the worst for each PTF evaluated), except when using other water contents as predictors. The poor performance of the

French PTFs was not related to the ranges of soil properties in our dataset, because all of them fell within the domain of applicability of the PTFs tested. Moreover, a metric distance representing a PTF's domain of applicability, developed by Tranter et al. (2009), was calculated for two of the published PTFs whose training dataset was available (M9_Co and M10_Co). Overall, 97% of the data in our dataset belonged to the domain of applicability these PTFs, which confirmed that they could be applied to our dataset.

The poor prediction of water-retention curve parameters by parametric PTFs agrees with results of Ghorbani Dashtaki et al. (2010), who reported that parametric PTFs generally perform worse than point PTFs, as relations between water-retention curve parameters and basic soil properties are complex. The same basic soil properties do not necessarily describe the variability in water content in the wet range and the dry range of the curve, which makes it difficult to capture the relation with them (Tomasella et al., 2003; Ghorbani Dashtaki et al., 2010).

To predict water content better, some authors suggested including other water contents at given matric heads in the PTFs (Al Majou et al., 2008a; Rawls et al., 1982; Vereecken et al., 2010). In our study, predictions of such PTFs were slightly better than those of PTFs that included only soil properties, but with differences depending on the specific water content included in the PTF. As observed by Al Majou et al. (2008a), water content prediction improved when the other water content included was that at field capacity (in this case, θ_{330}), but not that at the wilting point ($\theta_{15\,000}$), as observed by Borgesen and Schaap (2005). The improvement in prediction when using the field capacity water content was related to the shape of soil water-retention curves, which inflected strongly near field capacity. However, determining water content at field capacity in order to include it in PTFs remains unsatisfactory, as doing so, mainly in laboratories, is time-consuming and costly. Other authors suggest that information on soil structure, which is often considered through bulk density, should be included to improve PTF performance. In the study of Al Majou et al. (2008b), including bulk density kept bias low and improved prediction of water content. In our study, predictions of θ_{330} had errors similar to or larger than those of Al Majou et al. (2008b), but unlike their results, including bulk density did not improve predictions. Soil bulk density in conservation tillage systems is generally higher than that in conventional systems, which results in lower total porosity than that in tilled soils but, conversely, generally higher saturated and near-saturated hydraulic conductivity (Green et al., 2003). While, bulk density is a good proxy of hydraulic dynamics (Blanco-Canqui et al., 2004; Alletto et al., 2010) and AWC in conventionally tilled soils, it is less effective in conservation agriculture (Alletto et al., 2010; Chen et al., 1998), probably due to greater pore connectivity and proportion of macro- and mesopores in the latter. This disconnection between hydraulic properties and bulk density in conservation agriculture can indeed be attributed to major changes in pore-size distribution and connectivity when tillage intensity is reduced (Strudley et al., 2008; Alletto et al., 2010), thus leading to changes in AWC. Furthermore, as mentioned by several authors (e.g., Nakano and Miyazaki, 2005; Lilly and Nemes, 2008), the cylindrical core method used to measure bulk density does not predict pore connectivity well, so complementary methods must be used to assess it.

4.2 Development of new pedotransfer functions

Multiple linear regression is commonly used to develop PTFs (Wösten et al., 2001; Al Majou et al., 2008a; Tóth et al., 2015; Román Dobarco et al., 2019), unlike regression trees or random forests. Regression trees have been used to predict water content, but without considering cropping practices: Tóth et al. (2015) predicted θ_{330} and $\theta_{15\,000}$ using textural and taxonomic information (Table 1), while Rawls and Pachepsky (2002) did the same using textural and structural classes. To our knowledge, our study is the first to use random forests to predict water content. Vos et al. (2019) used random forests to highlight the influence of land use or land-use history classes, clay content and electrical conductivity on predicting topsoil carbon stock. In our study, random forests highlighted that some predictors not usually used in PTFs, such as CEC and rotation length, could help predict water content at a given matric head. Some properties have been suggested as important for predicting water content due to an indirect influence, such as organic carbon, which plays both an indirect role, by improving soil structure, and a direct role, through its adsorption properties (Tóth et al., 2015). Cropping practices influence soil properties greatly, especially soil structure (Strudley et al., 2008), and can thus influence water content indirectly. Román Dobarco et al. (2019) suggest that land use should be considered in future PTFs, even though PTFs are generally suitable for most agricultural soils.

However, given the similar cross-validation performances of PTFs developed from random forests and multiple linear regression (which were even better than regression trees), our results do not support the hypothesis that cropping practices are essential for predicting water content in the topsoil (0-30 cm). We also set new parameters for two multiple linear regressions (M22_Co and M28_Co), developed by Román Dobarco et al. (2019), that were among the published PTFs that predicted best; thus, recalibrating existing PTFs rather than developing new ones may be sufficient. Finally, when we developed PTFs from regression trees and random forests without including cropping practices, we obtained nearly identical results.

In terms of quality of adjustment, random forests performed the best, with almost perfect fits. This was likely due to the nature of machine-learning methods, which "learn" from the dataset provided and thus perform well with it. Consequently, we also expected regression trees to have high quality of adjustment, but their results were similar to those of multiple linear regressions. This result was likely related to the pruning, as adjustment to the training data is purposely reduced so that the model performs better with a test dataset. In our study, however, performance of regression trees and random forests decreased between adjustment (i.e. the training dataset) and cross-validation (the test dataset) (Fig. 5). While the poor prediction by the regression trees can be explained easily by their well-known instability (i.e. a small difference in the training dataset can result in a different tree) (Gey and Poggi, 2006; Yang et al., 2016), the instability of the random forests was more surprising. Conversely, multiple linear regression was a stable method whose quality of prediction was as good or better than that of the machine-learning methods. The similarity between its adjustment and cross-validation performances demonstrates its robustness. Overall, however, the cross-validation quality of prediction remained unsatisfactory in this study, mainly for EF_{CV}, which never reached satisfactory values for any of the PTFs despite having satisfactory ME_{CV} (close to 0).

In France, few water-retention data are available in conservation agriculture, and the small size of the dataset may have contributed to unsatisfactory predictions. Indeed, our study was located in a single French region and contained data for relatively few soils (140 samples from 61 agricultural fields). The dataset thus may not represent the wide range of French soil diversity. Moreover, the lack of an independent dataset to validate the new PTFs led us to use cross validation, which estimated only the quality of prediction of the modelling approach. Indeed, as predicted parameter values of the PTFs changed for each soil, the structure of the model could not be tested. Cross validation revealed that even the highly performing random forest method was unstable, which may have resulted from the small sample size. Supplementing the scarce water-retention data would advance development of reliable tools for conservation agriculture. In particular, more data could have helped us better assess the quality of prediction of the PTFs developed. The unsuitability of basic soil properties for predicting water retention remains a major limitation in the development of PTFs (Vereecken et al., 2010). As demonstrated by the study, more relevant predictors of water retention still need to be identified, as using three methods to select the best predictors objectively still yielded unsatisfactory results.

Conclusions

We tested the performance of several published PTFs and newly developed PTFs using multiple linear regressions, regression trees and random forests to predict water content at field capacity (h= -100 or -300 cm) and wilting point (h= -15 000 cm). Although some PTFs approached satisfactory performance according to the three criteria, none of them managed to reach it, which limits their use in soil transport models for conservation agriculture. Most of our soil samples belonged to the domain of applicability of the PTFs, so the poor results obtained are likely related to (i) the use of unsuitable predictors, (ii) the use of PTFs developed at an inappropriate scale or (iii) differences in soil management between databases.

This study, the first to develop PTFs specifically calibrated for conservation agriculture, demonstrated that cropping practices were not necessary to predict water contents. The small size of our dataset was a major obstacle and probably partly explains the unsatisfactory performance of our PTFs, despite using methods designed to yield high performance. Future studies should use larger datasets of soils under conservation agriculture, at more locations, to verify the preliminary results of this study.

The machine-learning methods often selected CEC, which had not been used to develop the PTFs. However, because of low performance, even by random forests, the results suggest that the development of PTFs still lacks suitable predictors. Including more relevant soil properties when developing PTFs thus remains a research path for improving PTFs.

Acknowledgements

This study was performed with data obtained in the framework of the CASDAR TTSI project no. 8102 (coordinated by the Chambre Régionale d'Agriculture de Midi-Pyrénées). Data analysis was performed in the framework of the BAG'AGES and BAG'AGES CISOL projects and financed by the Agence de l'Eau Adour-Garonne and Occitanie Region. We thank Isabelle Cousin (INRAE UR Sol, France) for sharing the SOLHYDRO database, which allowed us to acquire complementary results.

References

Abdollahi, L., Munkholm, L.J., 2014. Tillage System and Cover Crop Effects on Soil Quality: I. Chemical, Mechanical, and Biological Properties. Soil Sci. Soc. Am. J. 78, 262–270.

https://doi.org/10.2136/sssaj

Ahuja, L.R., Fiedler, F., Dunn, G.H., Benjamin, J.G., Garrison, A., 1998. Changes in Soil Water Retention Curves Due to Tillage and Natural Reconsolidation. Soil Sci. Soc. Am. J. 62, 1228–1233.

Akaike, H., Parzen, E., Tanabe, K., Kitagawa, G., 1998. Selected Papers of Hirotugu Akaike. Springer Science & Business Media.

Al Majou, H., Bruand, A., Duval, O., 2008a. The use of in situ volumetric water content at field capacity to improve the prediction of soil water retention properties. Can. J. Soil Sci. 88, 533–541. https://doi.org/10.4141/CJSS07065

Al Majou, H., Bruand, A., Duval, O., Cousin, I., 2007. Comparaison de fonctions de pédotransfert nationales et européennes pour prédire les propriétés de rétention en eau des sols. Etude Gest. des Sols 14, 103–116.

Al Majou, H., Bruand, A., Duval, O., Le Bas, C., Vautier, A., 2008b. Prediction of soil water retention properties after stratification by combining texture, bulk density and the type of horizon. Soil Use Manag. 24, 383–391. https://doi.org/10.1111/j.1475-2743.2008.00180.x

Alletto, L., Coquet, Y., Benoit, P., Heddadj, D., Barriuso, E., 2010. Tillage management e ff ects on pesticide fate in soils. A review. Agron. Sustain. Dev. 30, 367–400.

Baumhardt, R.L., Lascano, R.J., 1999. Water Budget and Yield of Dryland Cotton Intercropped with Terminated Winter Wheat. Agron. J. 91, 922–927.

Bescansa, P., Imaz, M.J., Virto, I., Enrique, A., Hoogmoed, W.B., 2006. Soil water retention as affected by tillage and residue management in semiarid Spain. Soil Tillage Res. 87, 19–27. https://doi.org/10.1016/j.still.2005.02.028

Blanco-Canqui, H., Gantzer, C.J., Anderson, S.H., Alberts, E.E., 2004. Tillage and Crop Influences on Physical Properties for an Epiaqualf. Soil Sci. Soc. Am. J. 68, 567–576. https://doi.org/10.2136/sssaj2004.5670

Borgesen, C.D., Schaap, M.G., 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. Geoderma 127, 154–167. https://doi.org/10.1016/j.geoderma.2004.11.025

Bruand, A., Duval, O., Cousin, I., 2004. Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO: Une première proposition combinant le type d'horizon, sa texture et sa densité apparente. Étude Gest. des Sols 11, 323–334.

Bruand, A., Pérez Fernández, P., Duval, O., 2003. Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. Soil Use Manag. 19, 232–242. https://doi.org/10.1079/SUM2003196

Cameira, M.R., Fernando, R.M., Pereira, L.S., 2003. Soil macropore dynamics affected by tillage and irrigation for a silty loam alluvial soil in southern Portugal. Soil Tillage Res. 70, 131–140. https://doi.org/10.1016/S0167-1987(02)00154-X

Chen, Y., Tessier, S., Gallichand, J., 1998. Estimates of tillage effects on saturated hydraulic conductivity. Can. Agric. Eng. 40, 169–177.

Cornelis, W.M., Ronsyn, J., Meirvenne, M. Van, Hartmann, R., 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. Soil Sci. Soc. Am. J. 65, 638–648.

Devine, S., Markewitz, D., Hendrix, P., Coleman, D., 2014. Soil aggregates and associated organic matter under conventional tillage, no-tillage, and forest succession after three decades. PLoS One 9, 1–12. https://doi.org/10.1371/journal.pone.0084988

FAO, 2016. Conservation Agriculture, in: Save and Grow in Practice: Maize, Rice and Wheat.

Gey, S., Poggi, J., 2006. Boosting and instability for regression trees. Comput. Stat. Data Anal. 50, 533–550. https://doi.org/10.1016/j.csda.2004.09.001

Ghanbarian, B., Taslimitehrani, V., Dong, G.Z., Pachepsky, Y.A., 2015. Sample dimensions effect on prediction of soil water retention curve and saturated hydraulic conductivity. Journal of Hydrology 528, 127-137.

Ghorbani Dashtaki, S., Homaee, M., Khodaverdiloo, H., 2010. Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data. Soil Use Manag. 26, 68–74. https://doi.org/10.1111/j.1475-2743.2009.00254.x

Gozubuyuk, Z., Sahin, U., Ozturk, I., Celik, A., Adiguzel, M.C., 2014. Tillage effects on certain physical and hydraulic properties of a loamy soil under a crop rotation in a semi-arid region with a cool climate. Catena 118, 195–205. https://doi.org/10.1016/j.catena.2014.01.006

Green, T.R., Ahuja, L.R., Benjamin, J.G., 2003. Advances and challenges in predicting agricultural management effects on soil hydraulic properties. Geoderma 116, 3–27. https://doi.org/10.1016/S0016-7061(03)00091-0

Hastie, T., Tibshirani, R., Firedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second. ed. Springer Series in Statistics.

Hill, R.L., Horton, R., Cruse, R.M., 1985. Tillage effects on soil water retention and pore size distribution of two mollisols. Soil Sci. Soc. Am. J. 49, 1264–1270.

Hillel, D., 1971. Soil and Water: Physical Principles and Processes, Academic Press.

Hobbs, P.R., Sayre, K., Gupta, R., 2008. The role of conservation agriculture in sustainable agriculture. Philos. Trans. R. Soc. B Biol. Sci. 363, 543–555. https://doi.org/10.1098/rstb.2007.2169

IUSS Working Group WRB, 2015. World Reference Base for Soil Resources 2014: International soil classification system for naming soils and creating legends for soil maps, World Soil. ed.

Kay, B.D., VandenBygaart, A.J., 2002. Conservation tillage and depth stratification of porosity and soil organic matter. Soil Tillage Res. 66, 107–118.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R package. R News 2, 18–22. https://doi.org/10.1023/A

Ließ, M., Glaser, B., Huwe, B., 2012. Geoderma Uncertainty in the spatial prediction of soil texture Comparison of regression tree and Random Forest models. Geoderma 170, 70–79. https://doi.org/10.1016/j.geoderma.2011.10.010

Lilly, A., Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2008. Probabilistic Approach to the Identifi cation of Input Variables to Estimate Hydraulic Conductivity. Soil Sci. Soc. Am. J. 72, 16–24. https://doi.org/10.2136/sssaj2006.0391

Mapa, R.B., Green, R.E., Santo, L., 1986. Temporal Variability of Soil Hydraulic Properties with Wetting and Drying Subsequent to Tillage. Soil Sci. Soc. Am. J. 50, 1133–1138.

Nakano, K., Miyazaki, T., 2005. Predicting the saturated hydraulic conductivity of compacted subsoils using the non-similar media concept. Soil Tillage Res. 84, 145–153. https://doi.org/10.1016/j.still.2004.11.010

Nash, J.E., Sutcliffe, J. V, 1970. River Flow Forecasting Through Conceptual Models Part I-A Discussion of Principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67, 1093–1102. https://doi.org/10.2136/sssaj2003.1093

Obi, J.C., Ogban, P.I., Ituen, U.J., Udoh, B.T., 2014. Development of pedotransfer functions for coastal plain soils using terrain attributes. Catena 123, 252–262.

https://doi.org/10.1016/j.catena.2014.08.015

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181–199. https://doi.org/10.1007/s10021-005-0054-1

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rawls, W.J., Brakensiek, D.L., Miller, N., 1983. Green-ampt Infiltration Parameters from Soils Data. J. Hydraul. Eng. 109, 62–70. https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(62)

Rawls, W.J., Brakensiek, D.L., Saxton, K.E., 1982. Estimation of Soil Water Properties. Trans. ASAE. https://doi.org/10.13031/2013.33720

Rawls, W.J., Pachepsky, Y.A., 2002. Soil Consistence and Structure as Predictors of Water Retention. Soil Sci. Soc. Am. J. 66, 1115–1126.

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

Rubio, C.M., Llorens, P., Gallart, F., 2008. Uncertainty and efficiency of pedotransfer functions for estimating water retention characteristics of soils. Eur. J. Soil Sci. 59, 339–347. https://doi.org/10.1111/j.1365-2389.2007.01002.x

Shipitalo, M.J., Dick, W.A., Edwards, W.M., 2000. Conservation tillage and macropore factors that affect water movement and the fate of chemicals. Soil Tillage Res. 53, 167–183.

Silva, M.L.D., Libardi, P.L., Gimenes, F.H.S., 2018. Soil Water Retention Curve as Affected by Sample Height. Revista Brasileira De Ciencia Do Solo 42.

Strudley, M.W., Green, T.R., Ascough II, J.C., 2008. Tillage effects on soil hydraulic properties in space and time : State of the science. Soil Tillage Res. 99, 4–48. https://doi.org/10.1016/j.still.2008.01.007

Therneau, T., Atkinson, B., 2019. Recursive Partitioning and Regression Trees. R package.

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

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

Tranter, G., Mcbratney, A.B., Minasny, B., 2009. Using distance metrics to determine the appropriate domain of pedotransfer function predictions. Geoderma 149, 421–425. https://doi.org/10.1016/j.geoderma.2009.01.006

van Genuchten, M.T., 1980. A Closed-form Equation For Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Sci. Soc. Am. J. 44, 892–898.

van Genuchten, M.T., Leij, F.J., Yates, S.R., 1991. The RETC Code for Quantifying the Hydraulic Functions of Unsaturated Soils. EPA/600/2-91/065.

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., van Genuchten, M.T., 2010. Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil Hydraulic Properties: A Review. Vadose Zo. J. 9, 795–820. https://doi.org/10.2136/vzj2010.0045

Verhulst, N., François, I., Govaerts, B., 2010. Conservation agriculture, improving soil quality for sustainable production systems? Adv. soil Sci. food Secur. soil Qual. 1799267585, 137–208.

Vos, C., Don, A., Hobley, E.U., Prietz, R., Heidkamp, A., Freibauer, A., 2019. Factors controlling the variation in organic carbon stocks in agricultural soils of Germany. Eur. J. Soil Sci. 70, 550–564. https://doi.org/10.1111/ejss.12787

Williams, S.M., Weil, R.R., 2004. Crop Cover Root Chanels May Alleviate Soil Compaction Effects On Soybean Crop. Soil Sci. Soc. Am. J. 68, 1403–1409.

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185. https://doi.org/10.1016/S0016-7061(98)00132-3

Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251, 123–150. https://doi.org/10.1016/S0022-1694(01)00464-4

Yang, R., Zhang, G., Liu, F., Lu, Y., Yang, Fan, Yang, Fei, Yang, M., Zhao, Y.-G., Li, D.-C., 2016. Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. Ecol. Indic. 60, 870–878.

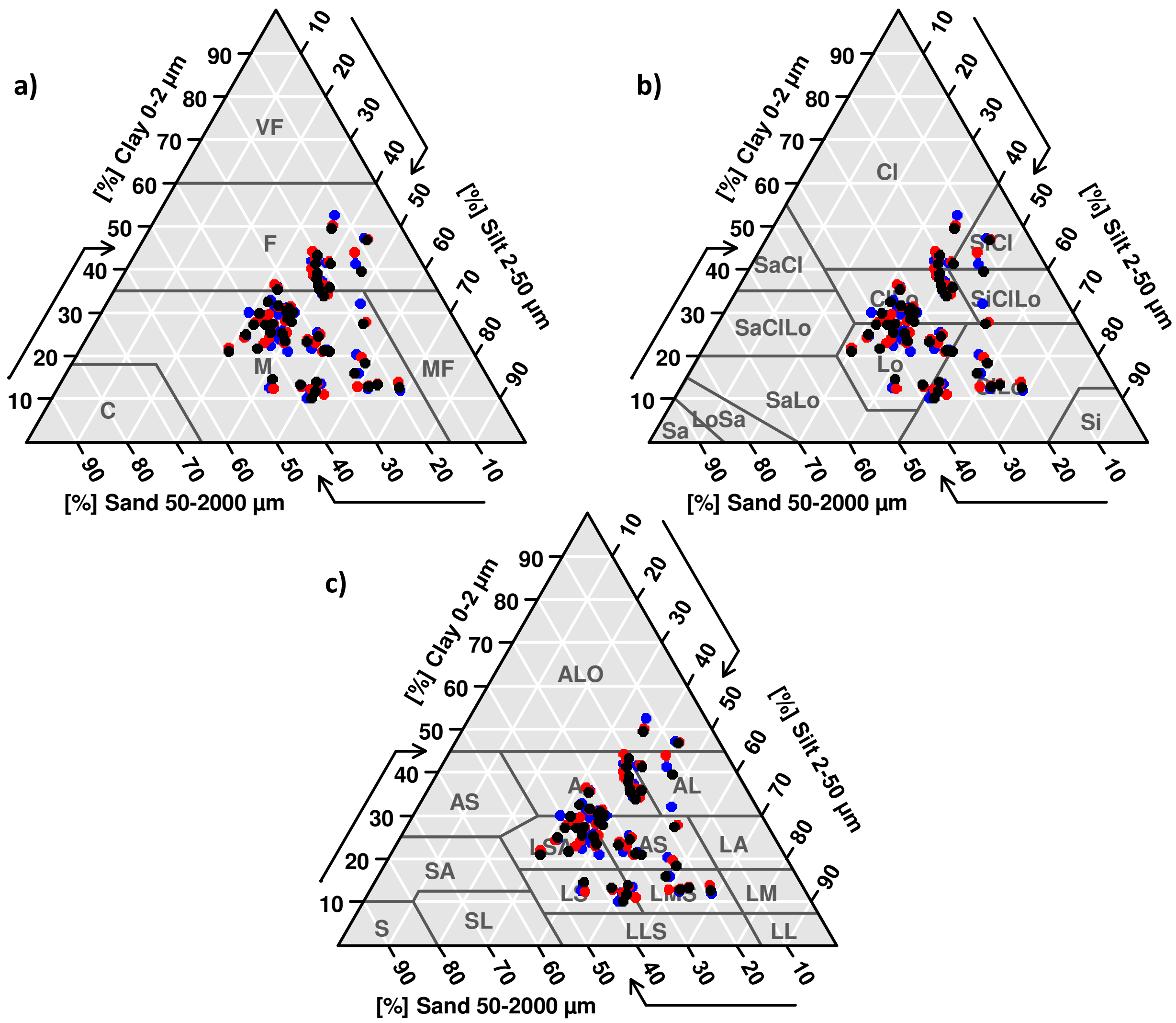

Fig. 1. Textures of the soil samples collected in 0-10, 10-20 and 20-30 cm deep soil layers (black, red and blue dots, respectively) in 61 conservation agriculture fields. The three texture triangles are based on the three classifications used to develop the pedotransfer functions found in the literature a) FAO, b) USDA and c) AISNE.

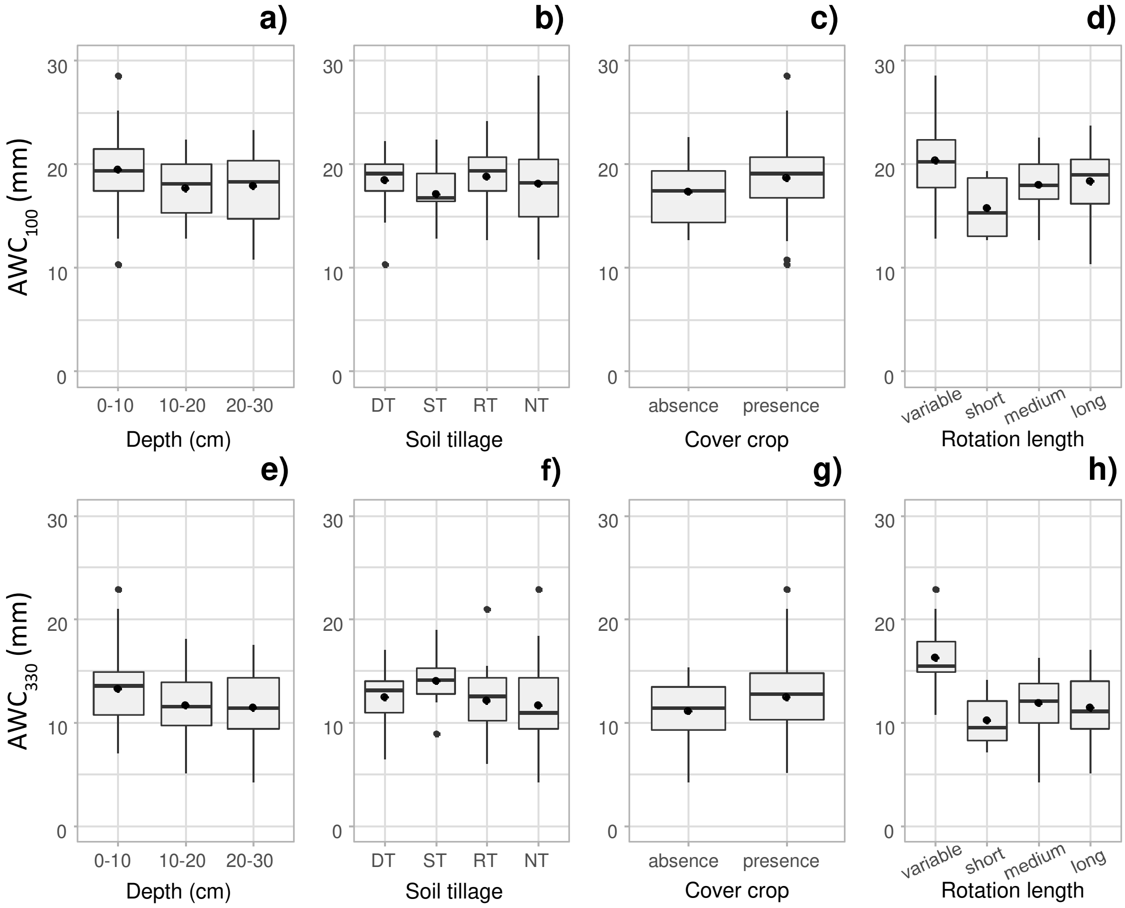

Fig. 2. Available water capacity (AWC) predicted assuming field capacity at a volumetric water content of -100 cm (AWC_{100}) or -330 cm (AWC_{330}) of matric head as a function of soil depth (a, e), soil tillage (b, f), cover-crop presence (c, g) and rotation length (d, h). For soil tillage, DT: deep tillage, RT: reduced tillage, ST: strip-till and NT: no tillage. For rotation length, variable: not fixed, short: ≤ 2 years, medium: > 2 years & ≤ 4 years, long: > 4 years.

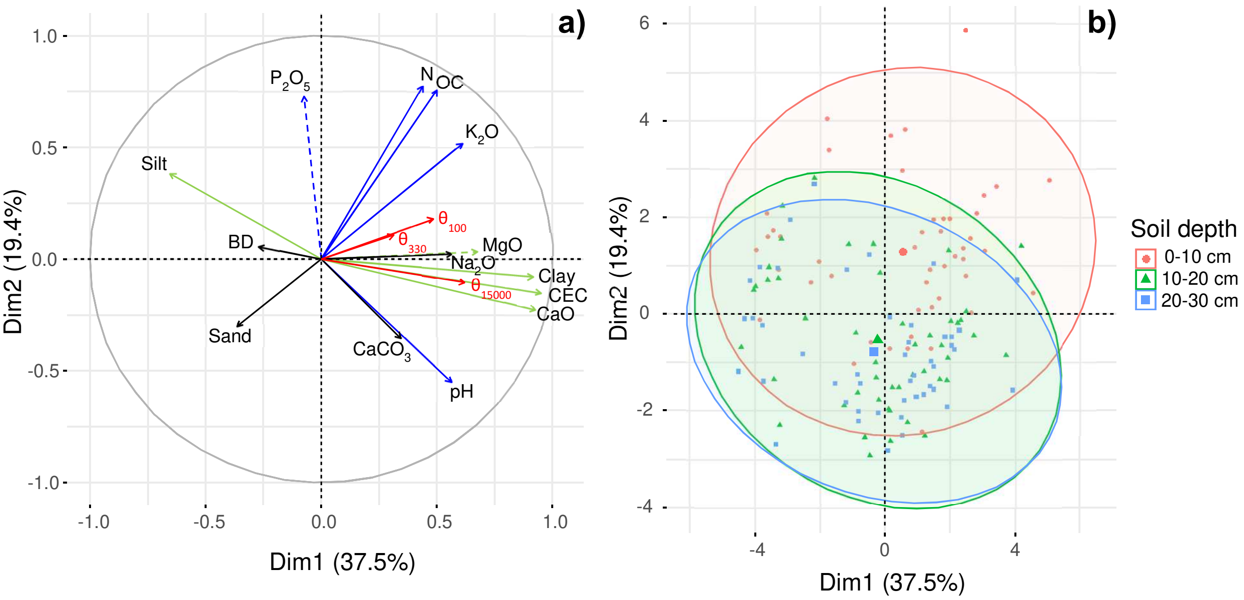

Fig. 3. Correlation circle of the (a) variables and (b) soil layers on the first two dimensions of the principal component analysis. (a) The variables that contributed significantly to the first and second axis are green and blue, respectively. Dashed arrows correspond to variables that did not contribute significantly to the first two axes. Black arrows correspond to variables that did not contribute significantly to any of the axes. Red arrows correspond to volumetric water contents at -100 cm (θ_{100}), -330 cm (θ_{330}) and -15 000 cm ($\theta_{15\,000}$) of matric head, which were not used to construct the axes. (b) Soil layers are coloured by depth, circles represent 95% confidence interval ellipses and larger symbols are centroids.

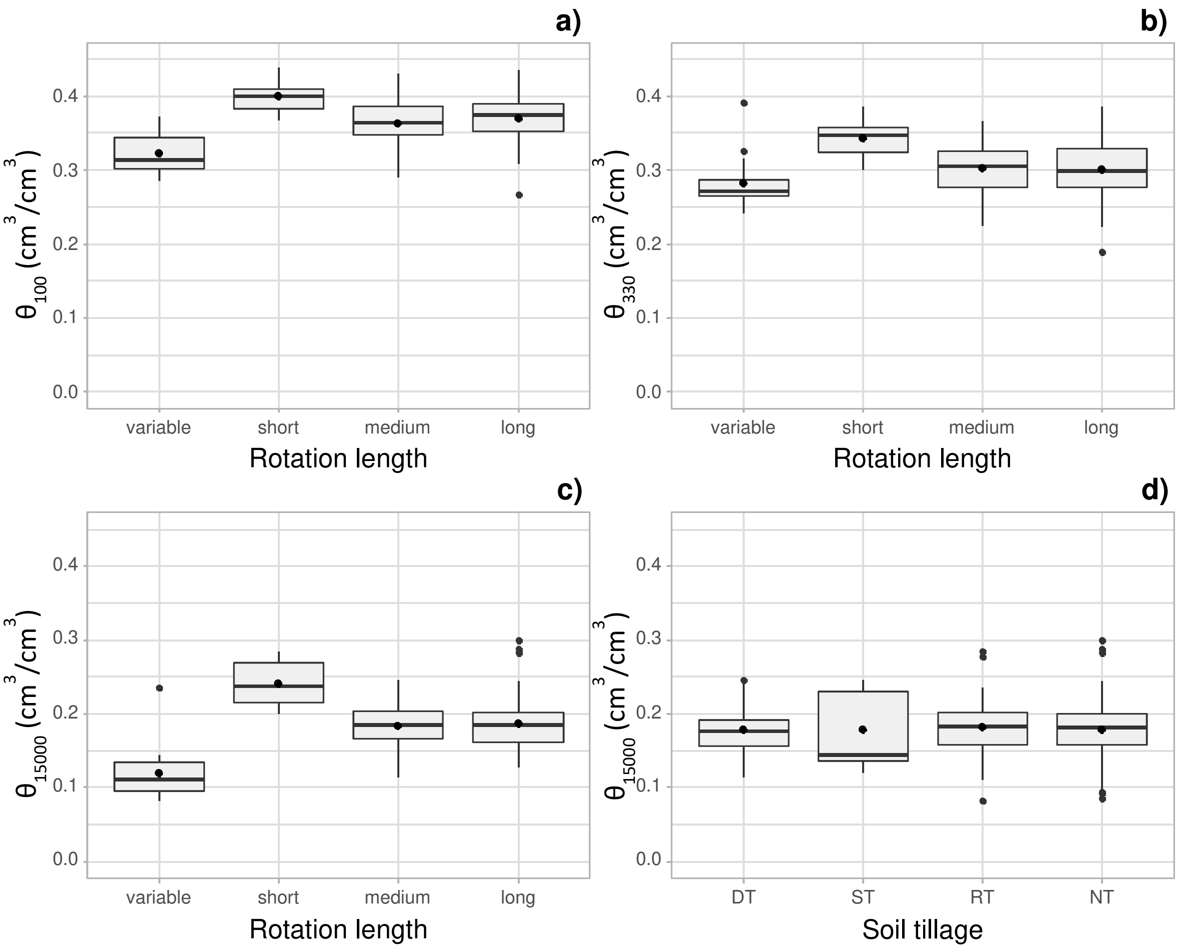

Fig. 4. Volumetric water content at (a) -100 cm (θ_{100}), (b) -330 cm (θ_{330}) and (c) -15 000 cm (θ_{15000}) of matric head as a function of rotation length (a, b, c) and (d) at -15 000 cm ($\theta_{15\,000}$) as a function of soil tillage. For soil tillage, DT: deep tillage, RT: reduced tillage, ST: strip-till and NT: no tillage. For length of rotation, variable: not fixed, short: ≤ 2 years, medium: > 2 years & ≤ 4 years, long: > 4 years.

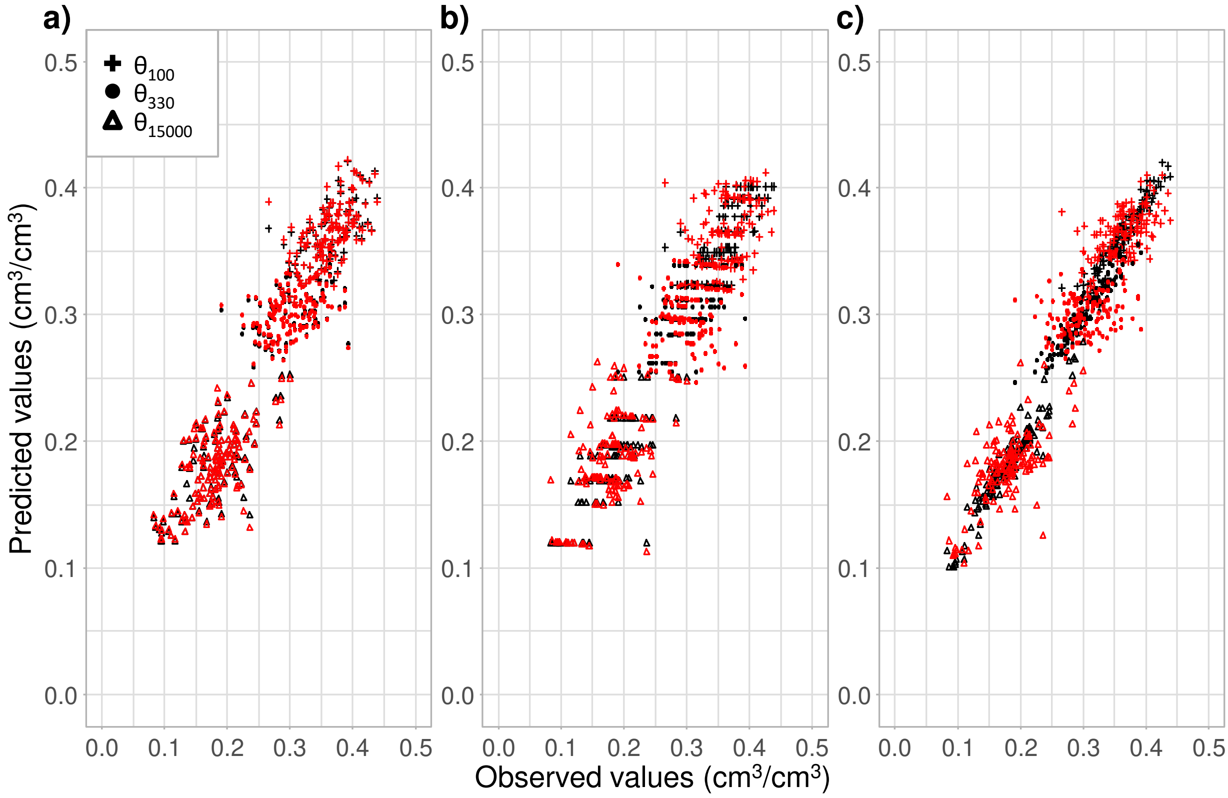

Fig. 5. Observed vs. predicted soil water content at -100 cm (θ_{100}), -330 cm (θ_{330}) and -15 000 cm (θ_{15000}) of matric head for (a) multiple linear regression, (b) regression tree and (c) random forest. Predicted adjustment values and cross-validation values are black and red, respectively.

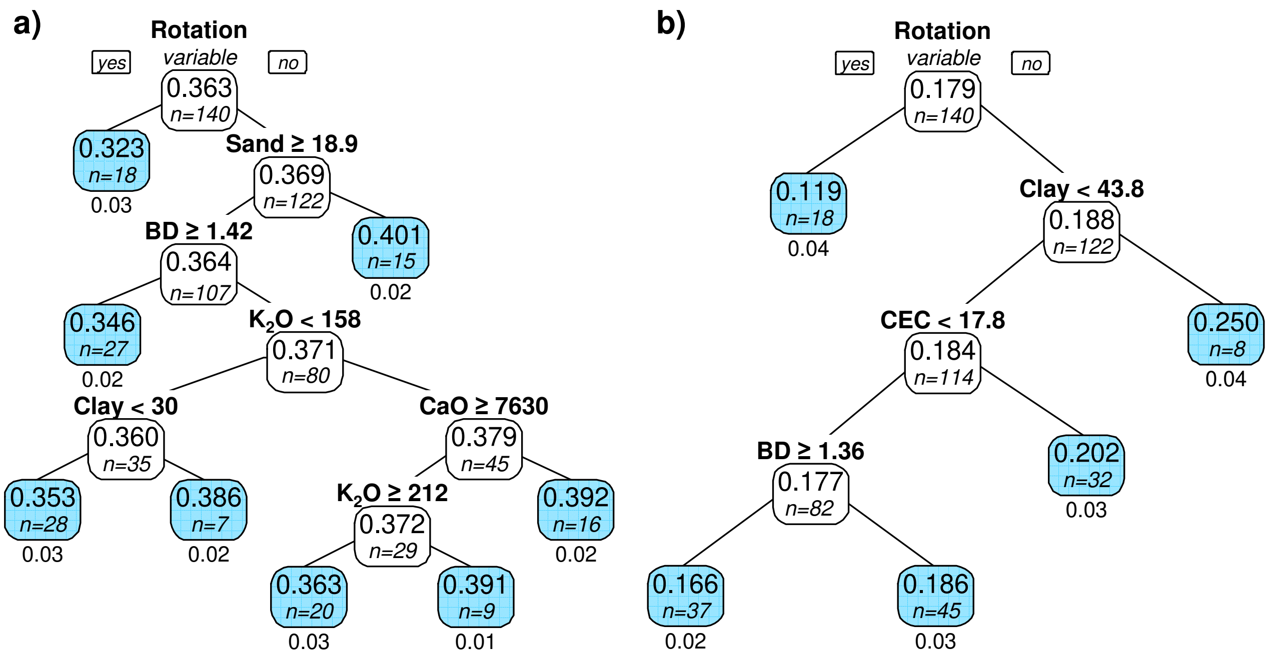

Fig. 6. Regression trees for the prediction of (a) θ_{100} and (b) $\theta_{15\,000}$. BD is bulk density (g/cm^3), CEC is cation exchange capacity (cmol/kg) and CaO and K_2O are exchangeable calcium and potassium (mg/kg), respectively. Values in boxes are mean water contents (cm^3/cm^3) of the n samples in the partition. The values below terminal leaves (blue boxes) are standard deviations.

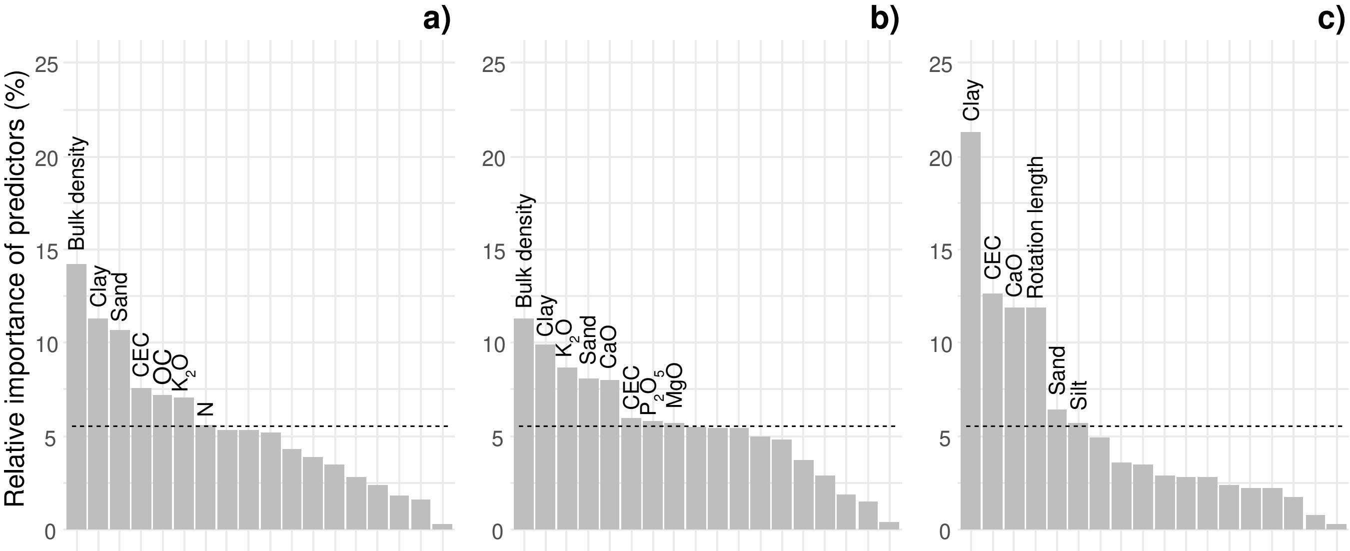

Fig. 7. Relative importance (%) of predictors in random forests of (a) θ_{100}, (b) θ_{330} and (c) θ_{15000}. Dashed lines represent the mean relative importance; only predictors above the mean are labelled. CEC is cation exchange capacity; CaO, K_2O and MgO are exchangeable calcium, potassium and magnesium, respectively; P_2O_5, OC and N are phosphorus, organic carbon and nitrogen content, respectively.

Table 1. Published pedotransfer functions (PTFs) used to predict soil volumetric water content (cm^3/cm^3) at a given matric head h=-100 cm, θ_{100}, h= -330 cm, θ_{330}, and h=-15 000 cm, $\theta_{15\,000}$. Cl, Si, OC and OM are contents (%) of clay, silt, organic carbon and organic matter, respectively. OC*=OC+1. BD is bulk density (g/cm^3). Co: continuous-PTFs, Cl: class-PTFs, Tr: tree-PTFs. When two PTFs are indicated in the PTF ID column, the first does not consider topsoil/subsoil separation, and the second considers only the topsoil.

Reference	Sampling location	N	Predictive variables / Equation		Variables predicted	PTF ID	
Rawls et al. (1982)	USA, 32 states	5350	$\theta_h = a + (b \times Sa) + (c \times Si) + (d \times Cl) + (e \times OM) + (f \times BD)$		θ_{40}, θ_{70}, θ_{100}, θ_{200}, (θ_{330}), θ_{600}, θ_{4000}, θ_{7000}, θ_{10000}, ($\theta_{15\,000}$)	M1_Co	
			$\theta_h = a + (b \times Sa) + (c \times Si) + (d \times Cl) + (e \times OM) + (f \times BD) + (h \times \theta_{15\,000})$			M2_Co	
			$\theta_h = a + (b \times Sa) + (c \times Si) + (d \times Cl) + (e \times OM) + (f \times BD) + (g \times \theta_{330}) + (h \times \theta_{15\,000})$			M3_Co	
Bruand et al. (2004)	France, Paris basin	340	- texture AISNE	(topsoil function)	θ_{10}, θ_{33}, θ_{100}, θ_{330}, θ_{1000}, θ_{3300}, $\theta_{15\,000}$	M4_Cl	
Al Majou et al. (2007)	France, Paris basin	320	- texture FAO	(topsoil function)	θ_{10}, θ_{33}, θ_{100}, θ_{330}, θ_{1000}, θ_{3300}, $\theta_{15\,000}$	M5_Cl	M6_Cl
			- texture FAO - bulk density	(topsoil function)		M7_Cl	M8_Cl
			- $\theta_h = a + (b \times Cl) + (c \times Si) + (d \times OC) + (e \times BD)$	(topsoil function)		M9_Co	M10_Co
Al Majou et al. (2008b)	France, Paris basin, Brittany, the western coastal marshlands and the Pyrenean piedmont plain	456	- texture FAO	(topsoil function)	θ_{10}, θ_{33}, θ_{100}, θ_{330}, θ_{1000}, θ_{3300}, $\theta_{15\,000}$	M11_Cl	M12_Cl
			- texture FAO - bulk density	(topsoil function)		M13_Cl	M14_Cl
Tóth et al. (2015)	18 European countries	18 537	- texture FAO & topsoil/subsoil		θ_{330}, $\theta_{15\,000}$	M15_Tr	
			- texture USDA & topsoil/subsoil			M16_Tr	
			$\theta_{330} = a_1 - (b_1 \times OC^{*-1}) + (c_1 \times Cl) + (d_1 \times Si) + (e_1 \times Si \times OC^{*-1}) - (f_1 \times Si \times Cl) + (g_1 \times Cl \times OC^{*-1})$ $\theta_{15\,000} = a_2 + (b_2 \times Cl) - (c_2 \times Si) - (d_2 \times OC^{*-1}) + (e_2 \times Si \times Cl) + (f_2 \times Cl \times OC^{*-1}) + (g_2 \times Si \times OC^{*-1})$			M17_Co	
Roman Dobarco et al. (2019)	France, northern half of the country, with little representation of more mountainous southern and eastern regions	689	- texture FAO	(topsoil function)	θ_{100}, θ_{330}, $\theta_{15\,000}$	M18_Cl	M19_Cl
			- texture FAO - bulk density	(topsoil function)		M20_Cl	M21_Cl
			$\theta_h = a + (b \times Cl) + (c \times Sa)$	(topsoil function)		M22_Co	M26_Co
			$\theta_h = a + (b \times Cl) + (c \times Sa) + (d \times OC)$	(topsoil function)		M23_Co	M27_Co
			$\theta_h = a + (b \times Cl) + (c \times Sa) + (e \times BD)$	(topsoil function)		M24_Co	M28_Co
			$\theta_h = a + (b \times Cl) + (c \times Sa) + (d \times OC) + (e \times BD)$	(topsoil function)		M25_Co	M29_Co

Table 2. Published pedotransfer functions (PTFs) used to evaluate the quality of prediction of the van Genuchten's water-retention curve parameters θ_s, α and n. Cl, Si, OC and OM are contents (%) of clay, silt, organic carbon and organic matter, respectively. OC*=OC+1. BD is bulk density (g/cm^3), CEC is cation exchange capacity (cmol/kg), T/S is topsoil/subsoil (T=1, S=0). θ_s is volumetric water content at saturation, α and n are shape parameters of van Genuchten's water retention curve.

Reference	Sampling location	Number of samples	Predictive variables / Equation	Predicted variables	PTF ID
Wösten et al. (1999)	12 European countries	4030	- texture FAO		P1_Cl
			$\theta_s = a_1 + (b_1 \times Cl) - (c_1 \times BD) - (d_1 \times Si^2) + (e_1 \times OM^2) + (f_1 \times Cl^{-1}) + (g_1 \times Si^{-1}) + (h_1 \times \ln(Si)) - (i_1 \times OM \times Cl) - (j_1 \times BD \times Cl) - (k_1 \times BD \times OM) - (l_1 \times T/S \times Si)$ $\ln(\alpha) = - a_2 + (b_2 \times Cl) + (c_2 \times Si) + (d_2 \times OM) + (e_2 \times BD) - (f_2 \times T/S) - (g_2 \times BD2) - (h_2 \times Cl^2) - (i_2 \times (OM^2)) + (j_2 \times OM^{-1}) + (k_2 \times \ln(Si)) + (l_2 \times \ln(OM)) - (m_2 \times BD \times Si) - (n_2 \times BD \times OM) + (o_2 \times T/S \times Cl)$ $\ln(n\text{-}1) = - a_3 - (b_3 \times Cl) + (c_3 \times Si) - (d_3 \times OM) + (e_3 \times BD) - (f_3 \times (BD^2)) + (g_3 \times (Cl^2)) + (h_3 \times (OM^2)) - (i_3 \times BD^{-1}) - (j_3 \times Si^{-1}) - (k_3 \times OM^{-1}) - (l_3 \times \ln(Si)) - (m_3 \times \ln(OM)) - (n_3 \times \ln(BD)) - (o_3 \times BD \times Cl) + (p_3 \times BD \times OM) + (q_3 \times T/S \times Cl)$		P2_Co
Al Majou et al. (2008a)	France, Paris basin	320	- texture FAO		P3_Cl
			$\theta_s = a_1 – (b_1 \times Cl) – (c_1 \times BD) + (d_1 \times Si^2) – (e_1 \times OC^2) + (f_1 \times Cl^{-1}) + (g_1 \times Si^{-1}) – (h_1 \times \ln(Si)) + (i_1 \times OC \times Cl) + (j_1 \times BD \times Cl) – (k_1 \times BD \times OC) – (l_1 \times Si)$ $\ln(\alpha) = a_2 + (b_2 \times Cl) + (c_2 \times Si) + (d_2 \times OC) + (e_2 \times BD) – (f_2 \times BD^2) – (g_2 \times Cl^2) – (h_2 \times OC^2) – (i_2 \times OC^{-1}) – (j_2 \times \ln(Si)) – (k_2 \times \ln(OC)) – (l_2 \times BD \times Si) – (m_2 \times BD \times OC)$ $\ln(n\text{-}1) = - a_3 – (b_3 \times Cl) + (c_3 \times Si) – (d_3 \times OC) + (e_3 \times BD) – (f_3 \times BD^2) + (g_3 \times Cl^2) + (h_3 \times OC^2) + (i_3 \times BD^{-1}) + (j_3 \times Si^{-1}) + (k_3 \times OC^{-1}) – (l_3 \times \ln(Si)) + (m_3 \times \ln(OC)) – (n_3 \times \ln(BD)) + (o_3 \times BD \times Cl) + (p_3 \times BD \times OC)$		P4_Co
Tóth et al (2015)	18 European countries	18 537	- texture FAO	θ_s, α, n	P5_Cl
			- texture USDA		P6_Cl
			$\theta_s = 0.5056 – (0.1437 \times 1/(OC+1)) + (0.0004152 \times Si)$ $\log10(\alpha) = -1.3050 – (0.0006123 \times Si) – (0.009810 \times Cl) + (0.07611 \times 1/(OC^*)) – (0.0004508 \times Si \times Cl) + (0.03472 \times Cl \times 1/(OC^*)) – (0.01226 \times Si \times 1/(OC+1))$ $\log10(n\text{-}1) = 0.01516 – (0.005775 \times 1/OC^*) – (0.24885 \times \log10(CEC)) – (0.01918 \times Cl) – (0.0005052 \times Si) – (0.007544 \times pH^2) – (0.02159 \times Cl \times 1/OC^*) + (0.01556 \times Cl \times \log10(CEC)) + (0.01477 \times 1/OC^* \times pH^2) +(0.0001121 \times Si \times Cl) – (0.33198 \times 1/OC^* \times \log10(CEC))$		P7_Co
			$\theta_s = 0.83080 – (0.28217 \times BD) + (0.0002728 \times Cl) + (0.000187 \times Si)$ $\log10(\alpha) = -0.43348 – (0.41729 \times BD) – (0.04762 \times OC) + (0.21810 \times T/S) – (0.01581 \times Cl) – (0.01207 \times Si)$ $\log10(n\text{-}1) = 0.22236 – (0.30189 \times BD) – (0.05558 \times T/S) – (0.005306 \times Cl) – (0.003084 \times Si) – (0.01072 \times OC)$		P8_Co
			$\theta_s = 0.63052 – (0.10262 \times BD^2) + (0.0002904 \times pH^2) + (0.0003335 \times Cl)$ $\log10(\alpha) = -1.16518 + (0.40515 \times 1/OC^*) – (0.16063 \times BD^2) – (0.008372 \times Cl) – (0.01300 \times Si) + (0.002166 \times pH^2) + (0.08233 \times T/S)$ $\log10(n\text{-}1) = -0.25929 + (0.25680 \times 1/OC^*) – (0.10590 \times BD^2) – (0.009004 \times Cl) – (0.001223 \times Si)$		P9_Co

Table 3. Summary statistics of particle size fractions (%), organic carbon (OC; %), nitrogen content (g/kg), bulk density (BD; g/cm^3), cation exchange capacity (CEC; cmol/kg), exchangeable CaO, MgO, K_2O, Na_2O (mg/kg), pH, total calcium carbonate $CaCO_3$ (g/kg), phosphorus content P_2O_5 (mg/kg) and volumetric water content at field capacity, θ_{100} and θ_{330}, and at wilting point, $\theta_{15\,000}$ (cm^3/cm^3) of the dataset used to evaluate published pedotransfer functions (PTFs) and develop new PTFs

N=140	Clay	Silt	Sand	OC	N	BD	CEC	CaO	MgO	K_20	Na_2O	pH	$CaCO_3$	P_2O_5	θ_{100}	θ_{330}	$\theta_{15\,0000}$
Mean	27.8	42.2	30.0	1.0	1.1	1.4	13.2	6700	236.0	189.3	13.4	7.6	41.0	35.2	0.363	0.301	0.179
Standard deviation	10.1	9.1	9.0	0.3	0.3	0.1	5.6	3274	135.6	103.3	6.1	0.9	54.1	29.1	0.035	0.037	0.042
Min	10.3	29.4	8.0	0.5	0.6	1.2	3.5	540	47.2	27.8	4.3	5.1	0.0	3.0	0.266	0.190	0.083
Median	28.0	39.1	31.9	1.0	1.0	1.4	13.4	6966	211.4	171.2	12.2	8.1	19.0	27.0	0.364	0.298	0.182
Max	52.6	68.7	49.0	2.2	2.2	1.7	24.6	13057	595.4	522.8	35.7	8.7	220.0	147.0	0.439	0.392	0.300

Table 4. Statistical criteria for the prediction of θ_{100}, θ_{330} and $\theta_{15\,000}$. $RMSE_P$: root mean squared error of prediction, ME_P: mean error of prediction, EF_P: Nash-Sutcliffe Efficiency of prediction. Co: continuous-PTFs, Cl: class-PTFs, Tr: tree-PTFs. 0.000 means < 1.10^{-3}

	θ_{100} (cm^3/cm^3)			θ_{330} (cm^3/cm^3)			$\theta_{15\,000}$ (cm^3/cm^3)		
PTF	**$RMSE_P$**	**ME_P**	**EF_P**	**$RMSE_P$**	**ME_P**	**EF_P**	**$RMSE_P$**	**ME_P**	**EF_P**
M1_Co	0.089	0.078	-5.68	0.073	0.049	-2.84	0.047	0.017	-0.29
M2_Co	0.262	0.259	-56.75	0.059	0.047	-1.53	-	-	-
M3_Co	0.034	-0.007	0.03	-	-	-	-	-	-
M4_Cl	0.065	-0.056	-2.24	0.037	-0.014	-0.11	0.040	-0.027	0.17
M5_Cl	0.056	-0.042	-1.60	0.042	-0.007	-0.28	0.044	0.003	-0.14
M6_Cl	0.052	-0.039	-1.26	0.038	0.002	-0.02	0.046	-0.013	-0.24
M7_Cl	0.049	-0.028	-1.01	0.045	0.001	-0.47	0.051	-0.004	-0.50
M8_Cl	0.042	-0.029	-0.48	0.038	0.009	-0.02	0.044	-0.01	-0.13
M9_Co	0.053	-0.044	-1.36	0.047	-0.025	-0.59	0.043	-0.023	-0.08
M10_Co	0.116	-0.112	-10.40	0.080	-0.069	-3.59	0.057	-0.045	-0.87
M11_Cl	0.057	-0.043	-1.73	0.042	-0.007	-0.28	0.045	0.011	-0.20
M12_Cl	0.049	-0.030	-1.04	0.043	0.009	-0.31	0.049	0.011	-0.38
M13_Cl	0.046	-0.028	-0.78	0.043	0.003	-0.31	0.050	0.008	-0.45
M14_Cl	0.039	-0.019	-0.30	0.044	0.017	-0.35	0.052	0.017	-0.56
M15_Tr	-	-	-	0.043	0.022	-0.33	0.045	-0.015	-0.20
M16_Tr	-	-	-	0.054	0.036	-1.07	0.042	0.001	-0.03
M17_Co	-	-	-	0.045	0.025	-0.45	0.036	0.000	0.23
M18_Cl	0.059	-0.045	-1.88	0.047	0.001	-0.57	0.050	0.018	-0.48
M19_Cl	0.051	-0.033	-1.20	0.041	0.014	-0.24	0.047	0.010	-0.29
M20_Cl	0.048	-0.031	-0.95	0.052	0.018	-0.91	0.056	0.020	-0.80
M21_Cl	0.042	-0.025	-0.50	0.041	0.016	-0.23	0.049	0.013	-0.37
M22_Co	0.066	-0.057	-2.63	0.04	0.005	-0.16	0.034	0.003	0.31
M23_Co	0.061	-0.051	-2.08	0.046	-0.011	-0.51	0.035	0.003	0.29
M24_Co	0.055	-0.046	-1.54	0.041	-0.001	-0.22	0.035	0.001	0.30
M25_Co	0.055	-0.045	-1.51	0.042	-0.002	-0.28	0.036	0.001	0.25
M26_Co	0.059	-0.048	-1.89	0.043	-0.007	-0.33	0.036	0.002	0.23
M27_Co	0.065	-0.056	-2.52	0.044	-0.016	-0.37	0.036	-0.002	0.25
M28_Co	0.060	-0.049	-2.02	0.042	-0.005	-0.25	0.036	0.002	0.24
M29_Co	0.074	-0.064	-3.55	0.048	-0.022	-0.65	0.038	-0.007	0.16

Table 5. Statistical criteria for the prediction of θs, α and n parameters**.** $RMSE_P$: root mean squared error of prediction, ME_P: mean error of prediction, EF_P: Nash-Sutcliffe efficiency of prediction. Co: continuous-PTFs, Cl: class-PTFs

	θs (cm^3/cm^{-3})			α (cm^{-1})			n (-)		
PTF	**$RMSE_P$**	**ME_P**	**EF_P**	**$RMSE_P$**	**ME_P**	**EF_P**	**$RMSE_P$**	**ME_P**	**EF_P**
P1_Cl	0.054	0.029	-2.05	0.232	-0.018	-0.01	0.331	-0.132	-0.23
P2_Co	0.439	-0.438	-198.53	0.232	-0.019	0.00	0.333	-0.110	-0.25
P3_Cl	0.038	0.010	-0.49	0.506	0.441	-3.78	0.361	-0.201	-0.47
P4_Co	0.376	-0.373	-144.92	-	-	-	0.366	-0.197	-0.51
P5_Cl	0.046	0.033	-1.14	0.232	-0.017	-0.01	0.326	-0.119	-0.19
P6_Cl	0.05	0.036	-1.59	0.232	0.012	0.00	0.339	-0.125	-0.29
P7_Co	0.037	0.019	-0.41	0.234	-0.036	-0.02	0.325	-0.068	-0.19
P8_Co	0.035	0.029	-0.27	0.233	-0.033	-0.02	0.305	0.003	-0.04
P9_Co	0.038	0.033	-0.56	0.234	-0.035	-0.02	0.316	-0.051	-0.12

Table 6. Multiple linear regression coefficients for estimating θ_{100}, θ_{330} and $\theta_{15\,000}$ from the non-stratified dataset and the dataset for the top and bottom soil layers. θ is the soil volumetric water content (cm^3/cm^3) at a given matric head. Clay: clay content (%), Silt: silt content (%), Sand: sand content (%), BD: bulk density (g/cm^3), N: nitrogen content (g/kg), CEC: cation exchange capacity (cmol/kg) and P_2O_5: phosphorus content (mg/kg)

θ_{100} = a + b × Clay + c × BD + d × Silt + e × N + f × Sand						
	Intercept	Clay	BD	Silt	N	Sand
Coefficients	-9.809	1.04×10^{-1}	-1.24×10^{-1}	1.03×10^{-1}	2.37×10^{-2}	1.02×10^{-1}
θ_{330} = a + b × Clay + c × BD + d × CEC + e × P_2O_5						
	Intercept	Clay	BD	CEC	P_2O_5	
Coefficients	0.386	2.54×10^{-3}	-9.27×10^{-2}	-2.71×10^{-3}	1.72×10^{-4}	
$\theta_{15\,000}$ = a + b × Clay + c × BD + d × pH + e × Silt						
	Intercept	Clay	BD	pH	Silt	
Coefficients	0.145	2.56×10^{-3}	-8.56×10^{-2}	7.00×10^{-3}	6.24×10^{-4}	

Table 7. Statistical criteria (root mean squared error (RMSE, cm^3/cm^3), mean error (ME, cm^3/cm^3) and Nash-Sutcliffe efficiency (EF)) of the quality of adjustment (subscript $_A$) or cross validation (subscript $_{CV}$) for the prediction of θ_{100}, θ_{330} and $\theta_{15\,000}$ by new pedotransfer functions developed from the non-stratified dataset and datasets of the top and bottom soil layers. Values less than 0.001 are expressed as 0.

Criterion	Multiple linear regression			Regression tree			Random forest		
	θ_{100}	θ_{330}	$\theta_{15\,000}$	θ_{100}	θ_{330}	$\theta_{15\,000}$	θ_{100}	θ_{330}	$\theta_{15\,000}$
$RMSE_A$	0.026	0.033	0.029	0.024	0.037	0.028	0.012	0.016	0.013
ME_A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EF_A	0.44	0.21	0.49	0.52	0.00	0.55	0.88	0.83	0.90
$RMSE_{CV}$	0.028	0.035	0.032	0.034	0.038	0.037	0.027	0.036	0.031
ME_{CV}	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000
EF_{CV}	0.34	0.14	0.41	0.01	-0.03	0.21	0.36	0.05	0.45