



HAL
open science

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

► To cite this version:

Sixtine Cueff, Yves Coquet, Jean-Noël Aubertot, Liliane Bel, Valérie Pot, et al.. Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions. *Soil and Tillage Research*, 2021, 209, pp.104967. 10.1016/j.still.2021.104967 . hal-03206553

HAL Id: hal-03206553

<https://hal.inrae.fr/hal-03206553>

Submitted on 10 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

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

3 Sixtine Cueff^{a,b,*}, Yves Coquet^b, Jean-Noël Aubertot^a, Liliane Bel^c, Valérie Pot^b, Lionel Alletto^{a,*}

4 ^a Université de Toulouse, INRAE, UMR AGIR, F-31326, Castanet-Tolosan, France

5 ^b Université Paris-Saclay, INRAE, AgroParisTech, UMR ECOSYS, 78850, Thiverval-Grignon,
6 France

7 ^c Université Paris-Saclay, INRAE, AgroParisTech, UMR MIA-Paris, 75005, Paris, France

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

10

11 **Abstract**

12 Conservation agriculture has been developed as a means to improve the sustainability of
13 agricultural systems and reduce drawbacks of conventional agricultural practices. Cropping
14 practices can have a large influence on soil properties such as water retention. Proper tools are
15 needed to assess and model effects of conservation agriculture on soil properties. As measuring soil
16 water retention is expensive and time consuming, pedotransfer functions (PTFs) are now commonly
17 used to predict them. The objectives of this study were to (i) present a new dataset of conservation
18 agriculture data, (ii) assess performances of existing PTFs in predicting soil water retention of soils
19 under conservation agriculture and (iii) develop new specific PTFs to predict water retention in
20 conservation agriculture more accurately. We used data collected only in fields under conservation
21 agriculture in France to evaluate several published PTFs with three evaluation criteria (RMSE,
22 prediction bias (ME) and Nash-Sutcliffe Efficiency (EF)). We then developed new PTFs using three
23 methods – multiple linear regression, regression tree and random forest – to predict soil water
24 content at matric heads of -100 (θ_{100} , field capacity for sandy soils), -330 (θ_{330} , field capacity for
25 other soils) and -15 000 cm ($\theta_{15\ 000}$, wilting point). Soil tillage, presence of a cover crop, rotation
26 length and previous reduced/no tillage were used as predictors in addition to basic soil properties
27 for regression trees and random forests. The quality of prediction (RMSE, ME and EF) was
28 calculated for each new PTF using a cross-validation procedure. Generally, predictions of wilting

29 point had lower absolute error than those of sandy-soil field capacity (RMSE = 0.044 and 0.066
30 cm^3/cm^3 , respectively). EF was usually negative for all water contents. The cross-validation
31 performance of the new PTFs was similar for multiple linear regression (RMSE: 0.028, ME: 0.000,
32 EF: 0.34 for θ_{100}) and random forest (RMSE: 0.027, ME: 0.000, EF: 0.36 for θ_{100}), and generally
33 worse for regression tree (especially EF). Multiple linear regression that did not consider cropping
34 practices performed as well as random forest and thus did not identify any major influence of
35 agricultural management on predicted water content. Future research on developing PTFs should
36 focus on identifying more relevant predictors.

37

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

40

41 1. Introduction

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

52 Effects of soil cultivation practices on soil properties has received much research attention in
53 recent decades, but clear trends have not been established due to differences in location, soils and
54 agricultural practices (Green et al., 2003; Strudley et al., 2008). Tillage tends to decrease bulk

55 density and increase macroporosity, thus increasing the saturated and near-saturated hydraulic
56 conductivity of the tilled layer. These effects are, however, strongly time-dependent and usually
57 disappear rapidly after tillage (Mapa et al., 1986), due to natural soil reconsolidation caused by
58 wetting and drying cycles (Ahuja et al., 1998). Simultaneously, tillage interrupts macropore
59 connectivity between the soil surface and the untilled deeper soil, thus decreasing water movement
60 throughout the entire soil profile (Cameira et al., 2003). Conversely, untilled soils have higher bulk
61 density and greater pore connectivity (Gozubuyuk et al., 2014). Cover crops may (partially)
62 counterbalance negative effects of no tillage on bulk density by, for example, creating stable
63 biopores through their root development during the growing season (Williams and Weil, 2004;
64 Abdollahi and Munkholm, 2014). Moreover, after cover crop destruction, the dead residues form a
65 mulch that physically protects the soil surface from crusting (Baumhardt and Lascano, 1999).
66 Maintaining crop residues on the soil surface also leads to accumulation of soil organic matter in
67 topsoil layers (Kay and VandenBygaart, 2002) and improves aggregate stability (Devine et al.,
68 2014). In parallel, increased macrofauna activity (especially of earthworms) in conservation
69 agriculture systems forms biopores that improve water infiltration (Shipitalo et al., 2000).
70 Finally, soils under conservation agriculture also tend to have a larger proportion of finer pores
71 (micropores) (Hill et al., 1985). These changes in pore-size distribution could improve the storage
72 of plant-available water (Bescansa et al., 2006).

73 The variety and complexity of the counteracting effects of conservation agriculture on soil
74 properties call for developing new tools to properly assess and model these effects. Development of
75 water- and solute-transport models has received much research attention in recent decades. The lack
76 of accurate data on soil hydraulic properties, especially for soils under conservation agriculture,
77 however, hinders the use of models, as they require water-retention and hydraulic conductivity data
78 as inputs (Wösten et al., 1999). Despite significant improvements in measuring techniques,
79 researchers agree that directly measuring water-retention curves remain expensive, time consuming

80 and impossible at a large scale (Wösten et al., 2001; Vereecken et al., 2010; Román Dobarco et al.,
81 2019).

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

98 Although PTFs have significantly facilitated widespread application of water- and solute-
99 transport models at the field scale and larger scales (Vereecken et al., 2010), some of their limits
100 have been identified. Several authors suggested that using information in addition to the commonly
101 used sand, silt and clay contents, bulk density and organic matter could improve prediction accuracy
102 (Vereecken et al., 2010). Water contents at selected matric heads (Rawls et al., 1983; Al Majou et
103 al., 2008a) or terrain attributes (Obi et al., 2014) have been proposed as additional information.
104 Land cover (Nemes et al., 2003) or soil management (Tóth et al., 2015) have also been proposed,
105 but they may create PTFs that are less applicable than those that use only soil properties as

106 parameters. Whether the available PTFs apply equally to soils under conservation or conventional
107 agriculture has not yet been explored. The type of agriculture under which the soils used to develop
108 a particular PTF is rarely specified, but most PTFs seem to have been developed from soils under
109 conventional agriculture. To our knowledge, no one has attempted to develop specific tools to
110 predict water content in conservation agriculture systems. Chen et al. (1998) did observe that the
111 relevant properties for describing hydraulic conductivity differed between tilled and untilled soil,
112 which highlights the importance of soil management and supports the need for additional data and
113 specific tools to predict water dynamics in soils under conservation agriculture.

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

118

119 **2. Materials and methods**

120 *2.1 Description of the dataset on conservation agriculture*

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

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

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

$$155 \quad \theta_h = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha h)^n]^m} \quad m = 1 - \frac{1}{n} \quad (1)$$

156 where θ_r [cm³/cm³] and θ_s [cm³/cm³] are the residual and saturated volumetric water content (θ)
 157 respectively, h is matric head [cm], and α [cm⁻¹], n [-], and m [-] are shape parameters of the curve.
 158 The fit of the curves to the data had a mean R² (\pm 1 SD) of 0.98 \pm 0.02.

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

$$160 \quad AWC = (\theta_{FC} - \theta_{WP}) \times H \quad (2)$$

161 where θ_{FC} and θ_{WP} are volumetric water content at field capacity and permanent wilting point
 162 (cm³/cm³), respectively, and H is the depth of each of the three layers (here, 100 mm).

163 According to the literature, θ_{FC} can equal either θ_{100} (for sandy soils) or θ_{330} (for other soils), and
 164 θ_{WP} equals $\theta_{15\,000}$ (Hillel, 1971). Both AWC₁₀₀ and AWC₃₃₀ were considered for the two definitions
 165 of θ_{FC} . However, PTFs are usually used to predict volumetric water content at several matric heads
 166 rather than AWC. The rest of the study thus focused only on the relation between θ_{100} , θ_{330} , $\theta_{15\,000}$
 167 and basic soil properties and/or cropping practices.

168 *2.2 Analysis of the dataset*

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

176 2.3 *Published pedotransfer functions*

177 Twenty nine existing PTFs that predict θ_{100} or θ_{330} , and/or $\theta_{15\,000}$ (Table 1) and eight PTFs
178 that predict three parameters (n , α , and θ_s) of the van Genuchten (1980) water-retention curve (Eq.
179 1, Table 2) were taken from the literature and applied to data for the 140 soils in this study. The
180 study used class-PTFs (Cl), continuous-PTFs (Co) and tree-PTFs (Tr). PTFs were calibrated using
181 several published databases (Table 1). Of the 26 PTFs that predict θ_{100} , 13 were Cl and 13 were Co.
182 Of the 28 PTFs that predict θ_{330} , 13, 13 and 2 were Cl, Co and Tr, respectively. Of the 27 PTFs that
183 predict $\theta_{15\,000}$, 13, 12 and 2 were Cl, Co and Tr, respectively. These published PTFs use different
184 variables as predictors, such as texture/granulometric fractions, bulk density and organic carbon
185 content. Two PTFs (M2_Co and M3_Co) also use θ_{FC} and/or θ_{WP} as predictors. However, as a water
186 content cannot be used to predict itself, M2_Co and M3_Co were not used to predict θ_{330} or $\theta_{15\,000}$.
187 Most publications identified in the literature (Table 1) also had PTFs for subsoil horizons (> 30 cm).
188 We used only the published PTFs developed for the topsoil as the dataset contained only topsoil
189 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 .

190 2.4 *Development of new pedotransfer functions*

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

195 2.4.1 Multiple linear regression

196 We developed multiple linear regressions using stepwise regression with forward selection,
197 which could include all soil properties as predictors. In this procedure, the Akaike information
198 criterion (AIC) (Akaike et al., 1998) was used to determine which set of predictors predicted water
199 content best. AIC is calculated at each step of the stepwise regression to determine the improvement

200 brought by adding the new predictor. The “best” model is the one that helps decrease AIC the most.
201 The procedure stops when no more improvement can be made by a new predictor or when all
202 predictors are included.

203 2.4.2 Regression tree

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

222 2.4.3 Random forest

223 Like for regression tree, random forest is also based on recursive partitions of the data. The
224 difference is that a forest of multiple decorrelated trees is grown by using a randomly bootstrapped

225 subset of data and a random subset of predictors (Hastie et al., 2009; Ließ et al., 2012). The
 226 “randomForest” R package (Liaw and Wiener, 2002) was used to build forests. The forest consisted
 227 of 500 trees, and six of 18 variables were randomly selected to grow each tree. Like for regression
 228 tree, the minimum size of a terminal node was 5 data points. Unlike for regression tree, however, a
 229 single tree cannot be extracted from the forest, but the relative importance of the predictors can be
 230 determined and used to help interpret the results. The relative importance of predictors was
 231 estimated according to how much worse the prediction would be if the data for that predictor were
 232 permuted randomly (Prasad et al., 2006).

233 2.5 Evaluation of pedotransfer functions

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

$$238 \quad RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N [f(x_k) - y_k]^2} \quad (3)$$

$$239 \quad ME = \frac{1}{N} \sum_{k=1}^N [f(x_k) - y_k] \quad (4)$$

$$240 \quad EF = 1 - \frac{\sum_{k=1}^N [f(x_k) - y_k]^2}{\sum_{k=1}^N [y_k - \bar{y}]^2} \quad (5)$$

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

244 RMSE = 0 indicates perfect prediction of the observed data, while the ME indicates whether the
 245 PTF overpredicts (positive ME) or underpredicts (negative ME) the observed data. The closer ME
 246 is to 0, the lower the bias is. EF=1 indicates perfect prediction of the observed data, while EF<0

247 indicates prediction worse than the that using the mean of observed values (for which $EF=0$). These
248 criteria have no thresholds that can be used to conclude whether a prediction is good or not;
249 nevertheless, to help interpret the results, we arbitrarily defined ranges to indicate satisfactory
250 prediction of AWC: less than $0.020 \text{ cm}^3/\text{cm}^3$ for RMSE and ME, and greater than 0.50 for EF.

251 The three criteria were used to assess the performance of the published and new PTFs. For
252 published PTFs, $f(x_k)$ corresponded to predictions using basic soil properties in the conservation
253 agriculture dataset as input data, assessed with the criteria $RMSE_P$, ME_P and EF_P . For the new
254 PTFs, two groups of criteria were used to evaluate their performance. One group of three criteria
255 ($RMSE_A$, ME_A , EF_A) evaluated the quality of adjustment to the data. In this case, $f(x_k)$
256 corresponded to predictions by the new PTF using basic soil properties in the same dataset from
257 which they had been developed. The second group of criteria ($RMSE_{CV}$, ME_{CV} , EF_{CV}) evaluated the
258 cross-validation quality of prediction. As the dataset contained too few soils ($N=140$) to split out an
259 independent validation dataset, leave-one-out cross validation (Hastie et al., 2009) was performed
260 instead. In it, the dataset was split 140 times into two datasets of 139 and 1 soils, respectively. The
261 140 datasets of 139 soils were used to calibrate 140 new PTFs. The 140 predictions were then
262 compared to their corresponding value in the dataset of observed values.

263 3. Results

264 3.1 Preliminary analysis of the dataset

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

271 length) (except for tillage for AWC_{330}) and of depth for both AWCs. Both AWC_{100} and AWC_{330}
 272 were highest (by a small degree) in the 0-10 cm layer (Fig. 2a, e).

273 The plane defined by the first two axes of the PCA of basic soil properties explained 57% of the
 274 variance of the dataset (Fig. 3a). Of the 14 basic soil properties, only 8 contributed significantly (i.e.
 275 more than if each one had contributed equally (i.e., 7%)) to the first two axes. Strong correlations
 276 were found between CEC, CaO content and clay content ($r_{CEC/CaO} = 1$,
 277 $r_{Clay/CaO} = r_{CEC/CaO} = 0.9$), which contributed the most to the first two axes due to their large
 278 contributions to the first axis (17%, 16% and 16%, respectively). Nitrogen, organic carbon and
 279 phosphorus contents contributed the most to the second axis (22%, 21% and 19%, respectively).
 280 Strong to very strong correlations were found between organic carbon, nitrogen and K_2O contents
 281 ($r_{OC/N} = 0.94$, $r_{K_2O/OC} = r_{K_2O/N} = 0.7$). Thus, soil layers above the second axis of the PCA had
 282 higher organic carbon, nitrogen and phosphorus contents, which was related to their depth, as most
 283 soil layers above the second axis were 0-10 cm deep (Fig. 3b). This is consistent with the low
 284 mechanical disturbance of the soil surface under conservation agriculture, which results in a thin
 285 horizon 5-10 cm deep that can exhibit different soil properties, especially organic matter. When
 286 projected as supplementary variables on the plane, water contents were poorly represented (Fig. 3a),
 287 which suggested that none of the basic soil properties were strongly related to them, as confirmed
 288 by correlation coefficients. The strongest significant correlations for θ_{100} were with clay content
 289 ($r = 0.5$), bulk density ($r = -0.4$), sand content ($r = -0.4$) and CEC ($r = 0.4$). Correlations for
 290 θ_{330} were weaker, not exceeding 0.3 with clay content or -0.3 with bulk density. Correlations for
 291 $\theta_{15\,000}$ were the strongest among those for the three water contents: 0.6 with clay content, CEC and
 292 CaO content.

293 We plotted θ_{100} , θ_{330} and $\theta_{15\,000}$ vs. cropping practices, rotation length, soil tillage and cover-
 294 crop presence to identify the influence of conservation agriculture on water contents. We also
 295 investigated the influence of depth, as the PCA indicated a difference between the 0-10 cm layer

296 and the other two layers. There were no clear differences between θ_{100} or θ_{330} as a function of
 297 agricultural practices, except for rotation length, with water content lower with variable rotations
 298 and higher with short rotations, compared to long or medium rotations (Fig. 4a, b). ANOVA
 299 confirmed a significant effect of rotation length on θ_{100} ($P < 0.001$) and θ_{330} ($P < 0.01$). For $\theta_{15\,000}$,
 300 water content was lower under strip-till than under the other types of tillage and had a trend similar
 301 to those of θ_{100} and θ_{330} for rotation length (Fig. 4c, d). All three cropping practices had a significant
 302 effect on $\theta_{15\,000}$ ($P < 0.01$ for soil tillage and $P < 0.001$ for cover-crop presence and rotation length).
 303 Unlike for AWC, depth had no significant effect on any of the water contents.

304 *3.2 Evaluation of the performance of published pedotransfer functions*

305 3.2.1 Prediction of volumetric water content at selected matric heads

306 For prediction of θ_{100} , $RMSEP$ varied from $0.034 \text{ cm}^3/\text{cm}^3$ (M3_Co) to $0.262 \text{ cm}^3/\text{cm}^3$
 307 (M2_Co) (Table 4). These extreme values were exceptions, however; mean ($\pm 1 \text{ SD}$) $RMSEP$ for
 308 most of the PTFs (22 of 26) was $0.055 \pm 0.009 \text{ cm}^3/\text{cm}^3$. Of the 26 PTFs, 24 underpredicted θ_{100} ,
 309 with MEP ranging from -0.112 to $-0.007 \text{ cm}^3/\text{cm}^3$. The same four PTFs that had extreme values of
 310 $RMSEP$ (M1_Co, M2_Co, M3_Co and M10_Co) had extremely high or low MEP . For EF_P ,
 311 negative or near-zero values showed that none of the PTFs tested predicted θ_{100} well. According to
 312 the three criteria, M3_Co, despite having been developed from samples from many locations in the
 313 USA, predicted θ_{100} the best, but used both θ_{330} and $\theta_{15\,000}$ as predictors. However, the other two
 314 PTFs developed from the same data (M1_Co, M2_Co) predicted θ_{100} the worst. Among the
 315 remaining PTFs, which used only basic soil properties, eight French Cl PTFs (M7_Cl, M8_Cl,
 316 M12_Cl, M13_Cl, M14_Cl, M19_Cl, M20_Cl, M21_Cl) had better $RMSEP$ (0.046 ± 0.004
 317 cm^3/cm^3) and MEP ($-0.028 \pm 0.004 \text{ cm}^3/\text{cm}^3$) than the others. However, ME remained
 318 unsatisfactory. All eight PTFs were Cl that used FAO texture or FAO texture and bulk density as
 319 classes.

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

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

335 3.2.2 Prediction of water-retention curve parameters

336 For predicting θ_s , $RMSE_P$ ranged from $0.035\text{-}0.439 \text{ cm}^3/\text{cm}^3$, while ME_P ranged from -0.438 to
 337 $0.010 \text{ cm}^3/\text{cm}^3$ (Table 5). P2_Co and P4_Co had large errors due to physically impossible values of
 338 θ_s (close to 0 or even negative). For the other PTFs that predicted θ_s , $RMSE_P$ and ME_P had
 339 satisfactory performances, with the best performance by P3_Cl, P7_Co, P8_Co and P9_Co
 340 ($RMSE_P= 0.037 \pm 0.001 \text{ cm}^3/\text{cm}^3$; $ME_P= 0.023 \pm 0.010 \text{ cm}^3/\text{cm}^3$; $EF_P= -0.43 \pm 0.12$). Negative EF_P
 341 values, however, indicated that none of the PTFs performed better than the mean of observed
 342 values.

343 For predicting α , the French PTF P3_Cl had particularly poor performance according to all
 344 criteria, and P4_Cl predicted physically impossible values. Thus, the best predictions were obtained

345 only with PTFs developed at the European scale, all of which performed similarly. For predicting n ,
346 $RMSE_P$ varied from 0.305-0.366. The nine PTFs always underpredicted n (negative ME_P), except
347 for P8_Co, which had the only satisfactory ME_P (0.003) and the best $RMSE_P$. The two PTFs
348 developed from soil samples from France performed slightly worse according to all criteria.

349 *3.3 Development of new pedotransfer functions*

350 3.3.1 Multiple linear regression

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

360 3.3.2 Regression tree

361 The maximum tree grown for θ_{100} , θ_{330} and $\theta_{15\,000}$ had 11, 9 and 8 partitions, respectively,
362 despite the inclusion of 18 potential predictors. After pruning, the θ_{330} tree was reduced to the
363 minimum size (no partitions); thus, the mean of θ_{330} was the best compromise between a suitable
364 tree size and low error in predicting the testing data. Consequently, only the trees that predicted θ_{100}
365 and $\theta_{15\,000}$ were evaluated (Fig. 6). The pruned θ_{100} and $\theta_{15\,000}$ trees were split 7 and 4 times,
366 respectively, and had three predictors in common: rotation length, clay content and bulk density.
367 Both trees were first split according to rotation length, which split variable length from the other
368 lengths. No other cropping practices appeared in the pruned trees. According to the criteria, all trees

369 had satisfactory quality of adjustment to observed data, with $ME_A=0$ and $EF_A \geq 0.5$ (Table 7). All
 370 criteria except ME_A were slightly higher for $\theta_{15\ 000}$ than for θ_{100} . The criteria for cross-validation
 371 quality of prediction had similar trends, with low ME_{CV} , but the trees did not predict well according
 372 to EF_{CV} (<0.21) (Table 7). Prediction performance thus decreased between adjustment and cross
 373 validation (Fig. 5b).

374 3.3.3 Random forest

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

386 **4. Discussion**

387 *4.1 Evaluation of the performance of published pedotransfer functions*

388 Most PTFs (24 of 26) underpredicted soil volumetric water content at -100 cm of matric
 389 head, while no clear trend (overestimation or underestimation) was observed at -330 and -15 000
 390 cm. The $RMSE_P$ for predicting volumetric water content were largest for the -100 and -330 cm
 391 matric heads. Although EF_P was higher for several PTFs at -15 000 cm, it was never satisfactory (\geq
 392 0.5). Overall, none of the 29 published PTFs provided satisfactory prediction of the volumetric

393 water content at the selected matric heads (-100, -330 and -15 000 cm) according to any of the
394 criteria, which limits their use in soil transport models under conservation agriculture.

395 The published PTFs may have had low-quality predictions for several reasons. First, differences
396 in the sampling or measurement protocol between the databases used to develop the PTFs and the
397 dataset that we used may be a source of uncertainty (Román Dobarco et al., 2019). For example, Al
398 Majou et al. (2008b) measured water content using undisturbed aggregates (10-15 cm³), whereas we
399 used undisturbed soil cylinders (50 cm³). Several studies have also highlighted the influence of
400 sample size on soil water retention and the quality of PTFs developed (Ghanbarian et al., 2015;
401 Silva et al., 2018). Furthermore, some of these PTFs were developed from large databases collated
402 in the USA or Europe and covered a wide range of sand, silt, clay and organic matter contents and
403 bulk densities (Rawls et al., 1982; Tóth et al., 2015). Like Cornelis et al. (2001), we calculated the
404 ranges of the soil properties of our samples and found that all lay within those in the databases from
405 the USA and Europe; nevertheless, the predictions were unsatisfactory according to the criteria.
406 Nemes et al. (2003) suggested that using a small set of relevant data rather than a larger, more
407 general dataset can produce more accurate PTFs. Indeed, for predicting Hungarian soils, they found
408 that PTFs that had been developed by neural networks from data from throughout the USA and
409 Europe performed worse than PTFs that had been developed from a smaller dataset that considered
410 the pedoclimatic context (e.g. the subset of Hungarian soils). Testing published PTFs developed
411 from large and general datasets with our dataset may explain the poor prediction in our study.
412 However, most of the PTFs tested were developed from French databases (Bruand et al., 2004; Al
413 Majou et al., 2007, 2008a, 2008b; Román Dobarco et al., 2019) and should have been more
414 appropriate for predicting water content of the soils in our dataset. These French PTFs, however,
415 did not necessarily perform better than those developed by Tóth et al. (2015) at the European scale.
416 They did, however, perform better than those of Rawls et al. (1982), which were developed from
417 soil samples from the USA, which appeared to be unsuitable (criteria among the worst for each PTF
418 evaluated), except when using other water contents as predictors. The poor performance of the

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

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

431 To predict water content better, some authors suggested including other water contents at given
432 matric heads in the PTFs (Al Majou et al., 2008a; Rawls et al., 1982; Vereecken et al., 2010). In our
433 study, predictions of such PTFs were slightly better than those of PTFs that included only soil
434 properties, but with differences depending on the specific water content included in the PTF. As
435 observed by Al Majou et al. (2008a), water content prediction improved when the other water
436 content included was that at field capacity (in this case, θ_{330}), but not that at the wilting point
437 ($\theta_{15\ 000}$), as observed by Borgesen and Schaap (2005). The improvement in prediction when using
438 the field capacity water content was related to the shape of soil water-retention curves, which
439 inflected strongly near field capacity. However, determining water content at field capacity in order
440 to include it in PTFs remains unsatisfactory, as doing so, mainly in laboratories, is time-consuming
441 and costly. Other authors suggest that information on soil structure, which is often considered
442 through bulk density, should be included to improve PTF performance. In the study of Al Majou et
443 al. (2008b), including bulk density kept bias low and improved prediction of water content. In our

444 study, predictions of θ_{330} had errors similar to or larger than those of Al Majou et al. (2008b), but
445 unlike their results, including bulk density did not improve predictions. Soil bulk density in
446 conservation tillage systems is generally higher than that in conventional systems, which results in
447 lower total porosity than that in tilled soils but, conversely, generally higher saturated and near-
448 saturated hydraulic conductivity (Green et al., 2003). While, bulk density is a good proxy of
449 hydraulic dynamics (Blanco-Canqui et al., 2004; Alletto et al., 2010) and AWC in conventionally
450 tilled soils, it is less effective in conservation agriculture (Alletto et al., 2010; Chen et al., 1998),
451 probably due to greater pore connectivity and proportion of macro- and mesopores in the latter. This
452 disconnection between hydraulic properties and bulk density in conservation agriculture can indeed
453 be attributed to major changes in pore-size distribution and connectivity when tillage intensity is
454 reduced (Strudley et al., 2008; Alletto et al., 2010), thus leading to changes in AWC. Furthermore,
455 as mentioned by several authors (e.g., Nakano and Miyazaki, 2005; Lilly and Nemes, 2008), the
456 cylindrical core method used to measure bulk density does not predict pore connectivity well, so
457 complementary methods must be used to assess it.

458 *4.2 Development of new pedotransfer functions*

459 Multiple linear regression is commonly used to develop PTFs (Wösten et al., 2001; Al
460 Majou et al., 2008a; Tóth et al., 2015; Román Dobarco et al., 2019), unlike regression trees or
461 random forests. Regression trees have been used to predict water content, but without considering
462 cropping practices: Tóth et al. (2015) predicted θ_{330} and $\theta_{15\,000}$ using textural and taxonomic
463 information (Table 1), while Rawls and Pachepsky (2002) did the same using textural and structural
464 classes. To our knowledge, our study is the first to use random forests to predict water content. Vos
465 et al. (2019) used random forests to highlight the influence of land use or land-use history classes,
466 clay content and electrical conductivity on predicting topsoil carbon stock. In our study, random
467 forests highlighted that some predictors not usually used in PTFs, such as CEC and rotation length,
468 could help predict water content at a given matric head. Some properties have been suggested as

469 important for predicting water content due to an indirect influence, such as organic carbon, which
470 plays both an indirect role, by improving soil structure, and a direct role, through its adsorption
471 properties (Tóth et al., 2015). Cropping practices influence soil properties greatly, especially soil
472 structure (Strudley et al., 2008), and can thus influence water content indirectly. Román Dobarco et
473 al. (2019) suggest that land use should be considered in future PTFs, even though PTFs are
474 generally suitable for most agricultural soils.

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

483 In terms of quality of adjustment, random forests performed the best, with almost perfect
484 fits. This was likely due to the nature of machine-learning methods, which “learn” from the dataset
485 provided and thus perform well with it. Consequently, we also expected regression trees to have
486 high quality of adjustment, but their results were similar to those of multiple linear regressions. This
487 result was likely related to the pruning, as adjustment to the training data is purposely reduced so
488 that the model performs better with a test dataset. In our study, however, performance of regression
489 trees and random forests decreased between adjustment (i.e. the training dataset) and cross-
490 validation (the test dataset) (Fig. 5). While the poor prediction by the regression trees can be
491 explained easily by their well-known instability (i.e. a small difference in the training dataset can
492 result in a different tree) (Gey and Poggi, 2006; Yang et al., 2016), the instability of the random
493 forests was more surprising. Conversely, multiple linear regression was a stable method whose

494 quality of prediction was as good or better than that of the machine-learning methods. The
495 similarity between its adjustment and cross-validation performances demonstrates its robustness.
496 Overall, however, the cross-validation quality of prediction remained unsatisfactory in this study,
497 mainly for EF_{CV} , which never reached satisfactory values for any of the PTFs despite having
498 satisfactory ME_{CV} (close to 0).

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

514 **Conclusions**

515 We tested the performance of several published PTFs and newly developed PTFs using multiple
516 linear regressions, regression trees and random forests to predict water content at field capacity ($h =$
517 -100 or -300 cm) and wilting point ($h = -15\ 000$ cm). Although some PTFs approached satisfactory
518 performance according to the three criteria, none of them managed to reach it, which limits their use

519 in soil transport models for conservation agriculture. Most of our soil samples belonged to the
520 domain of applicability of the PTFs, so the poor results obtained are likely related to (i) the use of
521 unsuitable predictors, (ii) the use of PTFs developed at an inappropriate scale or (iii) differences in
522 soil management between databases.

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

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

533 **Acknowledgements**

534 This study was performed with data obtained in the framework of the CASDAR TTSI project no.
535 8102 (coordinated by the Chambre Régionale d'Agriculture de Midi-Pyrénées). Data analysis was
536 performed in the framework of the BAG'AGES and BAG'AGES CISOL projects and financed by
537 the Agence de l'Eau Adour-Garonne and Occitanie Region. We thank Isabelle Cousin (INRAE UR
538 Sol, France) for sharing the SOLHYDRO database, which allowed us to acquire complementary
539 results.

540 **References**

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

- 543 <https://doi.org/10.2136/sssaj>
- 544 Ahuja, L.R., Fiedler, F., Dunn, G.H., Benjamin, J.G., Garrison, A., 1998. Changes in Soil Water
545 Retention Curves Due to Tillage and Natural Reconsolidation. *Soil Sci. Soc. Am. J.* 62, 1228–
546 1233.
- 547 Akaike, H., Parzen, E., Tanabe, K., Kitagawa, G., 1998. *Selected Papers of Hirotugu Akaike*.
548 Springer Science & Business Media.
- 549 Al Majou, H., Bruand, A., Duval, O., 2008a. The use of in situ volumetric water content at field
550 capacity to improve the prediction of soil water retention properties. *Can. J. Soil Sci.* 88, 533–
551 541. <https://doi.org/10.4141/CJSS07065>
- 552 Al Majou, H., Bruand, A., Duval, O., Cousin, I., 2007. Comparaison de fonctions de pédotransfert
553 nationales et européennes pour prédire les propriétés de rétention en eau des sols. *Etude Gest.*
554 *des Sols* 14, 103–116.
- 555 Al Majou, H., Bruand, A., Duval, O., Le Bas, C., Vautier, A., 2008b. Prediction of soil water
556 retention properties after stratification by combining texture, bulk density and the type of
557 horizon. *Soil Use Manag.* 24, 383–391. <https://doi.org/10.1111/j.1475-2743.2008.00180.x>
- 558 Alletto, L., Coquet, Y., Benoit, P., Heddadj, D., Barriuso, E., 2010. Tillage management effects on
559 pesticide fate in soils. A review. *Agron. Sustain. Dev.* 30, 367–400.
- 560 Baumhardt, R.L., Lascano, R.J., 1999. Water Budget and Yield of Dryland Cotton Intercropped
561 with Terminated Winter Wheat. *Agron. J.* 91, 922–927.
- 562 Bescansa, P., Imaz, M.J., Virto, I., Enrique, A., Hoogmoed, W.B., 2006. Soil water retention as
563 affected by tillage and residue management in semiarid Spain. *Soil Tillage Res.* 87, 19–27.
564 <https://doi.org/10.1016/j.still.2005.02.028>
- 565 Blanco-Canqui, H., Gantzer, C.J., Anderson, S.H., Alberts, E.E., 2004. Tillage and Crop Influences

- 566 on Physical Properties for an Epiqualf. *Soil Sci. Soc. Am. J.* 68, 567–576.
567 <https://doi.org/10.2136/sssaj2004.5670>
- 568 Borgesen, C.D., Schaap, M.G., 2005. Point and parameter pedotransfer functions for water retention
569 predictions for Danish soils. *Geoderma* 127, 154–167.
570 <https://doi.org/10.1016/j.geoderma.2004.11.025>
- 571 Bruand, A., Duval, O., Cousin, I., 2004. Estimation des propriétés de rétention en eau des sols à
572 partir de la base de données SOLHYDRO: Une première proposition combinant le type
573 d'horizon, sa texture et sa densité apparente. *Étude Gest. des Sols* 11, 323–334.
- 574 Bruand, A., Pérez Fernández, P., Duval, O., 2003. Use of class pedotransfer functions based on
575 texture and bulk density of clods to generate water retention curves. *Soil Use Manag.* 19, 232–
576 242. <https://doi.org/10.1079/SUM2003196>
- 577 Cameira, M.R., Fernando, R.M., Pereira, L.S., 2003. Soil macropore dynamics affected by tillage
578 and irrigation for a silty loam alluvial soil in southern Portugal. *Soil Tillage Res.* 70, 131–140.
579 [https://doi.org/10.1016/S0167-1987\(02\)00154-X](https://doi.org/10.1016/S0167-1987(02)00154-X)
- 580 Chen, Y., Tessier, S., Gallichand, J., 1998. Estimates of tillage effects on saturated hydraulic
581 conductivity. *Can. Agric. Eng.* 40, 169–177.
- 582 Cornelis, W.M., Ronsyn, J., Meirvenne, M. Van, Hartmann, R., 2001. Evaluation of pedotransfer
583 functions for predicting the soil moisture retention curve. *Soil Sci. Soc. Am. J.* 65, 638–648.
- 584 Devine, S., Markewitz, D., Hendrix, P., Coleman, D., 2014. Soil aggregates and associated organic
585 matter under conventional tillage, no-tillage, and forest succession after three decades. *PLoS*
586 *One* 9, 1–12. <https://doi.org/10.1371/journal.pone.0084988>
- 587 FAO, 2016. Conservation Agriculture, in: *Save and Grow in Practice: Maize, Rice and Wheat.*
- 588 Gey, S., Poggi, J., 2006. Boosting and instability for regression trees. *Comput. Stat. Data Anal.* 50,

- 589 533–550. <https://doi.org/10.1016/j.csda.2004.09.001>
- 590 Ghanbarian, B., Taslimitehrani, V., Dong, G.Z., Pachepsky, Y.A., 2015. Sample dimensions effect
591 on prediction of soil water retention curve and saturated hydraulic conductivity. *Journal of*
592 *Hydrology* 528, 127-137.
- 593 Ghorbani Dashtaki, S., Homaei, M., Khodaverdiloo, H., 2010. Derivation and validation of
594 pedotransfer functions for estimating soil water retention curve using a variety of soil data.
595 *Soil Use Manag.* 26, 68–74. <https://doi.org/10.1111/j.1475-2743.2009.00254.x>
- 596 Gozubuyuk, Z., Sahin, U., Ozturk, I., Celik, A., Adiguzel, M.C., 2014. Tillage effects on certain
597 physical and hydraulic properties of a loamy soil under a crop rotation in a semi-arid region
598 with a cool climate. *Catena* 118, 195–205. <https://doi.org/10.1016/j.catena.2014.01.006>
- 599 Green, T.R., Ahuja, L.R., Benjamin, J.G., 2003. Advances and challenges in predicting agricultural
600 management effects on soil hydraulic properties. *Geoderma* 116, 3–27.
601 [https://doi.org/10.1016/S0016-7061\(03\)00091-0](https://doi.org/10.1016/S0016-7061(03)00091-0)
- 602 Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining,*
603 *Inference, and Prediction*, Second. ed. Springer Series in Statistics.
- 604 Hill, R.L., Horton, R., Cruse, R.M., 1985. Tillage effects on soil water retention and pore size
605 distribution of two mollisols. *Soil Sci. Soc. Am. J.* 49, 1264–1270.
- 606 Hillel, D., 1971. *Soil and Water: Physical Principles and Processes*, Academic Press.
- 607 Hobbs, P.R., Sayre, K., Gupta, R., 2008. The role of conservation agriculture in sustainable
608 agriculture. *Philos. Trans. R. Soc. B Biol. Sci.* 363, 543–555.
609 <https://doi.org/10.1098/rstb.2007.2169>
- 610 IUSS Working Group WRB, 2015. *World Reference Base for Soil Resources 2014: International*
611 *soil classification system for naming soils and creating legends for soil maps*, World Soil. ed.

- 612 Kay, B.D., VandenBygaart, A.J., 2002. Conservation tillage and depth stratification of porosity and
613 soil organic matter. *Soil Tillage Res.* 66, 107–118.
- 614 Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R package. *R News* 2,
615 18–22. <https://doi.org/10.1023/A>
- 616 Ließ, M., Glaser, B., Huwe, B., 2012. Geoderma Uncertainty in the spatial prediction of soil texture
617 Comparison of regression tree and Random Forest models. *Geoderma* 170, 70–79.
618 <https://doi.org/10.1016/j.geoderma.2011.10.010>
- 619 Lilly, A., Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2008. Probabilistic Approach to the Identifi
620 cation of Input Variables to Estimate Hydraulic Conductivity. *Soil Sci. Soc. Am. J.* 72, 16–24.
621 <https://doi.org/10.2136/sssaj2006.0391>
- 622 Mapa, R.B., Green, R.E., Santo, L., 1986. Temporal Variability of Soil Hydraulic Properties with
623 Wetting and Drying Subsequent to Tillage. *Soil Sci. Soc. Am. J.* 50, 1133–1138.
- 624 Nakano, K., Miyazaki, T., 2005. Predicting the saturated hydraulic conductivity of compacted
625 subsoils using the non-similar media concept. *Soil Tillage Res.* 84, 145–153.
626 <https://doi.org/10.1016/j.still.2004.11.010>
- 627 Nash, J.E., Sutcliffe, J. V, 1970. River Flow Forecasting Through Conceptual Models Part I-A
628 Discussion of Principles. *J. Hydrol.* 10, 282–290. <https://doi.org/10.1016/0022->
629 [1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- 630 Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions
631 derived from different scales of data collection. *Soil Sci. Soc. Am. J.* 67, 1093–1102.
632 <https://doi.org/10.2136/sssaj2003.1093>
- 633 Obi, J.C., Ogban, P.I., Ituen, U.J., Udoh, B.T., 2014. Development of pedotransfer functions for
634 coastal plain soils using terrain attributes. *Catena* 123, 252–262.

- 635 <https://doi.org/10.1016/j.catena.2014.08.015>
- 636 Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques:
637 bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
638 <https://doi.org/10.1007/s10021-005-0054-1>
- 639 R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for
640 Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>
- 641 Rawls, W.J., Brakensiek, D.L., Miller, N., 1983. Green-ampt Infiltration Parameters from Soils
642 Data. *J. Hydraul. Eng.* 109, 62–70. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1983\)109:1\(62\)](https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(62))
- 643 Rawls, W.J., Brakensiek, D.L., Saxton, K.E., 1982. Estimation of Soil Water Properties. *Trans.*
644 *ASAE*. <https://doi.org/10.13031/2013.33720>
- 645 Rawls, W.J., Pachepsky, Y.A., 2002. Soil Consistence and Structure as Predictors of Water
646 Retention. *Soil Sci. Soc. Am. J.* 66, 1115–1126.
- 647 Román Dobarco, M., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for
648 predicting available water capacity in French soils, their applicability domain and associated
649 uncertainty. *Geoderma* 336, 81–95. <https://doi.org/10.1016/J.GEODERMA.2018.08.022>
- 650 Rubio, C.M., Llorens, P., Gallart, F., 2008. Uncertainty and efficiency of pedotransfer functions for
651 estimating water retention characteristics of soils. *Eur. J. Soil Sci.* 59, 339–347.
652 <https://doi.org/10.1111/j.1365-2389.2007.01002.x>
- 653 Shipitalo, M.J., Dick, W.A., Edwards, W.M., 2000. Conservation tillage and macropore factors that
654 affect water movement and the fate of chemicals. *Soil Tillage Res.* 53, 167–183.
- 655 Silva, M.L.D., Libardi, P.L., Gimenes, F.H.S., 2018. Soil Water Retention Curve as Affected by
656 Sample Height. *Revista Brasileira De Ciencia Do Solo* 42.
- 657 Strudley, M.W., Green, T.R., Ascough II, J.C., 2008. Tillage effects on soil hydraulic properties in

- 658 space and time : State of the science. *Soil Tillage Res.* 99, 4–48.
659 <https://doi.org/10.1016/j.still.2008.01.007>
- 660 Therneau, T., Atkinson, B., 2019. Recursive Partitioning and Regression Trees. R package.
- 661 Tomasella, J., Crestana, S., Rawls, W.J., 2003. Comparison of Two Techniques to Develop
662 Pedotransfer Functions for Water Retention. *Soil Sci. Soc. Am. J.* 67, 1085–1092.
- 663 Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., Tóth, G., 2015. New generation of
664 hydraulic pedotransfer functions for Europe. *Eur. J. Soil Sci.* 66, 226–238.
665 <https://doi.org/10.1111/ejss.12192>
- 666 Tranter, G., Mcbratney, A.B., Minasny, B., 2009. Using distance metrics to determine the
667 appropriate domain of pedotransfer function predictions. *Geoderma* 149, 421–425.
668 <https://doi.org/10.1016/j.geoderma.2009.01.006>
- 669 van Genuchten, M.T., 1980. A Closed-form Equation For Predicting the Hydraulic Conductivity of
670 Unsaturated Soils. *Soil Sci. Soc. Am. J.* 44, 892–898.
- 671 van Genuchten, M.T., Leij, F.J., Yates, S.R., 1991. The RETC Code for Quantifying the Hydraulic
672 Functions of Unsaturated Soils. EPA/600/2-91/065.
- 673 Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., van Genuchten,
674 M.T., 2010. Using Pedotransfer Functions to Estimate the van Genuchten–Mualem Soil
675 Hydraulic Properties: A Review. *Vadose Zo. J.* 9, 795–820.
676 <https://doi.org/10.2136/vzj2010.0045>
- 677 Verhulst, N., François, I., Govaerts, B., 2010. Conservation agriculture, improving soil quality for
678 sustainable production systems? *Adv. soil Sci. food Secur. soil Qual.* 1799267585, 137–208.
- 679 Vos, C., Don, A., Hobbey, E.U., Prietz, R., Heidkamp, A., Freibauer, A., 2019. Factors controlling
680 the variation in organic carbon stocks in agricultural soils of Germany. *Eur. J. Soil Sci.* 70,

- 681 550–564. <https://doi.org/10.1111/ejss.12787>
- 682 Williams, S.M., Weil, R.R., 2004. Crop Cover Root Channels May Alleviate Soil Compaction
683 Effects On Soybean Crop. *Soil Sci. Soc. Am. J.* 68, 1403–1409.
- 684 Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of
685 hydraulic properties of European soils. *Geoderma* 90, 169–185. [https://doi.org/10.1016/S0016-](https://doi.org/10.1016/S0016-7061(98)00132-3)
686 7061(98)00132-3
- 687 Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: Bridging the gap
688 between available basic soil data and missing soil hydraulic characteristics. *J. Hydrol.* 251,
689 123–150. [https://doi.org/10.1016/S0022-1694\(01\)00464-4](https://doi.org/10.1016/S0022-1694(01)00464-4)
- 690 Yang, R., Zhang, G., Liu, F., Lu, Y., Yang, Fan, Yang, Fei, Yang, M., Zhao, Y.-G., Li, D.-C., 2016.
691 Comparison of boosted regression tree and random forest models for mapping topsoil organic
692 carbon concentration in an alpine ecosystem. *Ecol. Indic.* 60, 870–878.

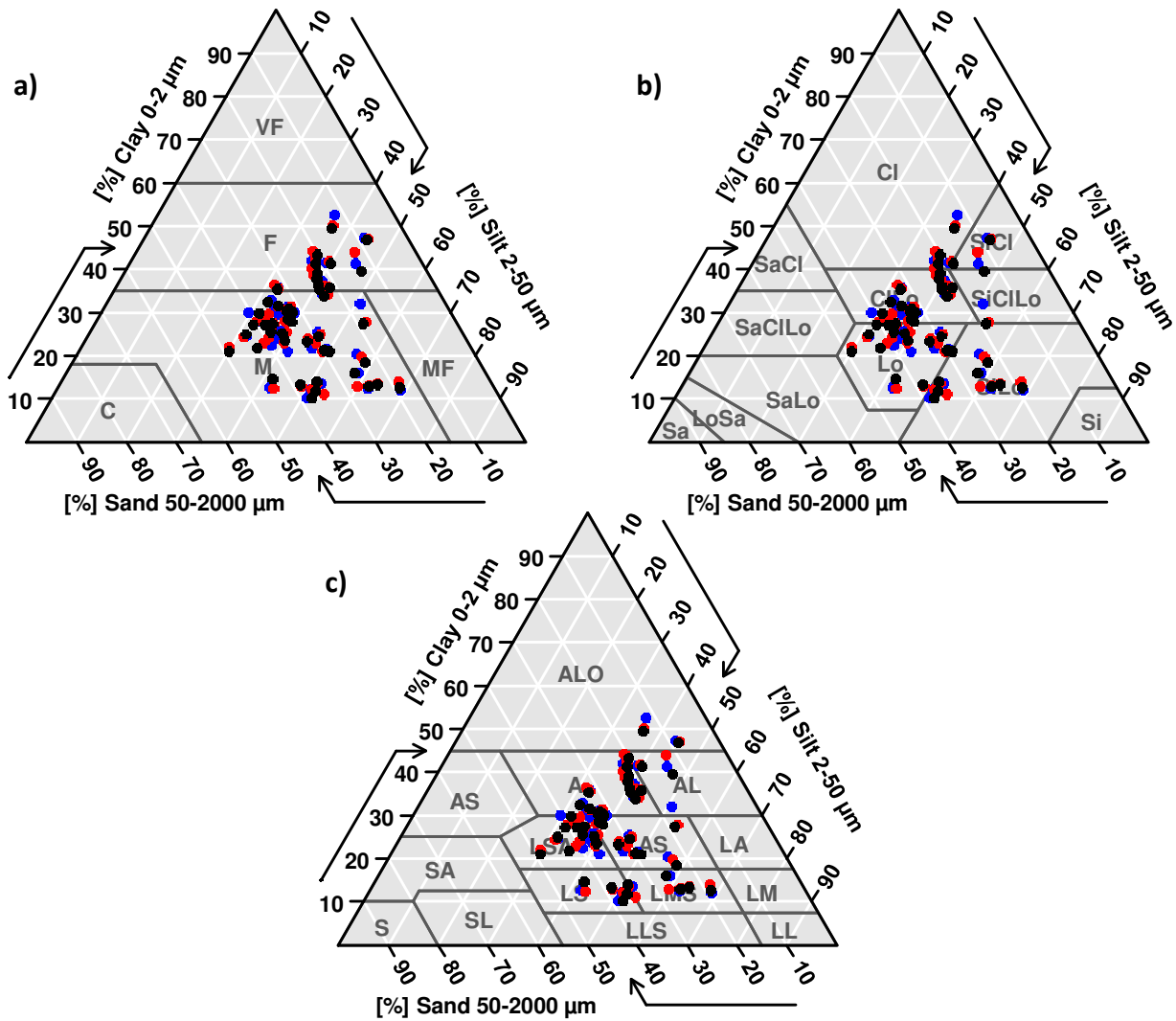


Fig. 1. Textures of the soil samples collected in 0-10, 10-20 and 20-30 cm deep soil layers (black, red and blue dots, respectively) in 61 conservation agriculture fields. The three texture triangles are based on the three classifications used to develop the pedotransfer functions found in the literature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

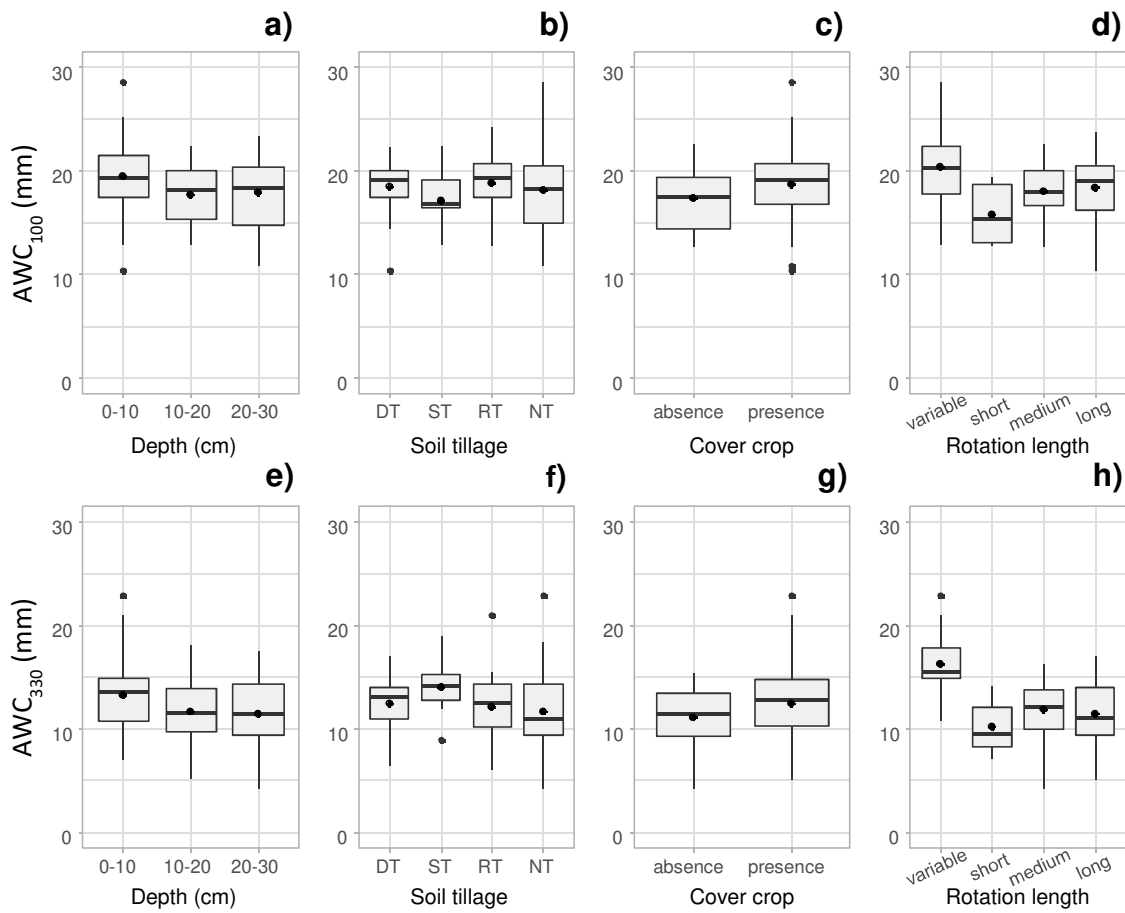


Fig. 2. Available water capacity (AWC) predicted assuming field capacity at a volumetric water content of -100 cm (AWC_{100}) or -330 cm (AWC_{330}) of matric head as a function of soil depth (a, e), soil tillage (b, f), cover-crop presence (c, g) and rotation length (d, h). For soil tillage, DT: deep 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.

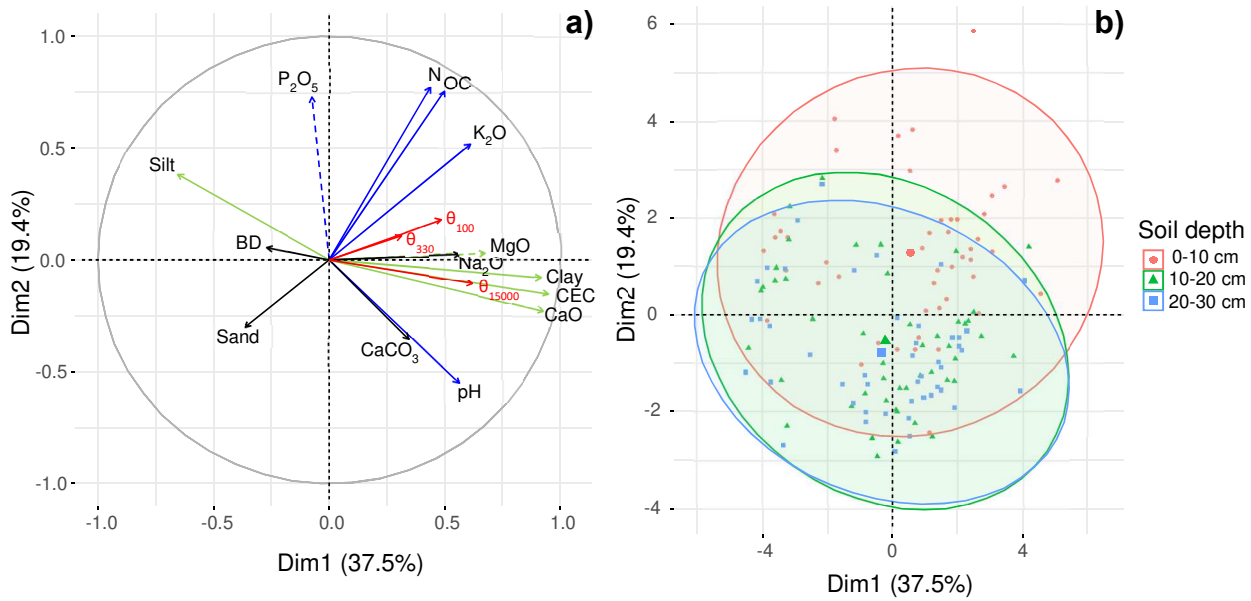


Fig. 3. Correlation circle of the (a) variables and (b) soil layers on the first two dimensions of the principal component analysis. (a) The variables that contributed significantly to the first and second axis are green and blue, respectively. Dashed arrows correspond to variables that did not contribute 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 (θ_{15000}) 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.

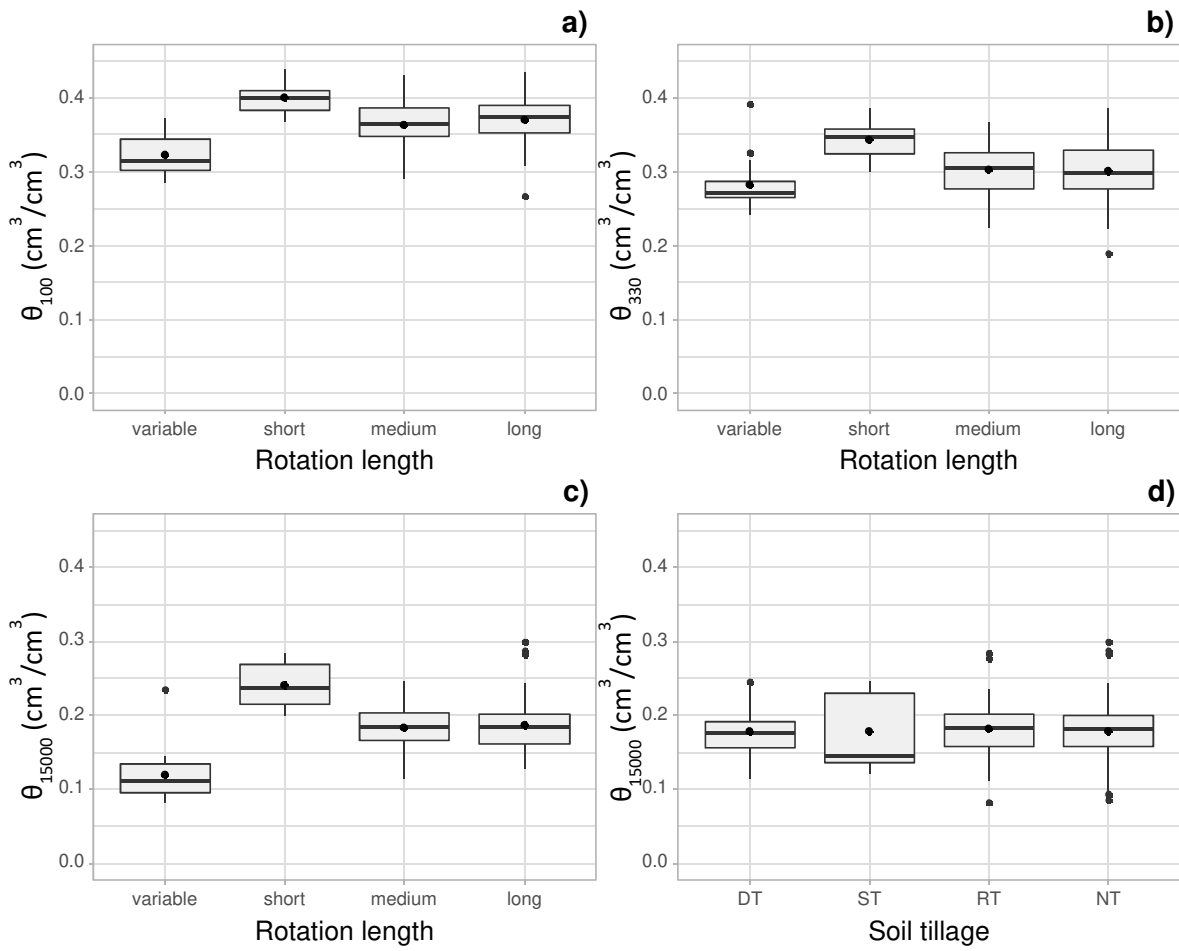


Fig. 4. Volumetric water content at (a) -100 cm (θ_{100}), (b) -330 cm (θ_{330}) and (c) -15 000 cm (θ_{15000}) of matric head as a function of rotation length (a, b, c) and (d) at -15 000 cm (θ_{15000}) as a function of soil tillage. For soil tillage, DT: deep tillage, RT: reduced tillage, ST: strip-till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

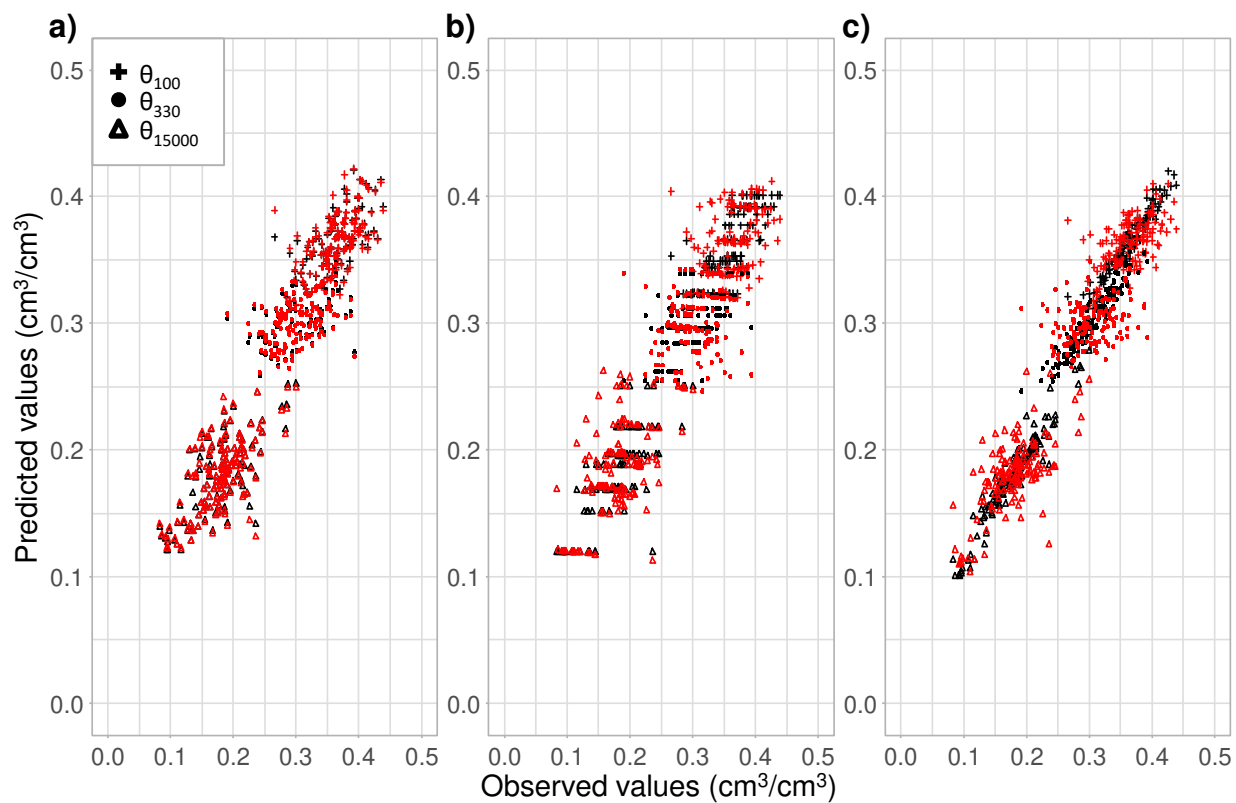


Fig. 5. Observed vs. predicted soil water content at -100 cm (θ_{100}), -330 cm (θ_{330}) and -15 000 cm (θ_{15000}) of matric head for (a) multiple linear regression, (b) regression tree and (c) random forest. Predicted adjustment values and cross-validation values are black and red, respectively.

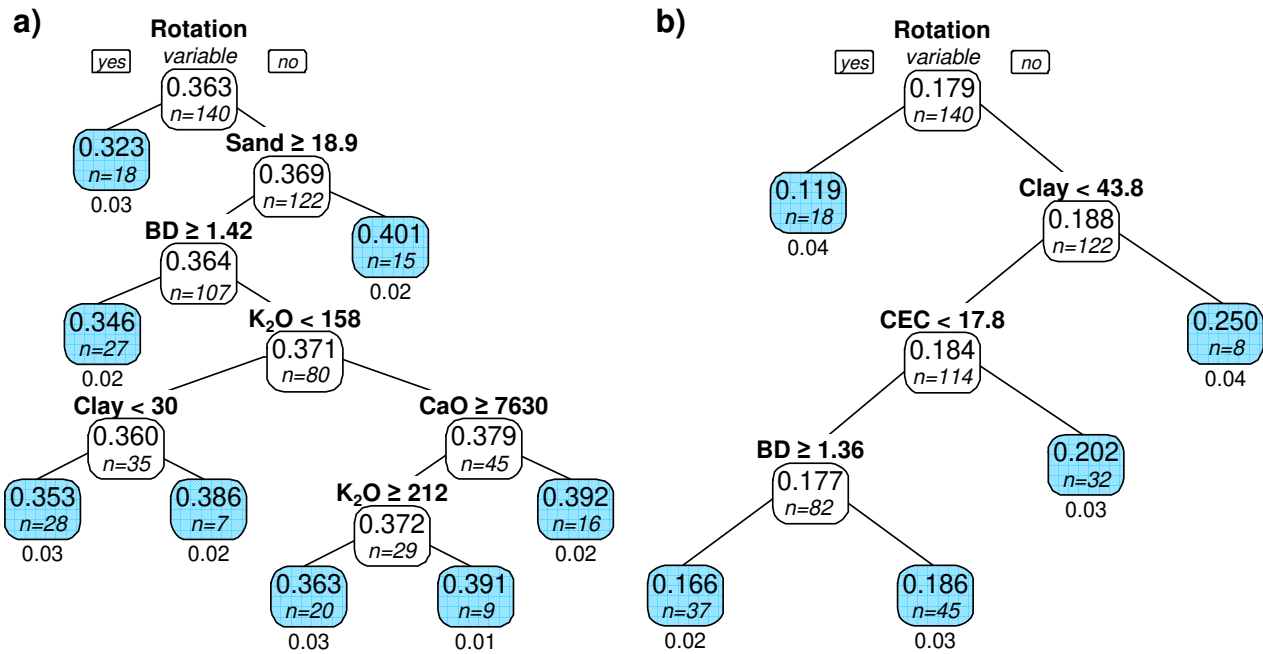


Fig. 6. Regression trees for the prediction of (a) θ_{100} and (b) $\theta_{15\,000}$. BD is bulk density (g/cm^3), CEC is cation exchange capacity (cmol/kg) and CaO and K_2O are exchangeable calcium and potassium (mg/kg), respectively. Values in boxes are mean water contents (cm^3/cm^3) of the n samples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

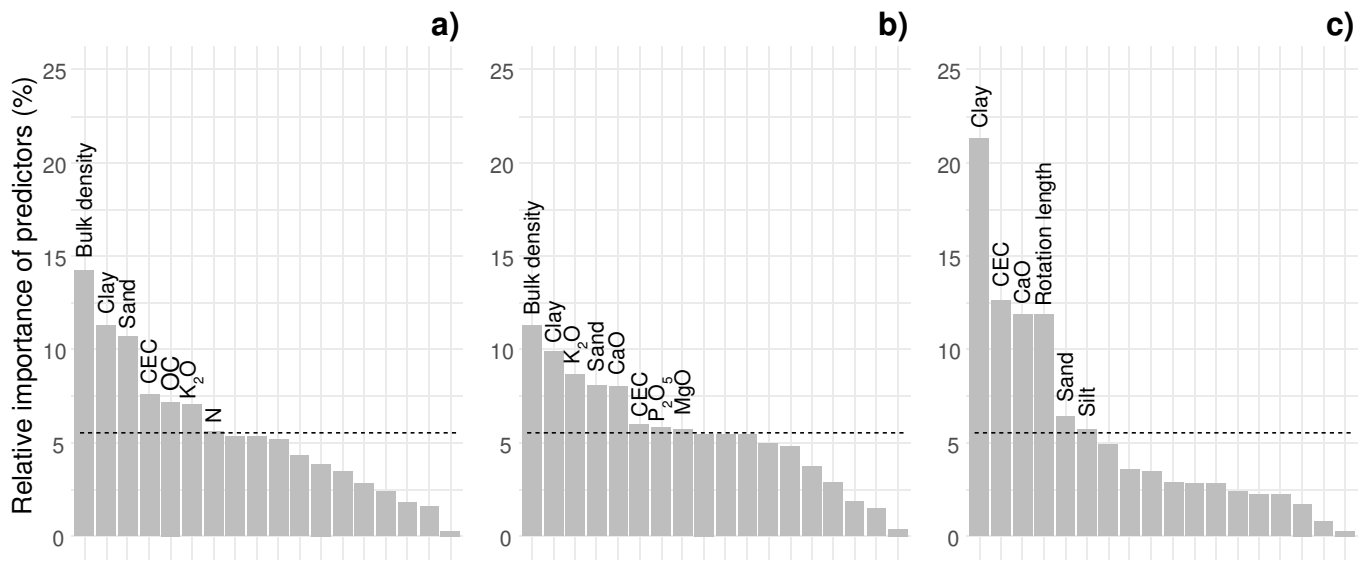


Fig. 7. Relative importance (%) of predictors in random forests of (a) θ_{100} , (b) θ_{330} and (c) θ_{15000} . Dashed lines represent the mean relative importance; only predictors above the mean are labelled. CEC is cation exchange capacity; CaO, K_2O and MgO are exchangeable calcium, potassium and 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. $\text{OC}^*=\text{OC}+1$. 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	$\theta_h = a + (b \times \text{Sa}) + (c \times \text{Si}) + (d \times \text{Cl}) + (e \times \text{OM}) + (f \times \text{BD})$	$\theta_{40}, \theta_{70}, \theta_{100}, \theta_{200},$	M1_Co
			$\theta_h = a + (b \times \text{Sa}) + (c \times \text{Si}) + (d \times \text{Cl}) + (e \times \text{OM}) + (f \times \text{BD}) + (h \times \theta_{15\ 000})$	$(\theta_{330}), \theta_{600}, \theta_{4000},$ $\theta_{7000}, \theta_{10000},$	M2_Co
			$\theta_h = a + (b \times \text{Sa}) + (c \times \text{Si}) + (d \times \text{Cl}) + (e \times \text{OM}) + (f \times \text{BD}) + (g \times \theta_{330}) + (h \times \theta_{15\ 000})$	$(\theta_{15\ 000})$	M3_Co
Bruand et al. (2004)	France, Paris basin	340	- texture AISNE (topsoil function)	$\theta_{10}, \theta_{33}, \theta_{100}, \theta_{330},$ $\theta_{1000}, \theta_{3300}, \theta_{15\ 000}$	M4_Cl
Al Majou et al. (2007)	France, Paris basin	320	- texture FAO (topsoil function)		M5_Cl M6_Cl
			- texture FAO (topsoil function)	$\theta_{10}, \theta_{33}, \theta_{100}, \theta_{330},$ $\theta_{1000}, \theta_{3300}, \theta_{15\ 000}$	M7_Cl M8_Cl
			- bulk density $-\theta_h = a + (b \times \text{Cl}) + (c \times \text{Si}) + (d \times \text{OC}) + (e \times \text{BD})$ (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)		M11_Cl M12_Cl
			- texture FAO (topsoil function)	$\theta_{10}, \theta_{33}, \theta_{100}, \theta_{330},$ $\theta_{1000}, \theta_{3300}, \theta_{15\ 000}$	M13_Cl M14_Cl
			- bulk density		
Tóth et al. (2015)	18 European countries	18 537	- texture FAO & topsoil/subsoil		M15_Tr
			- texture USDA & topsoil/subsoil		M16_Tr
			$\theta_{330} = a_1 - (b_1 \times \text{OC}^{*-1}) + (c_1 \times \text{Cl}) + (d_1 \times \text{Si}) + (e_1 \times \text{Si} \times \text{OC}^{*-1}) - (f_1 \times \text{Si} \times \text{Cl}) + (g_1 \times \text{Cl} \times \text{OC}^{*-1})$ $\theta_{15\ 000} = a_2 + (b_2 \times \text{Cl}) - (c_2 \times \text{Si}) - (d_2 \times \text{OC}^{*-1}) + (e_2 \times \text{Si} \times \text{Cl}) + (f_2 \times \text{Cl} \times \text{OC}^{*-1}) + (g_2 \times \text{Si} \times \text{OC}^{*-1})$	$\theta_{330}, \theta_{15\ 000}$	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)		M18_Cl M19_Cl
			- texture FAO (topsoil function)		M20_Cl M21_Cl
			- bulk density $\theta_h = a + (b \times \text{Cl}) + (c \times \text{Sa})$ (topsoil function)	$\theta_{100}, \theta_{330}, \theta_{15\ 000}$	M22_Co M26_Co
			$\theta_h = a + (b \times \text{Cl}) + (c \times \text{Sa}) + (d \times \text{OC})$ (topsoil function)		M23_Co M27_Co
			$\theta_h = a + (b \times \text{Cl}) + (c \times \text{Sa}) + (e \times \text{BD})$ (topsoil function)		M24_Co M28_Co
			$\theta_h = a + (b \times \text{Cl}) + (c \times \text{Sa}) + (d \times \text{OC}) + (e \times \text{BD})$ (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
			$\theta_s = a_1 + (b_1 \times Cl) - (c_1 \times BD) - (d_1 \times Si^2) + (e_1 \times OM^2) + (f_1 \times Cl^{-1}) + (g_1 \times Si^{-1}) + (h_1 \times \ln(Si)) - (i_1 \times OM \times Cl) - (j_1 \times BD \times Cl) - (k_1 \times BD \times OM) - (l_1 \times T/S \times Si)$ $\ln(\alpha) = -a_2 + (b_2 \times Cl) + (c_2 \times Si) + (d_2 \times OM) + (e_2 \times BD) - (f_2 \times T/S) - (g_2 \times BD^2) - (h_2 \times Cl^2) - (i_2 \times (OM^2)) + (j_2 \times OM^{-1}) + (k_2 \times \ln(Si)) + (l_2 \times \ln(OM)) - (m_2 \times BD \times Si) - (n_2 \times BD \times OM) + (o_2 \times T/S \times Cl)$ $\ln(n-1) = -a_3 - (b_3 \times Cl) + (c_3 \times Si) - (d_3 \times OM) + (e_3 \times BD) - (f_3 \times (BD^2)) + (g_3 \times (Cl^2)) + (h_3 \times (OM^2)) - (i_3 \times BD^{-1}) - (j_3 \times Si^{-1}) - (k_3 \times OM^{-1}) - (l_3 \times \ln(Si)) - (m_3 \times \ln(OM)) - (n_3 \times \ln(BD)) - (o_3 \times BD \times Cl) + (p_3 \times BD \times OM) + (q_3 \times T/S \times Cl)$	P2_Co	
Al Majou et al. (2008a)	France, Paris basin	320	- texture FAO		P3_Cl
			$\theta_s = a_1 - (b_1 \times Cl) - (c_1 \times BD) + (d_1 \times Si^2) - (e_1 \times OC^2) + (f_1 \times Cl^{-1}) + (g_1 \times Si^{-1}) - (h_1 \times \ln(Si)) + (i_1 \times OC \times Cl) + (j_1 \times BD \times Cl) - (k_1 \times BD \times OC) - (l_1 \times Si)$ $\ln(\alpha) = a_2 + (b_2 \times Cl) + (c_2 \times Si) + (d_2 \times OC) + (e_2 \times BD) - (f_2 \times BD^2) - (g_2 \times Cl^2) - (h_2 \times OC^2) - (i_2 \times OC^{-1}) - (j_2 \times \ln(Si)) - (k_2 \times \ln(OC)) - (l_2 \times BD \times Si) - (m_2 \times BD \times OC)$ $\ln(n-1) = -a_3 - (b_3 \times Cl) + (c_3 \times Si) - (d_3 \times OC) + (e_3 \times BD) - (f_3 \times BD^2) + (g_3 \times Cl^2) + (h_3 \times OC^2) + (i_3 \times BD^{-1}) + (j_3 \times Si^{-1}) + (k_3 \times OC^{-1}) - (l_3 \times \ln(Si)) + (m_3 \times \ln(OC)) - (n_3 \times \ln(BD)) + (o_3 \times BD \times Cl) + (p_3 \times BD \times OC)$	P4_Co	
Tóth et al (2015)	18 European countries	18 537	- texture FAO		θ_s, α, n P5_Cl
			- texture USDA		P6_Cl
			$\theta_s = 0.5056 - (0.1437 \times 1/(OC+1)) + (0.0004152 \times Si)$ $\log_{10}(\alpha) = -1.3050 - (0.0006123 \times Si) - (0.009810 \times Cl) + (0.07611 \times 1/(OC^*)) - (0.0004508 \times Si \times Cl) + (0.03472 \times Cl \times 1/(OC^*)) - (0.01226 \times Si \times 1/(OC+1))$ $\log_{10}(n-1) = 0.01516 - (0.005775 \times 1/OC^*) - (0.24885 \times \log_{10}(CEC)) - (0.01918 \times Cl) - (0.0005052 \times Si) - (0.007544 \times pH^2) - (0.02159 \times Cl \times 1/OC^*) + (0.01556 \times Cl \times \log_{10}(CEC)) + (0.01477 \times 1/OC^* \times pH^2) + (0.0001121 \times Si \times Cl) - (0.33198 \times 1/OC^* \times \log_{10}(CEC))$	P7_Co	
			$\theta_s = 0.83080 - (0.28217 \times BD) + (0.0002728 \times Cl) + (0.000187 \times Si)$ $\log_{10}(\alpha) = -0.43348 - (0.41729 \times BD) - (0.04762 \times OC) + (0.21810 \times T/S) - (0.01581 \times Cl) - (0.01207 \times Si)$ $\log_{10}(n-1) = 0.22236 - (0.30189 \times BD) - (0.05558 \times T/S) - (0.005306 \times Cl) - (0.003084 \times Si) - (0.01072 \times OC)$	P8_Co	
			$\theta_s = 0.63052 - (0.10262 \times BD^2) + (0.0002904 \times pH^2) + (0.0003335 \times Cl)$ $\log_{10}(\alpha) = -1.16518 + (0.40515 \times 1/OC^*) - (0.16063 \times BD^2) - (0.008372 \times Cl) - (0.01300 \times Si) + (0.002166 \times pH^2) + (0.08233 \times T/S)$ $\log_{10}(n-1) = -0.25929 + (0.25680 \times 1/OC^*) - (0.10590 \times BD^2) - (0.009004 \times Cl) - (0.001223 \times Si)$	P9_Co	

Table 3. Summary statistics of particle size fractions (%), organic carbon (OC; %), nitrogen content (g/kg), bulk density (BD; g/cm³), cation exchange capacity (CEC; cmol/kg), exchangeable CaO, MgO, K₂O, Na₂O (mg/kg), pH, total calcium carbonate CaCO₃ (g/kg), phosphorus content P₂O₅ (mg/kg) and volumetric water content at field capacity, θ_{100} and θ_{330} , and at wilting point, $\theta_{15\ 000}$ (cm³/cm³) 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 ₂ O	Na ₂ O	pH	CaCO ₃	P ₂ O ₅	θ_{100}	θ_{330}	$\theta_{15\ 000}$
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 $< 1.10^{-3}$

PTF	θ_{100} (cm ³ /cm ³)			θ_{330} (cm ³ /cm ³)			$\theta_{15\,000}$ (cm ³ /cm ³)		
	RMSE _P	ME _P	EF _P	RMSE _P	ME _P	EF _P	RMS E _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 θ_s , 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

PTF	θ_s (cm ³ /cm ³)			α (cm ⁻¹)			n (-)		
	$RMSE_P$	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)

$\theta_{100} = a + b \times \text{Clay} + c \times \text{BD} + d \times \text{Silt} + e \times \text{N} + f \times \text{Sand}$						
	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}
$\theta_{330} = a + b \times \text{Clay} + c \times \text{BD} + d \times \text{CEC} + e \times \text{P}_2\text{O}_5$						
	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}	
$\theta_{15\ 000} = a + b \times \text{Clay} + c \times \text{BD} + d \times \text{pH} + e \times \text{Silt}$						
	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, cm³/cm³), mean error (ME, cm³/cm³) 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