https://doi.org/10.17221/123/2022-PSE

A comparison of measured and estimated saturated hydraulic conductivity of various soils in the Czech Republic

Kamila Báťková, Svatopluk Matula, Eva Hrúzová, Markéta Miháliková, Recep Serdar Kara, Cansu Almaz*

Department of Water Resources, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Prague, Czech Republic

**Corresponding author: batkova@af.czu.cz*

Citation: Báťková K., Matula S., Hrúzová E., Miháliková M., Kara R.S., Almaz C. (2022): A comparison of measured and estimated saturated hydraulic conductivity of various soils in the Czech Republic. Plant Soil Environ., 68: 338–346.

Abstract: The study aims to indirectly determine the saturated hydraulic conductivity (Ks). The applicability of recently-published pedotransfer functions (PTFs) based on a machine learning approach has been tested, and their performance has been compared with well-known hierarchical PTFs (computer software Rosetta) for 126 soil data sets in the Czech Republic. The quality of estimates has been statistically evaluated in comparison with the measured Ks values; the root mean squared error (RMSE), the mean error (ME) and the coefficient of determination (R^2) were considered. The eight tested models of PTFs were ranked according to the RMSE values. The measured results reflected high Ks variability between and within the study areas, especially for those areas where preferential flow occurred. In most cases, the tested PTFs overestimated the measured Ks values, which is documented by positive ME values. The RMSE values of the Ks estimate ranged on average from 0.5 (coarse-textured soils) to 1.3 (medium to fine-textured soils) for log-transformed Ks in cm/day. Generally, the models based on Random Forest performed better than those based on Boosted Regression Trees. However, the best estimates were obtained by Neural Network analysis PTFs in Rosetta, which scored for four best rankings out of five.

Keywords: soil parameter; soil texture; soil property; prediction; comparative assessment

The saturated hydraulic conductivity of soil (Ks) is one of the most important and most widely-used soil parameters and is commonly applied in a number of different geotechnical, environmental, and water investigations and models (Schaap et al. 2001, Mbonimpa et al. 2002, Araya and Ghezzehei 2019, Tuffour et al. 2019). Ks refers to the ease with which the pores of saturated soil/rock transmit water (United States Department of Agriculture 2022). Ks is reported as one of the most important soil properties during the precipitation, snowmelt, flooding and irrigation events, as it determines the water flow behaviour, infiltration rate, runoff generation and deep drainage (Gamie and De Smedt 2018, Araya and Ghezzehei 2019). Various methods have been developed to determine Ks in the field and the laboratory (Klute 1986). However, for larger areas or heterogeneous areas, an unreasonably high number of replicates need to be carried out in order to account for the spatial variability of Ks. Estimates of Ks by means of pedotransfer functions (PTFs) have been researched widely over the last 30 years. Large databases of basic soil properties (i.e. the European Soil Database (ESDB), the Soil Survey Geographic Database (SSURGO)), together with a range of approaches, including high-performance computing, have been used to obtain reasonable Ks estimates. Bouma (1989) introduced the term pedotransfer function, and Minasny et al. (1999) described PTFs as "translating data we have into what we need". The concept of PTFs was based on easily measured and easily-available soil properties, such as soil texture

Supported by the Ministry of Agriculture of the Czech Republic, Czech National Agency for Agricultural Research (NAZV), the Czech Republic, Project No. QK1910086.

https://doi.org/10.17221/123/2022-PSE

and dry bulk density (BD), which were used as predictors to estimate desirable hydraulic properties (e.g. Ks). More recently, numerous PTFs have been proposed for a variety of purposes. Reviews discussing already published PTFs can be found in the works of Wösten et al. (2001), Pachepsky and Rawls (2004) and Vereecken et al. (2010). These works were mainly aimed at predicting soil water retention parameters. In the review of Zhang and Shaap (2019), a detailed description of the statistical techniques leading to the PTFs development for Ks predictions is presented.

Generally, the first types of PTFs were in the form of tabular values based on the soil texture class (e.g. Wösten et al. 1995) and linear/nonlinear regression equations (e.g. Wösten et al. 1995, Minasny et al. 1999). A more recent approach utilises Neural Network analysis (NN), which relates the basic soil properties (predictors) to the required output data (Ks) by an iterative calibration procedure. This approach has been implemented into the user-friendly Rosetta computer program, in which the models published by Shaap and Leij are utilised (Schaap et al. 1998, Schaap and Leij 2000). The current technical progress of high-performance computing and in hydraulic data collection of large databases has enabled the development of data-driven methods such as machine learning (ML). Araya and Ghezzehei (2019) presented ML-based PTFs for Ks prediction using various types of ML algorithms (K-Nearest Neighbours, Support Vector Regression, Random Forest and Boosted Regression Trees). The availability of large background soil databases implemented into the Rosetta program (Schaap et al. 2001) and ML-based PTFs (Araya and Ghezzehei 2019) made them widely applicable. In this study, the hypothesis that PTFs are robust enough to predict Ks of soils of the Czech Republic with acceptable accuracy is tested.

MATERIAL AND METHODS

Background Ks data. A total of 126 Ks measurements, together with information about soil texture, BD and organic carbon (C_{org}) content, were utilised for this study. The Ks data summarised within the HYPRESCZ database (Miháliková et al. 2013) were enriched by 46 recent own measurements. The data originates from agricultural soils in 13 localities in the Czech Republic (Figure 1). The basic information, together with the relevant soil characteristics, is presented in Table 1. The soil classification is presented in Figure 2. The USDA textural triangle consists of 12 texture classes; however, the FAO textural triangle defines 5 texture classes only. In the Czech Republic, the 12 USDA classes are grouped into 5 "grouped texture classes," according to Němeček et al. (2001), which are similar to the FAO texture classes.

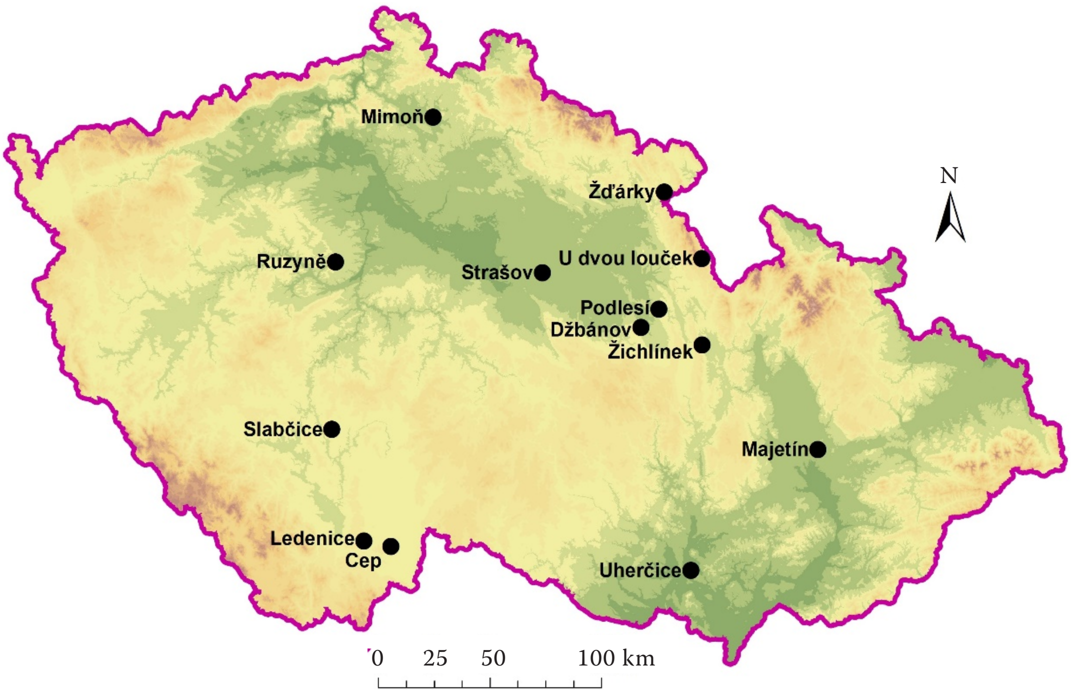

Figure 1. Location of the sites under investigation within the Czech Republic (background map: Czech Office for Surveying, Mapping and Cadastre)

https://doi.org/10.17221/123/2022-PSE

Table 1. A description of the soils used for pedotransfer function (PTF) application; data for a total of 126 soils are grouped and described in terms of dry bulk density (BD), organic carbon (C_{org}) and saturated hydraulic conductivity (Ks)

USDA texture class	Grouped texture class	Records No.	Averaged BD (g/cm³)	BD range (g/cm³)	C_{org} range (%)	Averaged C_{org} (%)	Averaged measured Ks (cm/day)
Sand	1	5	1.41	1.25–1.53	0.46–1.02	0.62	503.29
Loamy sand	1	6	1.34	1.07–1.70	0.27–1.32	0.81	178.18
Sandy loam	2	13	1.47	1.07–1.89	0.17–2.65	1.42	44.09
Loam	3	14	1.57	1.39–1.79	0.06–1.62	0.64	33.13
Silt loam	3	26	1.38	1.01–1.62	0.00–2.90	1.22	245.33
Silt	3	0	na	na	na	na	na
Sandy clay loam	4	15	1.45	1.22–1.73	0.06–3.31	2.34	87.02
Clay loam	4	16	1.55	1.26–1.75	0.06–1.69	0.61	7.42
Silty clay loam	4	23	1.39	1.13–1.74	0.08–1.83	1.02	214.04
Sandy clay	5	0	na	na	na	na	na
Silty clay	5	5	1.27	1.13–1.35	1.72–2.61	1.95	128.43
Clay	5	3	1.29	1.18–1.50	0.41–1.95	1.10	11.71

na – not applicable, as no data for this texture class was available

The Ks data were measured by different laboratory and field methods; the constant head apparatus, the falling head apparatus, pressure ring infiltrometer (Matula and Kozáková 1997) and Hood infiltrometer (Umwelt Geräte Technik, GmbH, Müncheberg, Germany) were employed. The possible effect of the measurement method was not evaluated due to the non-existence of any reference method for Ks determination. The predictors were measured by standard procedures; particle size distribution

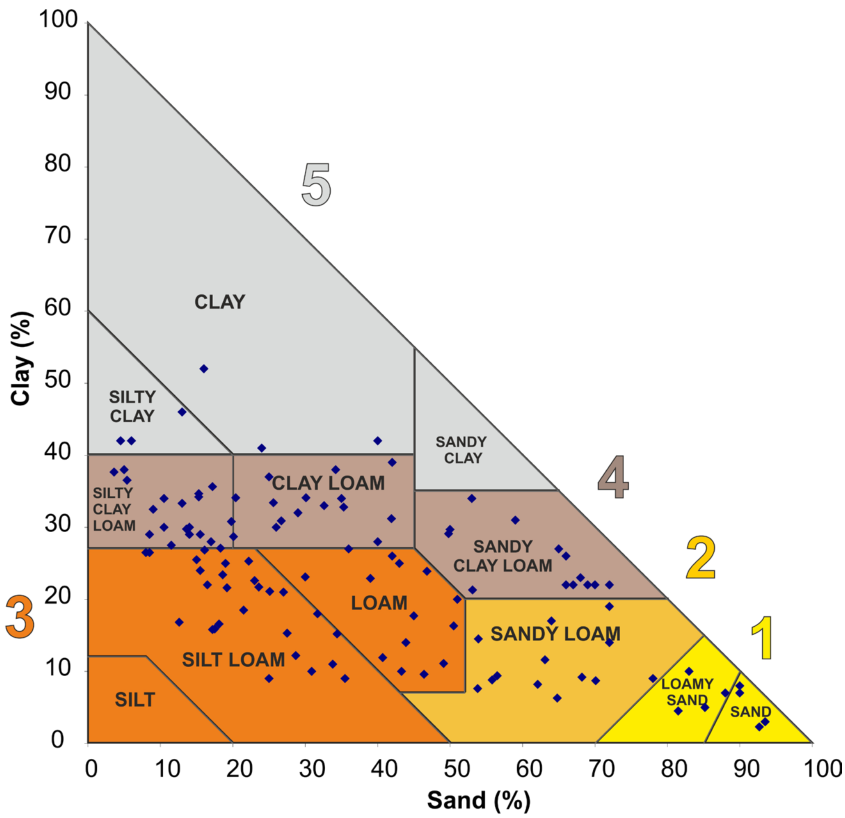

Figure 2. Particle size distribution data of soils used in this study within the USDA soil texture triangle, with coloured indications of the five grouped texture classes (from 1 to 5) according to Němeček et al. (2001)

https://doi.org/10.17221/123/2022-PSE

analysis by the Hydrometer Method, particle density by the Pycnometer Bottle Method, organic carbon C_{org} by the Walkley and Black oxidometric method, and bulk density on the basis of undisturbed soil core samples (100 cm^3 and/or 250 cm^3).

Applied PTFs. The performance of eight models of PTFs with different predictors was evaluated within this study (Table 2). Aray and Ghezzehei (2019) developed ML-based PTFs on over 18 000 soils based on four types of ML-algorithms, two of which were selected for testing within this study: Random Forest (RF) and Boosted Regression Trees (BRT). The RF method combines (averages) the decisions of the large number of individual decision trees that are "grown" individually by searching for a predictor that ensures the best split that results in the smallest model error. The RF method is reported to be relatively robust to errors and outliers (Gunarathna et al. 2019). BRT provides a form of a decision tree model ensemble with an enhancing procedure by a gradient boosting algorithm that creates additive regression models by sequentially fitting the decision trees (or any different type of "simply based learner") to the current pseudo-residuals at each iteration (Friedman 2002). Thanks to their operating principle, BRT methods are attractive in works where the training data originates from different measurement methods, as in the case of Ks measurements in the field/laboratory when different methods have been applied (Araya and Ghezzehei 2019).

Rosetta (Schaap et al. 2001) is a public domain Windows-based modelling tool for water and solute transport within a variably saturated medium. In total, 1 306 soil samples with a measured Ks value are incorporated within the Rosetta database. It offers five hierarchical PTF models for Ks prediction; two of them were tested in this study (Table 2). Neural Network can be described as a highly interconnected network consisting of many simple processing units that are referred to as neurons (by analogy with the biological neurons in the human brain). Neurons that have similar characteristics are arranged in the NN in groups that are referred to as layers. The neurons in one layer are not mutually connected, but they are connected to the neurons in the adjacent layer. The connection strength of the neurons in the adjacent layers is represented by a parameter referred to as the connection strength or the weight. The NN normally consists of three layers: the input layer, the hidden layer and the output layer (Parasuraman et al. 2006, Arshad et al. 2013).

Statistical evaluation. Ks values expressed in cm/day are presented and evaluated, as it enables comparisons with other published studies. Prior to any statistical evaluation, all Ks values were log-transformed in order to obtain their normal distribution. The performance of the tested PTFs was measured in terms of the root mean squared error (RMSE), the mean error (ME) and the coefficient of determination (R^2), as follows:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i) \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

$$R^2 = \left\{\frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \; \sum_{i=1}^{n} y_i}{\sqrt{[n\sum_{i=1}^{n} {x_i}^2 - (\sum_{i=1}^{n} x_i)^2]\,[n\sum_{i=1}^{n} {y_i}^2 - (\sum_{i=1}^{n} y_i)^2]}}\right\}^2 \tag{3}$$

Table 2. List of applied pedotransfer functions (PTFs) and their predictors

PTF model	Predictor	Reference
BRT 3-0	% sand, % silt, % clay	Araya and Ghezzehei (2019)
BRT 3-1	% sand, % silt, % clay, BD (g/cm^3)	
BRT 3-2	% sand, % silt, % clay, BD (g/cm^3, C_{org} (%)	
RF 3-0	% sand, % silt, % clay	
RF 3-1	% sand, % silt, % clay, BD (g/cm^3)	
RF 3-2	% sand, % silt, % clay, BD (g/cm^3), C_{org} (%)	
Rosetta-SSC	% sand, % silt, % clay	Schaap et al. (2001)
Rosetta-SSC, BD	% sand, % silt, % clay, BD (g/cm^3)	

BRT – Boosted Regression Trees; RF – Random Forest; BD – dry bulk density; C_{org} – organic carbon

https://doi.org/10.17221/123/2022-PSE

where: x_i – measured Ks data; y_i – predicted Ks data; n – number of x_i y_i data pairs.

The RMSE indicates the average deviation of the predicted Ks values from the measured Ks. The smaller the RMSE value is, the better the performance of the PTF prediction. The performance of each PTF model was evaluated according to its rank on a scale from 1 to 8; the best ranking value (1) was attributed to the applied PTF with the smallest RMSE value. The ME is negative if the prediction underestimates the Ks value and is positive if the PTF overestimates the measured Ks. The correspondence between the measured and predicted data is indicated by the R^2 value: the higher the R^2, the better the correspondence.

RESULTS AND DISCUSSION

A total of 126 Ks values were predicted by eight models of PTFs. The soils investigated are rather heterogeneous and involve soils from two to six USDA soil texture classes. Evaluation and ranking of each applied PTF model were carried out in terms of the individual localities and also in terms of the five grouped texture classes (Němeček et al. 2001). The data distributions through their quartiles are graphically displayed in Box and Whisker plots (Figure 3). Generally, a quite high natural variability within and between the localities was observed, especially in the case of agricultural fields, where the tillage operations can temporarily affect the topsoil hydraulic properties. Relatively low variability in measured Ks and relatively good agreement between predicted and observed Ks were found for soils with a coarser texture (Figure 3, texture groups 1 and 2). Relatively high variability in measured Ks was found for soils with medium-to-fine textures (Figure 3, texture groups 3, 4 and 5), where Ks ranged approx. from 0.1 to 1 000 cm/day. For these groups, Rosetta SSC was not able to predict the wide range of measured Ks data (light green).

The quality of the predictions can be observed on the correlation graphs, where predicted and measured Ks data are plotted. The performance of the individual applied models of PTFs for each grouped textural class is displayed in Figure 4, while the comparison for the individual localities is displayed in Figure 5. Stronger correlations can be observed for models using NN analysis and the RF algorithm for coarse-textured soils (texture groups 1 and 2). The

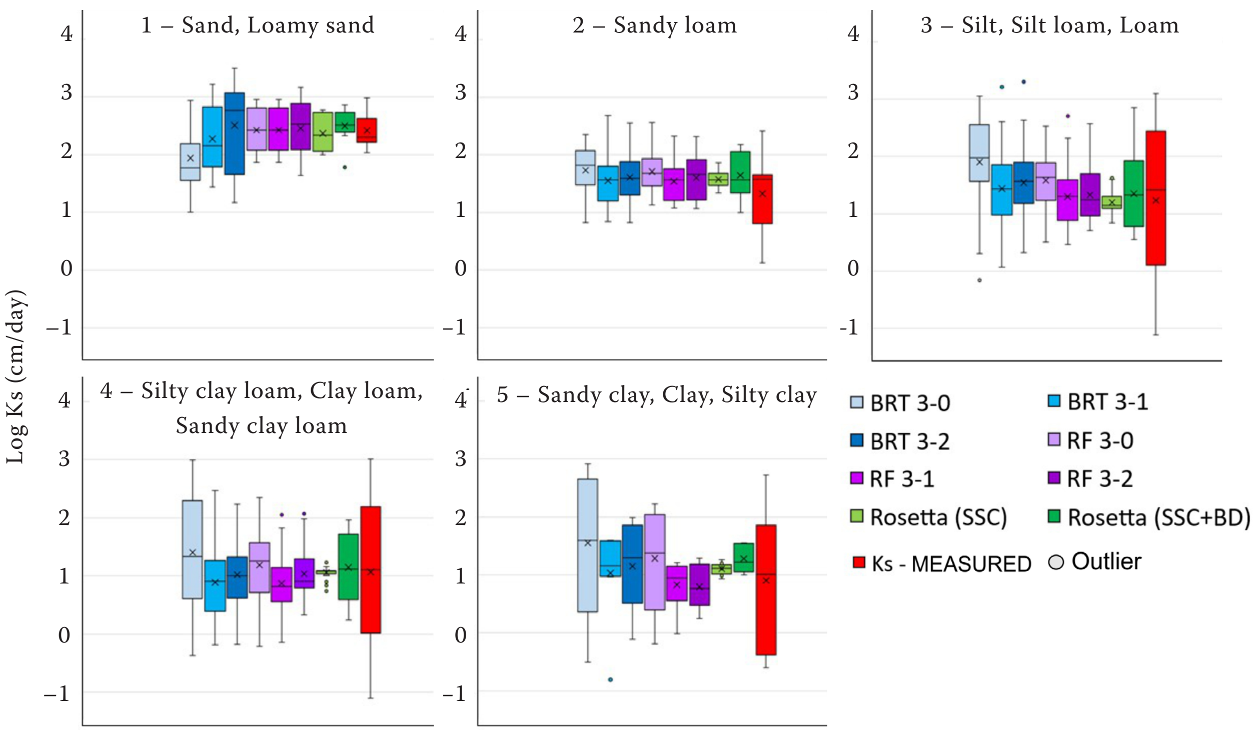

Figure 3. Comparison of the measured (in red colour) and predicted saturated hydraulic conductivity (Ks) values by means of Box and Whisker plots. BRT – Boosted Regression Trees; RF – Random Forest; BD – dry bulk density

https://doi.org/10.17221/123/2022-PSE

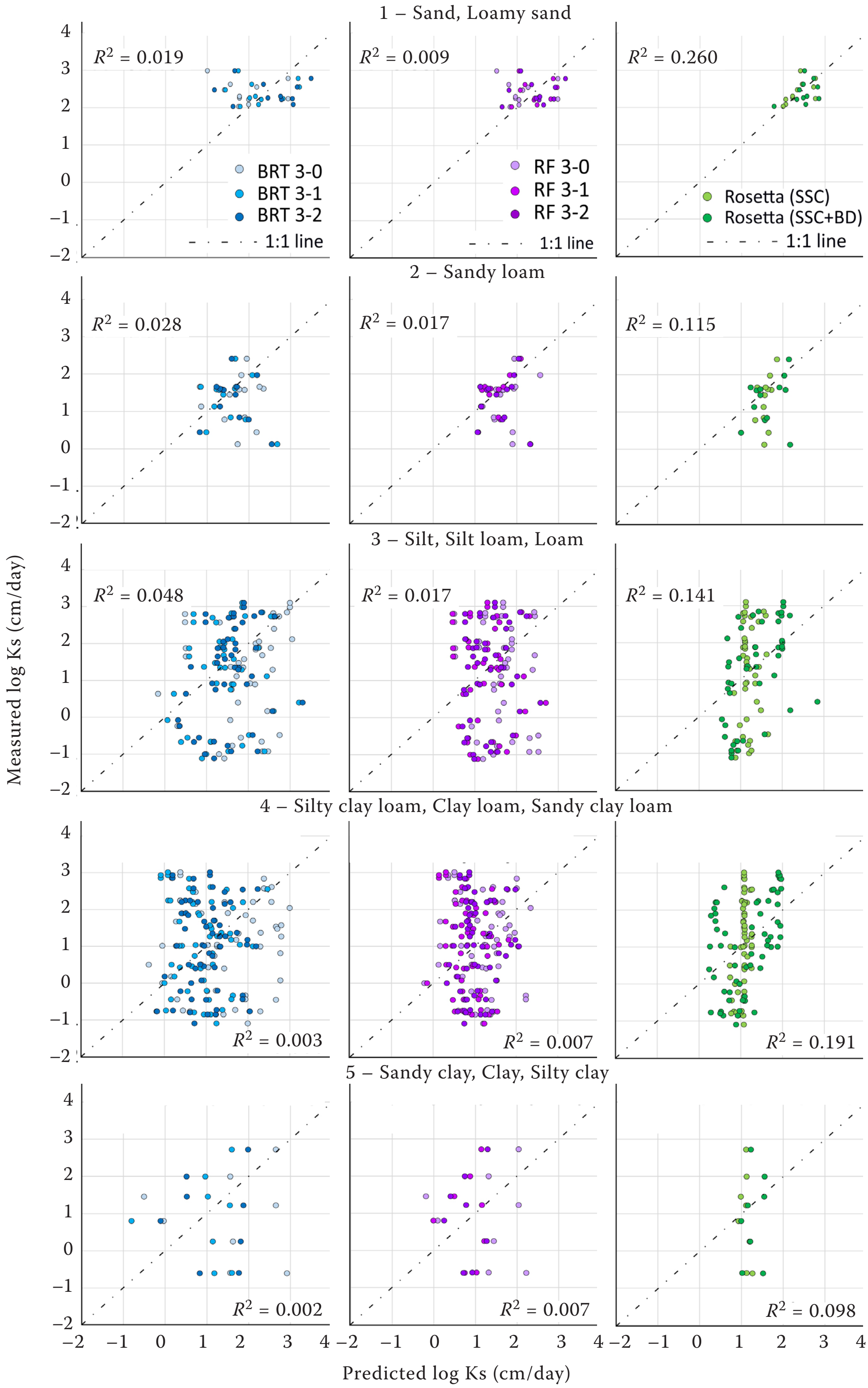

Figure 4. Correlations between the measured and predicted log-transformed Ks data for the soils in the Czech Republic with respect to their attribution to the grouped texture classes (1–5)

https://doi.org/10.17221/123/2022-PSE

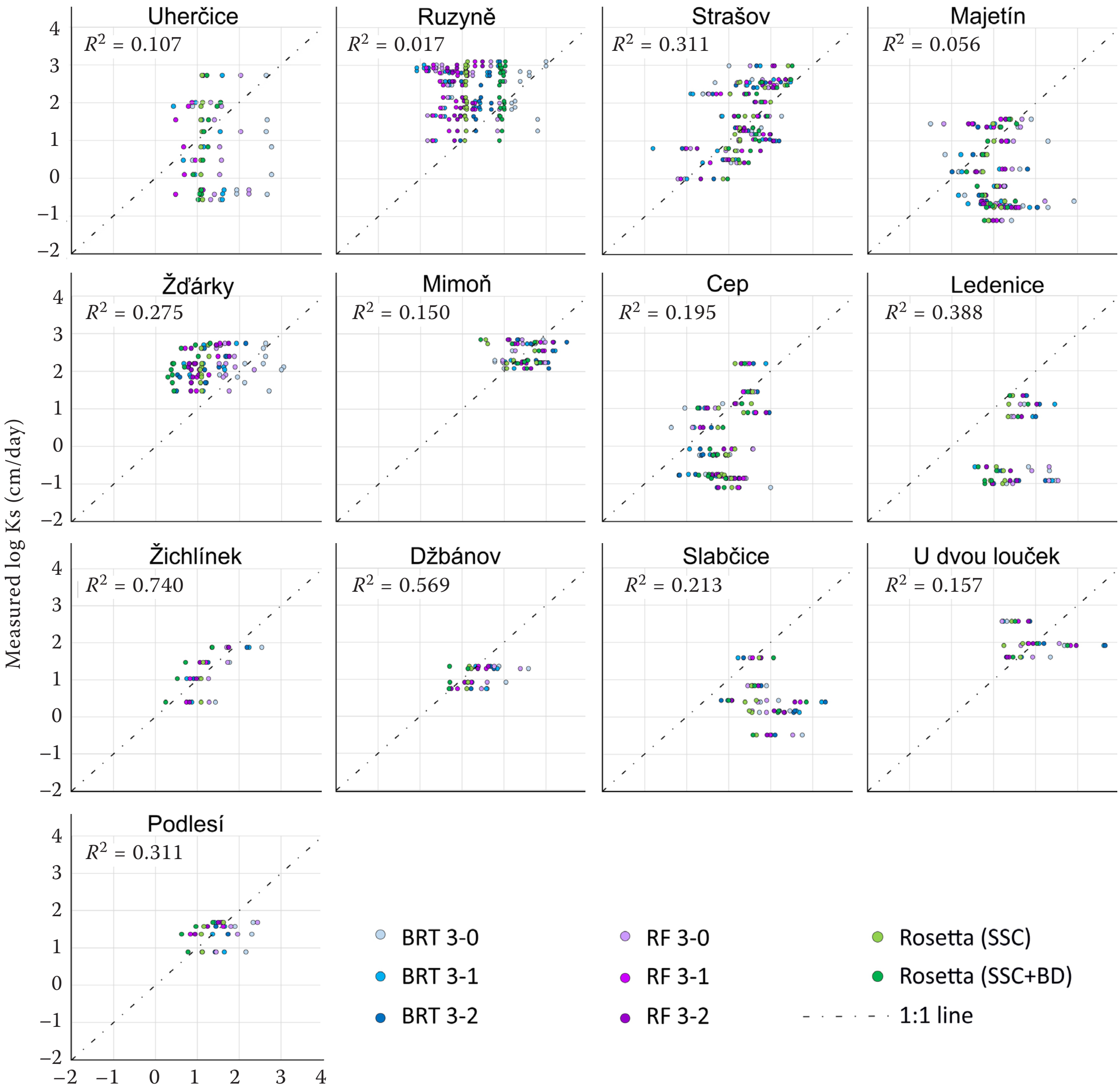

Figure 5. Correlations between the measured and predicted log-transformed saturated hydraulic conductivity (Ks) data for each of the localities in the Czech Republic. BRT – Boosted Regression Trees; RF – Random Forest; BD – dry bulk density

R^2 coefficients ranged from 0.002 (BRT models for texture class 5) to 0.260 (Rosetta models for texture class 1). Very good predictions were observed for the Žichlínek locality (R^2 = 0.740). However, a high R^2 coefficient does not always point to high-quality predictions. This is well illustrated in Figure 5, in the case of the Ledenice locality, where the R^2 coefficient reached a relatively high value of 0.388, but Ks was significantly overestimated in practically all cases. For this reason, the final evaluation and ranking of the applied PTFs were made on the basis of RMSE (Table 3). The best ranking (1) is attributed to the PTF, with the smallest RMSE value summarised for all five grouped texture classes. The effect of overestimation or underestimation of the Ks values is shown in Figure 6, where the ME for each applied PTF and grouped texture class is plotted. Sparse underestimated Ks values originated randomly from all five grouped texture classes; no trends or texture dependency can be observed for the ME values.

In conclusion, the best performance was by the Neural Network models in Rosetta, followed by the

https://doi.org/10.17221/123/2022-PSE

Table 3. Performance and the final ranking of the tested pedotransfer functions (PTFs) based on root mean squared error (RMSE)

Grouped texture class*	BRT 3-0	BRT 3-1	BRT 3-2	RF 3-0	RF 3-1	RF 3-2	Rosetta (SSC)	Rosetta (SSC + BD)
RMSE values (log Ks in cm/day)								
1 (11)	0.825	0.605	0.783	0.597	0.430	0.493	0.256	0.318
2 (13)	0.841	0.881	0.857	0.745	0.734	0.754	0.621	0.700
3 (40)	1.499	1.367	1.307	1.513	1.418	1.306	1.265	1.122
4 (54)	1.520	1.399	1.306	1.326	1.341	1.240	1.153	1.072
5 (8)	1.791	1.322	1.365	1.493	1.157	1.134	1.168	1.146
Ranking according to RMSE for each Grouped texture class								
1 (11)	8	6	7	5	3	4	1	2
2 (13)	6	8	7	4	3	5	1	2
3 (40)	7	5	4	8	6	3	2	1
4 (54)	8	7	4	5	6	3	2	1
5 (8)	8	5	6	7	3	1	4	2
Sum of rankings**	37	31	28	29	21	16	10	8
Ranking 1–5 (126)	**8**	**7**	**5**	**6**	**4**	**3**	**2**	**1**

*The values in brackets denote the number of soils within each grouped texture class. **The best ranking (1) is attributed to the PTF with the smallest value of the sum of the individual rankings within the grouped texture classes. BRT – Boosted Regression Trees; RF – Random Forest; BD – dry bulk density

Random Forest models, while the ranking of the Boosted Regression Trees models was the poorest. The prediction quality increased with an increasing number of predictors, which corresponds with the findings of Schaap et al. (2001). The Rosetta SSC-BD model, based on the known % content of clay, silt and sand particles, together with information on BD, outperformed all other models (Table 3). However, machine learning techniques have great potential and show promising results (Tóth et al. 2015, Araya and Ghezzehei 2019). The RMSE values for the models using RT reported by Lilly et al. (2008) were on an average 0.97; Tóth et al. (2015) reported an RMSE range from 0.90 to 1.36, while RMSE reported by Araya and Ghezzehei (2019)

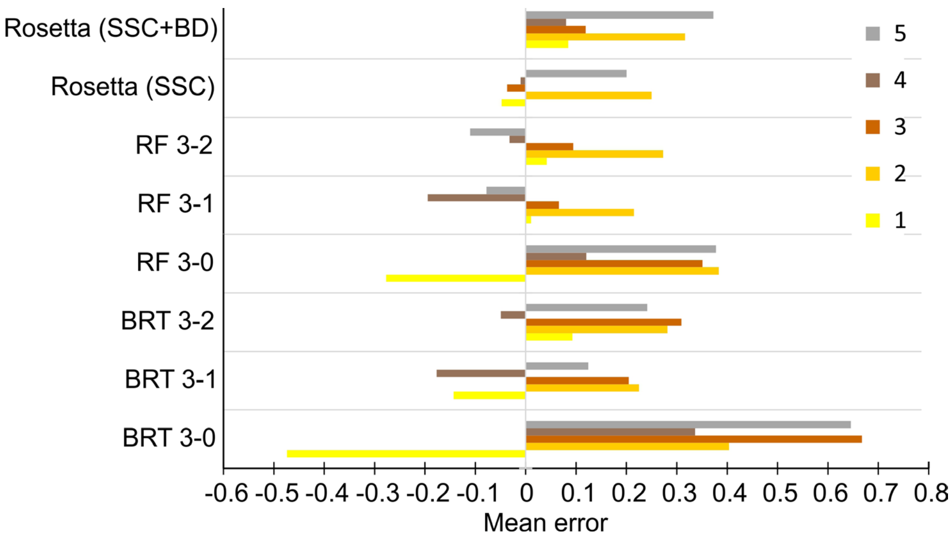

Figure 6. Performance evaluation of the tested pedotransfer functions (PTFs) by means of mean error (ME) for the grouped textural classes (1–5); negative values of ME refer to an underestimation in comparison with the measured values (log Ks in cm/day). BRT – Boosted Regression Trees; RF – Random Forest; BD – dry bulk density

https://doi.org/10.17221/123/2022-PSE

reached 0.34–0.44 for the BRT models and 0.37–0.44 for RF. In our study, comparable results with RMSE < 1 were obtained by all eight applied models of PTFs only for the grouped soil texture classes 1 and 2 (sand, loamy sand and sandy loam). A possible reason for not scoring higher might be the properties of the soils within the background soil database of PTFs published by Araya and Ghezzehei (2019), which contains mostly soils with a coarse texture; sand, loamy sand, sandy loam, sandy clay loam. In our upcoming work, we therefore plan to involve soil data from this study into the background database and repeat the performance testing of the PTFs.

Acknowledgement. We would like to thank to Mr Robin Arthur Healey for his throughout English editing of the manuscript.

REFERENCES

Araya S.N., Ghezzehei T.A. (2019): Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations. Water Resources Research, 55: 5715–5737.

Arshad R.R., Sayyad G., Mosaddeghi M., Gharabaghi B. (2013): Predicting saturated hydraulic conductivity by artificial intelligence and regression models. ISRN Soil Science, 2013: 308159.

Bouma J. (1989): Using soil survey data for quantitative land evaluation. Advances in Soil Sciences, 9: 177–213.

Friedman J.H. (2002): Stochastic gradient boosting. Computational Statistics and Data Analysis, 38: 367–378.

Gamie R., De Smedt F. (2018): Experimental and statistical study of saturated hydraulic conductivity and relations with other soil properties of a desert soil. European Journal of Soil Science, 69: 256–264.

Gunarathna M.H.J.P., Sakai K., Nakandakari T., Momii K., Kumari M.K.N. (2019): Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. Water, 11: 1940.

Klute A.E. (1986): Methods of Soil Analysis, Part 1. Physical and Mineralogical Methods. Monograph 9. Madison, ASA and SSSA.

Lilly A., Nemes A., Rawls W.J., Pachepsky Y.A. (2008): Probabilistic approach to the identification of input variables to estimate hydraulic conductivity. Soil Science Society of America Journal, 72: 16–24.

Matula S., Kozáková H. (1997): A simple pressure infiltrometer for determination of soil hydraulic properties by *in situ* infiltration measurements. Rostlinná výroba, 43: 405–413.

Mbonimpa M., Aubertin M., Chapuis R.P., Bussière B. (2002): Practical pedotransfer functions for estimating the saturated hydraulic conductivity. Geotechnical and Geological Engineering, 20: 235–259.

Miháliková M., Matula S., Doležal F. (2013): HYPRESCZ – database of soil hydrophysical properties in the Czech Republic. Soil and Water Research, 8: 34–41.

Minasny B., McBratney A.B., Bristow K.Y. (1999): Comparison of different approaches to the development of pedotransfer functions for water retention curves. Geoderma, 93: 225–253.

Němeček J., Macků J., Vokoun J., Vavříček D., Novák P. (2001): The Taxonmic Classification System of Soils in the Czech Republic. Prague, Czech University of Life Sciences Prague, Research Institute for Soil and Water Conservation. ISBN 80-238-8061-6 (In Czech)

Pachepsky Y.A., Rawls W.J. (2004): Development of Pedotransfer Functions in Soil Hydrology. Developments in Soil Science. Amsterodam, Elsevier.

Parasuraman K., Elshorbagy A., Si B. (2006): Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. Soil Science Society of America Journal, 70: 1851–1859.

Schaap M.G., Leij F.J. (2000): Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model. Soil Science Society of America Journal, 64: 843–851.

Schaap M.G., Leij F.J., van Genuchten M.T. (1998): Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Science Society of America Journal, 62: 847–855.

Schaap M.G., Leij F.J., van Genuchten M.T. (2001): Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology, 251: 163–176.

Tóth B., Weynants M., Nemes A., Makó A., Bilas G., Tóth G. (2015): New generation of hydraulic pedotransfer functions for Europe. European Journal of Soil Science, 66: 226–238.

Tuffour H., Abubakari A., Agbeshie A., Khalid A., Tetteh E., Keshavarzi A., Bonsu M., Quansah C., Oppong J., Danso L. (2019): Pedotransfer functions for estimating saturated hydraulic conductivity of selected benchmark soils in Ghana. Asian Soil Research Journal, 2: 1–11.

United States Department of Agriculture, Natural Resources Conservation Service. National soil survey handbook, title 430-VI. http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/ref/?cid=nrcs142p2_054242 (accessed 4 March 2022).

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 Zone Journal, 9: 795–820.

Wösten J.H.M., Finke P.A., Jansen M.J.W. (1995): Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma, 66: 227–237.

Wösten J.H.M., Pachepsky Y., Rawls W.J. (2001): Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. Journal of Hydrology, 251: 123–150.

Zhang Y., Schaap M.G. (2019): Estimation of saturated hydraulic conductivity with pedotransfer functions: a review. Journal of Hydrology, 575: 1011–1030.

Received: April 12, 2022
Accepted: June 21, 2022
Published online: July 25, 2022