DATASET FOR CREATING PEDOTRANSFER FUNCTIONS TO ESTIMATE ORGANIC LIQUID RETENTION OF SOILS

Hilda Hernádi[*], András Makó

Department of Crop Production and Soil Science, Georgikon Faculty, University of Pannonia, Deák F. u. 16,
H-8360 Keszthely, Hungary
*Corresponding author, e-mail: hhilduci@gmail.com

Abstract

Soil properties characterising pressure-saturation relationships (P-S), such as the fluid retention values or the fitting parameter of retention curves are basic input parameters for simulating the behaviour and transport of nonaqueous phase liquids (NAPLs) in subsurface. Recent investigations have shown the limited applicability of the commonly used estimation methods for predicting NAPL retention values in environmental practice. Alternatively, building pedotransfer functions (PTFs) based on the easily measurable properties of soils might give more accurate and reliable results for estimating hydraulic properties of soils and enable the utilisation of the wide range of data incorporated in Hungarian and international datasets. In spite of the availability of several well-established PTFs to predict the water retention of soils only a limited amount of research has been done concerning the NAPL retention of soils. Thus, in our study, data from our recent NAPL and water retention measurements were collected into a dataset containing the basic soil properties as well. Relationships between basic soil properties and fluid retention of soils with water or an organic liquid (Dunasol 180/220) were investigated with principal component analysis. NAPL retention of soil samples were determined with PTFs, based on basic soil properties and their derived values, and using a scaling method. Result of the statistical analysis (SPSS 13.1) revealed that using PTFs could be a promising alternative and could give more accurate results compared to the scaling method both for determining the NAPL saturation or the volumetric NAPL retention values of soils.

Keywords: NAPL retention, pedotransfer function, hydraulic characteristics, Leverett equation

INTRODUCTION

From the 1980s increased attention has been received to improve understanding the relevant characteristics and processes in flow and transport of subsurface nonaqueous phase liquids (NAPL) firstly in petrol industry and later to help design remedial strategies. Knowledge of pressure-saturation (P-S) relationships is essential for simulating the fate and transport of nonaqueous phase liquids (NAPLs) in subsurface with any type of models. Nowadays, as the measurements of hydraulic parameters are costly and time consuming the development of accurate estimation method is preferred.

In environmental practice the NAPL retention of soils is commonly determined by the table of average empirical pressure-saturation (P-S) values (dePastrovitch et al., 1979), table of average fitting parameters of the van Genuchten equations proposed by Carsel and Parrish (1988) (e.g. in RETZ model – van Genuchten et al., 1991; HSSM model – Weawer et al., 1994), the different modified versions of the Leverett function (Leverett, 1941) or their combination. In addition, Beckett and Joy (2003) created a dataset of calculated fitting parameters based on the modified

scaling method suggested by Lenhard and Parker (1987). Nevertheless, tables of average NAPL retention or fitting parameter values of fluid retention curves do not accurately represent the variability of soils with different physical and chemical properties. All type of the modified Leverett function is valid for ideal porous systems. These estimation methods do not take into account the different interactions between the various fluids and the porous media, thus their application might be limited for natural soils (i.e. well aggregated, higher organic matter or clay content, etc.) soils. Furthermore, they have not been properly validated (only a few methods validated with column experiments, but these were carried out in most cases with glass beads and/or sands) (Rathfelder and Abriola, 1996; Makó and Hernádi, 2013).

In the commonly used fluid retention measurement methods usually only the main drainage curves are measured. The more accurate determination of the main drainage curve of soil has key importance in determining the transport parameters (e.g. residual organic liquid content, penetration depth and time). In addition, main drainage curves could be the basis of determining the hysteretic and scanning curves in hysteretic models.

Predicting of the physical, chemical and biological properties of soils with pedotransfer functions (PTFs) is a fast-developing field and several well-established PTFs are available to predict the water retention of soils both in Hungary (Rajkai, 2004; Rajkai et al., 1996; 2004; Nemes, 2003; Makó et al., 2005 and Tóth, 2011) and abroad (Minasny et al., 1999; Wösten et al., 1999; Rawls et al., 2001, etc.). Most of these estimation methods have already been incorporated into numerical algorithms such as SOILPAR 2.0 (Acutis and Donatelli, 2003), Neuro Multistep (Minasny et al., 2004), TALA-JTANonc 1.0 (Fodor and Rajkai, 2011) or the k-Nearest (Nemes et al., 2008), etc. However, only a few studies had begun focusing on creating PTFs concerning the soils organic liquid retention capacity.

Experiences in creating PTFs for water retentions may be essential tool for obtain the best possible estimation method for predicting the NAPL retention. PTFs for water retention (as response variable) may be created for the measured point values of the pressure saturation curves (point estimation) or for the fitting parameters of the hydraulic functions (parameter estimation) (Brooks and Corey, 1964; Brutsaert, 1966; van Genuchten, 1980) based on the easily measurable basic soil properties or their derived values (predictor variables) (Wösten, 1995).

The parameter estimation methods are widely used in environmental and soil hydrological practices because the simulation models mainly use these fitting parameters as input variable. Moreover, these parameters are equal to those of the equations predicting the soil hydraulic conductivity and relative permeability (van Genuchten, 1980; Lenhard and Parker, 1987; Chen et al., 1999). If the results of two point PTFs were to be compared or if no measured water retention points are available for a particular PTF, (or vice versa) the comparison of the retention curves with using the fitted hydraulic parameters (Minasny et al., 1999; Rawls et al., 2001) or calculating the estimated water contents at the desired pressure heads by linear interpolation were suggested (Tietje and Tapkenhinrichs 1993; Schaap and Leij, 2001).

For parametric methods it is common to predict logarithmic transformed values of α (ln α) and n (ln n-1) to convert the distribution of the parameters into a more statistically normal distribution (e.g., Rawls and Brakensiek, 1985; Wösten et al, 1999). Both in case of water and NAPL retention estimation, application of similar soil properties (bulk density or texture class information, organic matter and carbonate content, etc.) or their inherited values (e.g. the averaged values of particle size data) were suggested as independent variables (Makó, 2004; Makó and Elek, 2006). In case of predicting the water retention the texture, morphology etc. are commonly used as a grouping factor in developing PTFs (Wösten et al. 1995; Schaap et al. 1999, 2001; Pachepsky and Rawls 2004). Statistical attributes of comparing PTFs for predicting water retention can potentially be adapted for the investigation of NAPL retention estimation methods. Recently, the determination of accuracy (with the working dataset), uncertainty and reliability (with test data) of the predictions, R, R2, mean error (ME), mean square error (MSE), root mean square error (RMSE), the unbiased root mean square errors (URMSE), etc. and their complete calculation are recommended, to acchieve a comprehensive verification (Pachepsky and Rawls, 2004). For the comparison of the accuracy of the fitted fluid retention curves the calculation of ZAPF values were suggested by Rajkai (2004). The AIC (Akaike Information Criterion) value offers the possibility to compare the efficiency of different estimation or fitting methods with various numbers of parameters and the models with lower error can be selected with Fisher's test (Rajkai, 2004). The uncertainty in input data can be evaluated using Monte Carlo analysis (Minasny et al. 1999), procedures based on fuzzy rules (McBratney et al., 2002) or with the Bootstrap Method (Carsel and Parish, 1988).

Nowadays, the development of inference systems (e.g. SINFER) to select the proper PTFs with minimum variance based on logical rules (McBratney et al., 2002)., or the application of data driven methods, e.g. support vector machines (Lamorsky et al., 2008) might be a challenging topic. The stability of the estimated coefficients can be investigated using the double cross-validation techniques (K-fold, Leave-one-out, Jackknife or Delete-d methods) with randomly split the data (Pachepsky and Rawls, 2004). According to Tóth et al., (2013) it may be sufficient to random spilt the data in proportion of 90:10 (working and test data). Besides, many authors suggested the detection of outliers for calibrating PTFs. To our current knowledge, there is only one research had begun for creating PTFs concerning the soils organic liquid retention in Europe (Makó, 1995; Makó, 2002; 2004). In spite of the large number of measurements to determine the NAPL retention of soils, databases from measured NAPL retention data have not been created until now. Beckett and Joy (2003) created a dataset of calculated fitting parameters based on the modified scaling method suggested by Lenhard and Parker (1987) and Parker et al. (1987) but this contains the scaled NAPL retention values calculated from the fitted water retention values of HYPRES database. Creating PTFs for NAPL-retention might be promising because a large amount of measured basic data is available in national and Hungarian databases (HYPRES, UNSODA v2.0, HUNSODA, EU HYDI, MARTHA etc.), which have already been used effectively in developing PTFs for estimating water retention (Wösten et al., 1995; Nemes et al., 2003, 2008; Makó et al., 2005; Tóth et al., 2006, 2013; Lilly, 2010). Moreover, the up-to-date hydrodynamic and transport models enable the adaptation of GIS (Geographic Information Systems) datasets (e.g. GMS - Groundwater Modelling System and Argus Open Numerical Environments - ARGUS ONE), which allow for cartographic representation with different commonly used software applications in the environmental engineer-

ing practice (such us SURFER-GRAPHER or AUTOCAD). In the 1990s a series of investigations for measuring NAPL retention of soils and mineral mixtures with the pressure plate method to create PTFs for organic liquid retention, in Hungary were started (Makó, 1995; Makó, 2002).

In this study a dataset from these recent measurements was created and analysed by statistical methods (SPSS 13.1). After the preliminary analysis (descriptive statistics and outlier detections), PTFs were built for predicting the fitting parameters of NAPL retention curves. Then, the NAPL retention of soils was predicted with classPTFs (for selected four texture groups) and using the scaling method of Lenhard and Parker (1987). Afterwards, the accuracy and reliability of predicting NAPL retention with different estimation methods were compared.

DESCRIPTION OF THE DATASET

The dataset contains the physical and chemical properties and the fluid retention data of five measurement series collected from 1991 to 2011 (*Table 1-2*). Dataset contains 369 disturbed and undisturbed samples, 40 soil profiles with 107 genetic layers and various types of soils (*Fig. 1*) for 10 texture classes are represented (*Fig. 2*).

Table 1 The subsets of dataset

Subset	N	%	Origin
1	111	30.1	undisturbed samples of 9 soil profiles of the *research* program of the *Hungarian* National Long-term Fertilization (Hernádi and Makó, 2011a)
2	123	33.3	undisturbed samples of 12 soil profiles of an investigation for the Hungarian Gas & Oil Company Plc (MOL Rt.) (Makó, 2002; 2005)
3	45	12.2	disturbed samples of mineral mixture series (Hernádi et al., 2011b) (Makó and Marczali, 1999)
4	60	16.3	disturbed samples of aggregate series separated from the upper „A" layer of selected soils (2.0 mm>, 1.0 mm>, 0,5 mm>, 0,25 mm> and 0,056 mm>) (Makó and Elek, 2006)
5*	30	8.1	disturbed samples investigated in the course of the TÁMOP-4.2.1/B-09/1/KONV-2010-0003 Mobility and Environment Project (Hernádi et al., 2011)
Sum	369	100	

* before fluid retention measurements, samples were held 24hr in water for desaggregation and dried on 40°C/24hr

Table 2 Soil properties in the dataset

	N	N%	water retention			NAPL retention		
			vol.%	N	N%	vol.%	N	N%
Surface area (BET) ($cm^2 g^{-1}$) (Brunaer et al., 1938)	144	39.0	pF 0.0	308	83.5	0 mbar	245	66.4
Plasticity according to Arany (%) (Buzás, 1993)	339	91.9	pF 0.2	170	46.1	2 mbar	361	97.9
Particle size distribution (%)*	312	84.6	pF 1.0	111	30.1	20 mbar	242	65.6
Particle size distribution (%) **	369	100.0	pF 1.3	59	16.0	50 mbar	245	66.4
Organic matter content (%) ***	369	100.0	pF 1.5	279	75.6	100 mbar	116	31.4
$CaCO_3$ (%)***	369	100.0	pF 1.7	59	16.0	150 mbar	245	66.4
Acidity (Na_4OAc) (cmol (+) kg^{-1})***	66	17.9	pF 2.0	156	42.3	200 mbar	116	31.4
Acidity (KCl) (cmol (+) kg^{-1})***	66	17.9	pF 2.2	59	16.0	400 mbar	245	66.4
Salt content (mass %)***	93	25.2	pF 2.3	111	30.1	500 mbar	116	31.4
Exchangeable Na (cmol(+)/kg)***	87	23.6	pF 2.5	168	45.5	1000 mbar	361	97.8
Base saturation (w %)***	24	6.5	pF 2.6	59	16.0	1500 mbar	30	8.1
Aggregate size distribution (%)*	105	28.5	pF 2.7	111	30.1			
pH (distilled water 1 : 2,5)***	273	74.0	pF 3.0	59	16.0			
pH (KCl suspension 1 : 2,5)***	189	51.2	pF 3.2	30	8.1			
Bulk density (water) (g cm^{-3})*	348	94.3	pF 3.4	111	30.1			
Bulk density (NAPL) (g cm^{-3})*	369	100.0	pF 4.2	279	75.6			
			pF 6.2	369	100.0			

* Physical properties of soils were determined according to MSZ 08 0205:1978 standard (PSD – pipette method)
**PSD of soils is determined according to ISO 11277:1995 with pipette method, after the total desaggregation of soil samples (eliminating the soil organic matter and $CaCO_3$ content and the iron oxides).
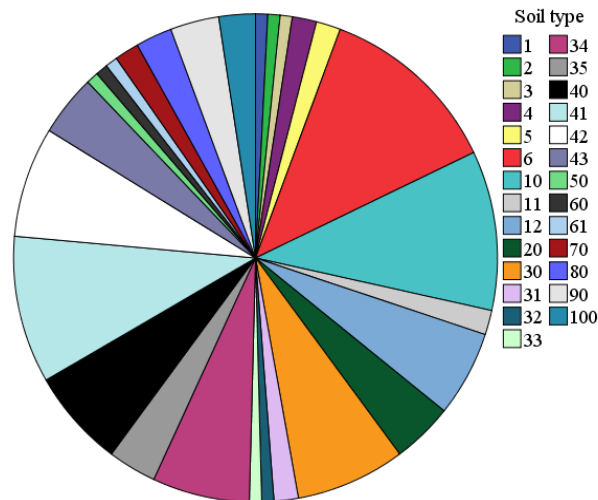***Chemical properties of soils were determined according to MSZ 08 0206/2:1978 standard

Fig. 1 WRB (World Reference Base) soil types, clay minerals and mineral mixtures in the dataset
(1: Benite, 2: Kaolin, 3: Illite, 4: Loess, 5: Pannon sand, 6: Mineral mixtures, 10: Eutric Cambisol, 11: Hortic Terric Cambisol, 12: Dystric Cambisol, 20: Haplic Arenosol, 30: Gleyic Luvisol, 40: Calcic Phaeosem, 41: Luvic Phaeozem, 42: Calcaric Phaeozem, 43: Haplic Phaeozem, 50: Vermic Calcic Chernozem, 60: Vertic Stagnic Solonetz, 61: Orthic Solonetz, 70: Gleyic Vertisol, 80: Calcaric Gleysol, 90: Eutric Regosol, 100: Calcaric Fluvisol)
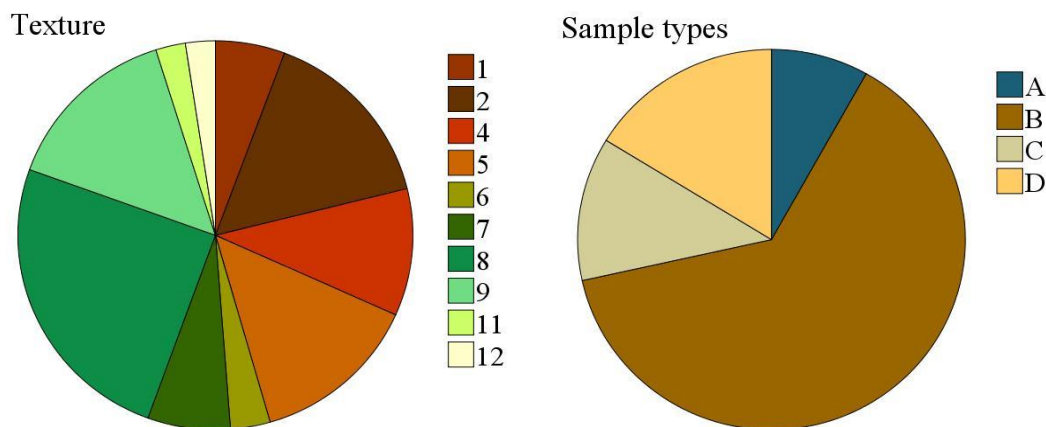

Fig.2 Texture and sample types of soil, clay minerals and mineral mixture samples
(1: Clay, 2: Silty clay, 4:Clay loam, 5: Silty clay loam, 6: Sandy clay loam, 7: Loam, 8: Silt loam, 9: Sandy loam, 11:Loamy sand, 12: Sand;
A: Disturbed samples, B: Undisturbed samples, C: Mineral mixtures, D: Aggregate series of soil samples)

METHODS

The basic soil properties, used in this investigation were measured according to the Hungarian standards (MSZ08 0205:1978 and MSZ08 0206/2:1978). The texture of soil samples was determined according to the ISSS (International Society of Soil Science) texture triangle (*Table 3*).

The fluid retention measurements were performed with distilled water and a special nonaromatic organic liquid, DUNASOL 180/220 (Hungarian Gas & Oil Company Plc. - MOL Rt.).

Water retention measurements were carried out with porous pressure plate extractors (Soilmoisture Corp. LAB 023). For the purpose of determining the NAPL retention of the samples, a modified version (as described by Makó, 1995) of these pressure plate extractors were used.

Samples showing extremely high or low NAPL retention values were selected with outlier detection. Only the samples, were the casewise diagnostic of the preliminary linear regression between soil parameters and fluid retention values showed significant differences with more than two times the standard deviation at a given pressure level, were eliminated. Then only the records where both the water and NAPL retention were measured were retained. Therefore, the final dataset contained 316 samples.

The fluid retention measurements were performed with distilled water and a special nonaromatic organic liquid, DUNASOL 180/220 (Hungarian Gas & Oil Company Plc. - MOL Rt.).

Water retention measurements were carried out with porous pressure plate extractors (Soilmoisture Corp. LAB 023). For the purpose of determining the NAPL retention of the samples, a modified version (as described by Makó, 1995) of these pressure plate extractors were used.

Table 3 Descriptive statistics of the selected basic soil properties used in this study

	Method	Unit	Mean	SD	SE	Med.	Min.	Max.	N
Clay			27.58	0.95	16.75	24.08	1.25	67.67	369
Silt	MSZ 08 0205:1978	w %	44.22	0.91	16.11	44.39	0.05	78.50	369
Sand			28.21	1.43	25.20	21.43	0.63	98.62	369
OM	MSZ 08 0206/2:1978	w %	1.43	0.06	1.24	1.21	0.00	5.66	369
CaCO$_3$	MSZ 08 0206/2:1978	w %	5.57	0.43	8.26	0.16	0.00	30.71	369
Bulk density (water)*	MSZ 08 0205:1978	g cm^{-3}	1.43	0.01	0.19	1.44	0.89	2.20	369
Bulk density* (NAPL)	MSZ 08 0205:1978	g cm^{-3}	1.42	0.01	0.26	1.45	0.80	1.99	369

*Samples for water and NAPL retention measurements had different bulk density.

Samples showing extremely high or low NAPL retention values were selected with outlier detection. Only the samples, were the casewise diagnostic of the preliminary linear regression between soil parameters and fluid retention values showed significant differences with more than two times the standard deviation at a given pressure level, were eliminated. Then only the records where both the water and NAPL retention were measured were retained. Therefore, the final dataset contained 316 samples.

For fitting the NAPL retention curves the van Genuchten equation with three parameters (Eq. 1) (van Genuchten et al., 1980) were used as initial values in nonlinear regression (SPSS 13.1, Nonlinear regression/quadratic sequential programming) with the average fitting parameters proposed by Carsel and Parish (1988) (used in e.g. RETZ and HSSM models).

$$\theta_{(h)} = \frac{\theta_s}{1 + \left((\alpha * h)^n\right)^{1-\frac{1}{n}}} \qquad (1)$$

where: $\theta_{(h)}$ is the volumetric fluid content at potential h (kPa); θ_s is the saturated fluid content; α, and n are fitting parameters.

Afterwards, NAPL retention values for 10 pressure levels (0.01, 0.3, 10, 33, 100, 200, 330, 1000, 15849 and 1584893 mbar) were calculated.

The accuracy of fitting was verified by the calculation of Pearson R^2, RMSE (Eq. 2) and ZAPF values (Eq. 3) (Rajkai, 2004).

$$RMSE = \sqrt{\frac{\sum_1^N (y_i - \hat{y}_i)^2}{N}} \qquad (2)$$

where: y$_i$ is the fitted NAPL retention, ŷ$_i$ the estimated NAPL retention and N is the number of samples.

In case of ZAPF values (Eq. 3.), the lower |ZAPF| than measurement error (ME) of NAPL retention was regarded as sufficiently fitted (as suggested by Rajkai, 2004).

$$ZAPF = \frac{\sum_{i=1}^n |\theta_e - \theta_m|}{n} \qquad (3)$$

where: θ_m is the measured fluid retention value θ_e estimated fluid retention values; n is the number of fitting point of fluid retention curves.

Before creating PTFs the dataset was randomly split with the ratio of 90:10 (calibration and validation data set), thus we had opportunity to investigate the accuracy and the also the reliability of estimations (*Table 4*). In further investigations the fitted fluid retention data was used, which enabled us to compare the retention data determined with both fluid types measured at different pressure levels. Relationships between basic soil properties and fluid retentions, as well as between soil properties and fitting parameters were preliminarily investigated with principal component analysis (PCA) (SPSS 13.1/Varimax rotation with Kaiser normalization).

Subsequently, parametric PTFs were built as class PTFs for selected four texture groups (silty clay, silty clay loam, silt loam and sandy loam) with multiple linear regression (SPSS 13.1 Linear regression/Stepwise method). In linear regression, bulk density, clay, silt, CaCO$_3$ and organic matter content and their derived values, such as the logarithmic, reciprocal, squared values and their interactions were used as independent variables (as suggested by Wösten, 1995).

The NAPL retention of the soil samples were predicted with a modified fitting method proposed by Lenhard and Parker (1987) (Eq. 4) which was a combination of the Leverett equation (1941) and a four-parametric fitting procedure suggested by van Genuchten et al. (1980).

$$S = \left[\frac{1}{1 + (\beta * \alpha * P)^n}\right]^m \qquad (4)$$

where: S is the fluid saturation; P is the pressure head; β is a scaling factor calculated from the interfacial tension of the phase pair and α, n and m are fitting parameters.

In scaling method the m = 1-1/n constraints was supposed as suggested by van Genuchten (1991).

In order to compare the accuracy and reliability of the different predictions, the RMSE and R^2 values were determined. The NAPL saturation values of soil samples were also calculated on the validation test in order to make a preliminary investigation to compare the reliability of PTFs with scaling method in predicting NAPL saturation values, versus the pressure.

Table 4 Statistics of the fitted fluid retention values in calibration and validation datasets

Pressure (mbar)	Calibration set						Validation set					
	N = 264						N = 26					
	Mean	SE	SD	Med.	Min.	Max.	Mean	SE	SD	Med.	Min.	Max.
NAPL retention												
0.0	41.0	1.0	10.0	41.0	22.0	68.0	38.0	2.0	10.0	37.0	25.0	57.0
0.33	41.0	1.0	10.0	41.0	22.0	68.0	38.0	2.0	10.0	36.0	25.0	57.0
10	36.0	1.0	10.0	35.0	17.0	64.0	33.0	2.0	8.0	32.0	21.0	49.0
33	31.0	1.0	8.0	31.0	9.0	62.0	28.0	1.0	5.0	29.0	20.0	38.0
100	25.0	0.0	6.0	25.0	5.0	60.0	24.0	1.0	5.0	24.0	16.0	33.0
200	22.0	0.0	6.0	21.0	3.0	55.0	22.0	1.0	6.0	22.0	10.0	32.0
330	20.0	0.0	6.0	19.0	2.0	49.0	21.0	1.0	6.0	22.0	7.0	32.0
1000	17.0	0.0	6.0	16.0	1.0	34.0	18.0	1.0	7.0	18.0	3.0	30.0
15849	11.0	0.0	7.0	11.0	0.0	29.0	14.0	1.0	7.0	14.0	0.0	26.0
1584893	7.0	0.0	6.0	6.0	0.0	22.0	10.0	1.0	6.0	9.0	0.0	21.0
Water retention												
0.0	50.0	1.0	11.0	47.0	27.0	81.0	49.0	2.0	9.0	47.0	37.0	70.0
0.33	50.0	1.0	11.0	47.0	26.0	81.0	49.0	2.0	9.0	47.0	37.0	70.0
10	48.0	1.0	11.0	46.0	16.0	78.0	47.0	2.0	9.0	46.0	36.0	69.0
33	45.0	1.0	11.0	43.0	12.0	73.0	45.0	2.0	10.0	42.0	31.0	67.0
100	41.0	1.0	12.0	39.0	5.0	73.0	42.0	2.0	10.0	40.0	26.0	63.0
200	37.0	1.0	12.0	36.0	1.0	72.0	39.0	2.0	10.0	39.0	23.0	62.0
330	35.0	1.0	12.0	33.0	0.0	70.0	37.0	2.0	10.0	38.0	21.0	61.0
1000	29.0	1.0	11.0	27.0	0.0	65.0	32.0	2.0	10.0	32.0	17.0	56.0
15849	17.0	0.0	8.0	16.0	0.0	41.0	19.0	1.0	7.0	19.0	7.0	37.0
1584893	7.0	0.0	4.0	6.0	0.0	19.0	8.0	1.0	3.0	8.0	1.0	15.0

RESULTS AND DISCUSSION

Fitting van Genuchten equation

As the determination coefficient showed NAPL retention curves can be accurately fitted with the hydraulic function proposed by van Genuchten (1980). Moreover, according to the ZAPF values, the differences between the fitted and measured NAPL retention values were lower than the measurement error in more than 98.4% of the samples (*Table 5*).

Table 5 Statistics of the fitted van Genucthen equation to the measured NAPL retention values

Equation	ZAPF (<0,75 – ME*)		RMSE	R	R^2
	N	Frequency (%)			
van Genuchten (3 parameter)	2286.0	98.4	1.2	0.99	0.99

* ME is the average error of NAPL retention measurement

Principal component analysis

Principal components were extracted from 17 variables which showed strong correlations between the selected soil properties. The maximum possible variance of 90.92% can be explained by the three components.

Scores of component I (CI) give information about 69.56% scores of component II (CII) about 15.48% and scores of component III (CIII) about 5.87% of the variance of variables.

After Kaiser normalization, all components with eigenvalues under 1.0 were eliminated, and only three components were taken into account. In *Table 6* and *7*, component scores less than the variance explained criteria (0.4) are shown in gray. Communality values show, that all the variances of each correlated variables might be accounted for by the components. The finer textured fractions belonged to the same component with the NAPL retention at higher pressure levels, where the organic liquid can be retained against gravity mainly by capillary and adsorptive forces. The joint variation of these soil properties were affected by CI as the unobserved latent variable. Moreover, the component scores of finer particle size fractions exceeded the 0.4 criteria only in CI.

Differences in the role of components on the NAPL retention of soils were observed. Increasing value of component scores with increasing pressure can refer to the strong correlations between particle size distribution and fluid retention, especially in the higher pressure range.

Higher component scores of CII were observed than CI on the 0-50mbar pressure range. In CII only the fluid retention less than 150 mbar were represented while the component scores of CI in case of NAPL retention at

higher than 400 mbar exceed the 0.9 value (*Table 6.*) This difference might be resulted from reaching a threshold pressure range, which can be in the range of 20-50 mbar. At this pressure range the larger pores might be drained by gravity during NAPL retention measurements with Dunasol 180/220, as similar results were observed by Makó et al. (2011a). This is lower than the pressure of the field capacity of soils, which is referred approximately 400 mbar (pF 2.5) in the case of water retention.

This is might not be contradicted to the results of other researchers for water retention (Vereecken et al., 1989; Saxton and Rawls, 2006). Vereecken et al., (1989) resulted that at pressure higher than 1500 kPa water retention is determined mainly by texture then in lower pressure range the fluid retention strongly influenced by the aggregation of soil particles and organic matter content as well.

However, in case of NAPL retention the variance of organic matter content affects the NAPL retention but in decreasing rate with increasing pressure in CI and CII, and calcium carbonate content of samples appeared only in CIII.

Variance in bulk densities can be explained by all three components, thus it may influence the NAPL retention at every pressure level. In all cases bulk density had negative correlation with other variables which refers inverse relationships between bulk density and fluid content of soils samples.

More complex and less correlation between parameters of water and NAPL retention curves related to basic soil properties were observed. In case of water retention only the 55.59% of the variance could be explained by the components as compared to NAPL retention (76.55% (*Table 7*).

Parameters α, n and θ_s have substantial roles in CI of NAPL retention in contrast to the parameters of water retention curves which belong to separated principal components. Descriptive statistics show (*Table 4*), that the mean saturated water content is significantly higher (50 vol %) than the possible average maximum NAPL retention (41 vol %), which might influence the variability of the other two fitting parameters.

Parameters α and n are shape factors, strongly influenced by the pore size distribution of soils and structure (van Genuchten et al., 1991; Vereecken et al., 1989). However, there is a difference in effective pore size distribution of soils saturated by fluids with various chemical and physical properties. In addition, interactions between water and the solid phase during drainage and imbibition processes, such as swelling and shrinking, have not occurred in case of the soil pores filled with NAPLs. Thus, the soil properties which might influence the shape of fluid retention curves might be difference.

Strong correlation between the fitting parameters was experienced when the van Genuchten equation was fitted to NAPL retention data. The correlation between α and θ_s of the selected four texture group varied between 0.553-0.887. This might refers the inability of using the same initial parameters or boundary conditions for fitting NAPL and water retention curves.

Lower communality values were experienced in the n and θ_s values of water retention curves than those of NAPL retention curves. This could be resulted from the variability in NAPL retention of soils might be explained better with the selected basic soil properties (communalities of n was 0.821, and θ_s was 0.739) than in case of water retention (where the cumulated variance was only 0.282 and 0.298).

Table 6 Components of variance and scores of the principal component analysis

	Variables	CI	CII	CIII	Communalities
Basic soil properties	0.0002 mm >	0.93	0.22	-0.19	0.94
	0.005-0.002 mm	0.87	0.36	0.11	0.89
	0.01-0.005 mm	0.91	0.30	0.18	0.94
	0.02-0.01 mm	0.75	0.35	0.50	0.93
	0.05-0.02 mm	0.02	0.41	0.88	0.95
	0.05-0.25 mm	-0.94	0.05	-0.33	0.99
	Organic matter (w %)	0.56	0.46	0.02	0.53
	CaCO$_3$ (w %)	-0.07	0.05	0.95	0.91
	Bulk density (g cm^{-3})	-0.51	-0.53	-0.53	0.82
NAPL retention	0mbar (w %)	0.48	0.66	0.46	0.88
	2mbar (w%)	0.49	0.65	0.45	0.87
	20mbar (w %)	0.46	0.76	0.39	0.94
	50mbar (w %)	0.56	0.59	0.54	0.96
	150mbar (w %)	0.78	0.46	0.40	0.98
	400mbar (w %)	0.91	0.34	0.14	0.96
	1000mbar (w %)	0.93	0.34	0.04	0.98
	1500mbar (w %)	0.90	0.41	-0.05	0.98
Statistics	Variance %	69.56	15.48	5.87	

Table 7 Components of variance and scores of the principal component analysis with parameters of fluid retention curves

Variables	Water retention				NAPL retention			
4	CI	CII	CIII	Communalities	CI	CII	CIII	Communalities
a	0.025	-0.072	0.753	0.665	-0.738	0.451	-0.041	0.749
n	-0.155	0.581	-0.101	0.282	0.468	-0.775	0.021	0.821
θ_s	-0.602	0.171	-0.228	0.298	0.841	0.155	0.085	0.739
Clay content (w %)	0.847	-0.163	-0.237	0.933	0.409	0.657	-0.455	0.806
Silt content (w %)	0.469	0.478	0.077	0.932	0.097	0.815	0.337	0.788
Organic matter (w %)	0.014	0.123	0.784	0.579	-0.187	0.074	0.643	0.454
$CaCO_3$ (w %)	0.084	0.801	0.137	0.722	0.426	0.03	0.726	0.709
Bulk density (g cm^{-3})	-0.638	-0.315	-0.034	0.686	-0.889	0.016	0.136	0.808
Variance %	22.768	17.792	22.768		33.191	29.179	14.184	
Cumulated variance %	22.768	40.560	55.599		33.191	62.370	76.554	
Bartlett's Test of Sphericity				0.000				0.000
Kaiser-Meyer-Olkin Measure of Sampling Adequacy				0.597				0.634

In both cases bulk density is presented in CI with negative effect on saturated fluid content, which can be a consequence the bulk density and saturated fluid content are closely related.

In case of water retention the role of organic matter content is apparent only in CIII separately from clay content. This can be a consequence of the effect of high clay content might mask the effects of increasing organic matter for water retention, as Saxton and Rawls (2006) and Rawls et al. (2003) experienced as well.

Similar relationships observed between organic matter content and the fitting parameters of water retention curves than Petersen et al. (1968) and Vereeckeen et al. (1989) that is the organic matter content of soils affects primary the α parameter of water retention curves, which refers the fluid retention values at inflection point. In contrary with texture, organic matter content of soils influence the position of the retention curves, rather than its shape.

Communality values showed the same important role of calcium carbonate (0.722, 0.709) and organic matter (0.579, 0.454) on water and NAPL retention. However, a separated component (CIII) was composed by organic matter, $CaCO_3$ and clay content in case of NAPL retention data. In addition the role of clay content on water retention seemed to be higher than its role on NAPL retention. All of these observed relationships refer to the different rate of interactions between various fluids and the solid phase considering these express mainly the linear relationships between these variables.

In spite of the cumulated variance was only 55.59 % in case of water retention, the Kaiser-Meyer-Olkin test showed the applicability of PCA, sampling adequacy (as it refers that the correlations between pairs of variables can be explained with other variables) was higher than 0.5, and the Barthlett's test indicated significant correlation between variables (p<0.05).

Pedotransfer functions

Accurate prediction was given by classPTFs for the selected four texture groups for estimating NAPL retention of soils (*Table 8*).

Table 8 Accuracy of PTFs to estimate fitting parameters of NAPL retention curves

Parameter	R^2	RMSE
α	0.978	13.59
n	0.900	17.75
θ_s	0.988	7.50

According to the RMSE values classPTFs give more accurate estimations to saturated fluid content than the shape parameters α or n. The lower efficiency of estimating parameter α might be caused by the strong correlations experienced between parameters α and θ_s and might also resulted from the weaker correlation with independent variables.
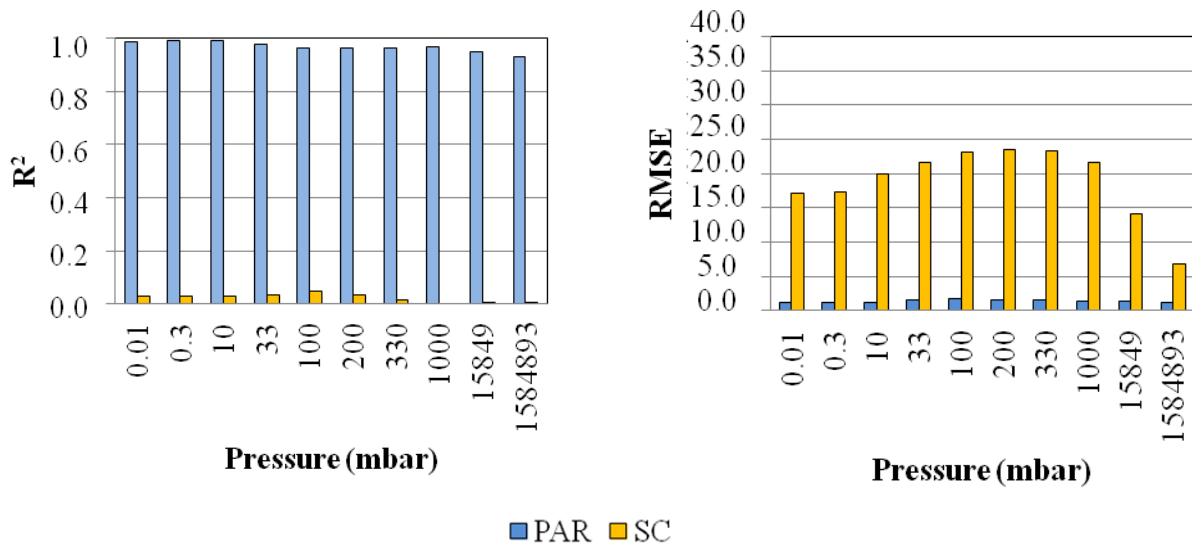
Fig. 3 Accuracy of estimating volumetric NAPL retention values with different estimation methods
PAR: the parametric method based on basic soil properties and their derived values;
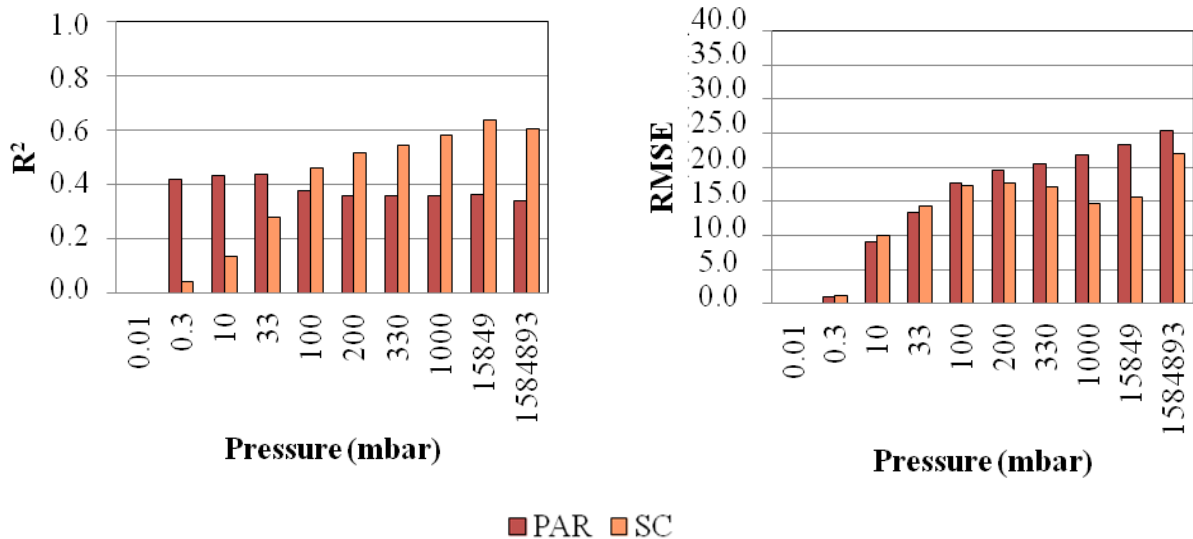SC: Scaling method (Lenhard and Parker, 1987)

Fig. 4 Reliability of estimating NAPL saturation values with PTF and scaling methods
PAR: the parametric method based on basic soil properties and their derived values;
SC: Scaling method (Lenhard and Parker, 1987)

The applicability of PTFs to predict the NAPL retention values was proven by the statistical attributes (R^2 and RMSE). In comparison the accuracy of parametric PTFs with the scaling method, NAPL retention might be better predicted with PTFs at a given pressure level (*Fig. 3*).

Applicability of these estimation methods for predicting NAPL saturation values of soils was compared with the determination of R^2 and RMSE values on validation data (10 % of the dataset, 26 samples). Lower reliability of PTFs was found to estimate the NAPL saturation with PTFs than the scaling method at higher than 100 mbar pressure level. However, PTFs could provide estimates sufficiently reliable for NAPL retention at all pressure level, especially near saturation (*Fig. 4*). On *Fig. 4* the statistics of saturated fluid content were not presented because their values are 1 for both fluids.

Supposedly, scaling method gave increasing better estimation for NAPL retention as a consequence of decreasing difference between fluid retentions with NAPL and water. As it shown formerly from descriptive statistics (*Table 4*), the difference between the average water and NAPL retention of the investigated soil samples was 26 vol. % at 100 mbar but practically 0 vol. % at 1584893 mbar.

At the same time RMSE values indicated lower difference in reliability of estimation methods for predicting the NAPL saturation.

Only moderate reliability of PTFs was found when predicting the volumetric NAPL retention values (*Fig. 5*). In spite of the fact that higher RMSE values were observed for PTFs in lower pressure ranges, the R^2 values show that the reliability of PTFs exceed that of the scaling method at lower pressure levels (< 330 mbar) and is al-
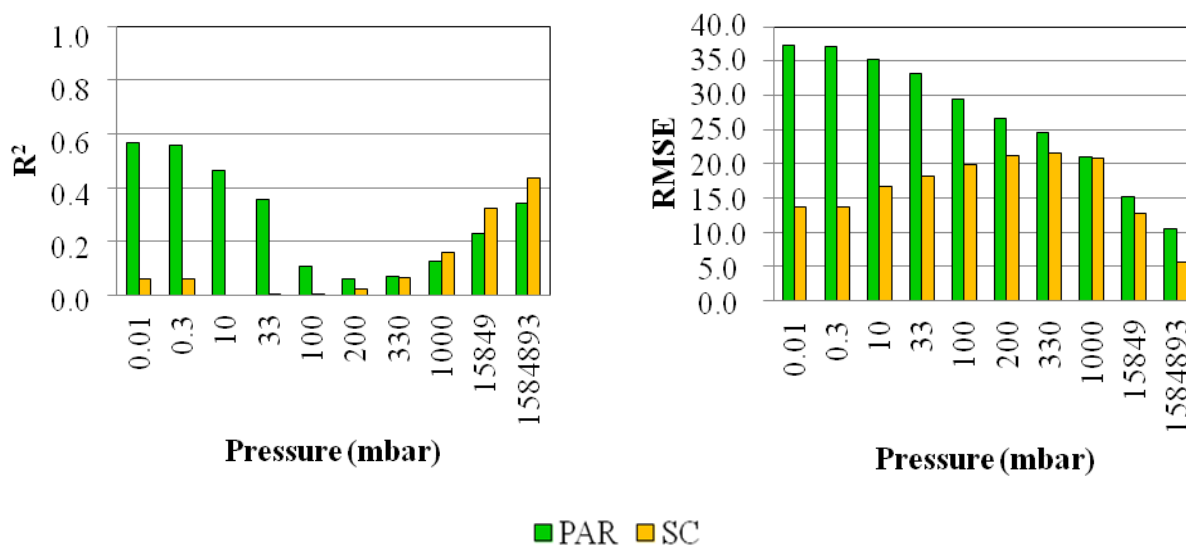
Fig. 5 Reliability of estimating volumetric NAPL retention values with PTF and scaling methods
PAR: the parametric method based on basic soil properties and their derived values;
SC: Scaling method (Lenhard and Parker, 1987)

most equal to it at pressure higher than 1000 mbar (*Fig. 5*). Moreover both the decreasing R^2 values of the scaling method and the increasing RMSE values in the mid-pressure range (near the inflection point of the NAPL retention curves) suggest the inability of scaling methods for predicting NAPL retention of soil based on water retention.

CONCLUSION

In the last fifteen years several investigations were performed in order to create estimation methods to predict the NAPL retention of soils. However, the results of these measurements have not been collected into comprehensive and well established databases, yet.

Building PTFs using the data of easily measured soil property data might be a promising alternative for predicting the NAPL retention of soils. By the creation of an organic liquid retention database representative for most soil types we can generate well established point and parametric PTFs based on basic soil properties. After the validation of PTFs with the results of column experiments it might be inserted to transport models. Moreover, databases of soil hydraulic and basic properties which have already been used in developing PTFs for estimating water retention might be applicable to create NAPL pollution sensitivity maps.

In our study an organic liquid retention dataset from recently measured NAPL and water retention data of soil and mineral mixture samples was created. The fluid retention measurements were carried out with the pressure plate extractor method. Furthermore, the basic soil properties, such as particle size distribution, organic matter and $CaCO_3$ contents, and bulk density, etc. were also measured.

According to our results, the van Genuchten equation, having three parameters, could be perfectly fitted to the measured NAPL retention data as well as to water retention curves.

Preliminary analysis of the filtered dataset of 316 soil samples with PCA showed strong correlations between the selected basic soil properties as well as fluid retentions of water and NAPL. The important role of organic matter and $CaCO_3$ content in fluid retention of soils beyond the particle size distribution and bulk density was proved.

The estimation methods for predicting the volumetric NAPL retention values of soils PTFs gave better results than the scaling method. As the parameters of NAPL retention curves were correlated (e. g. fitting with nonlinear regression might offer more than one possible solution), further research is needed to investigate the role of initial parameter values in fitting hydraulic functions to NAPL retention data.

Reliability of PTFs and the scaling method to predict the saturation values versus the pressure were compared. Similar reliable estimations found when predicting the NAPL saturation values with scaling method and with PTFs. However, PTFs has been successfully applied to estimate the volumetric NAPL retention, more reliable than the scaling method.

References

Acutis, M., Donatelli, M. 2003. SOILPAR 2.00: Software to estimate soil hydrological parameters and functions. *European Journal of Agronomy* 8 (3–4), 373–377. DOI: 10.1016/s1161-0301(02)00128-4

Beckett, G.D., Joy, S. 2003. Light Non-Aqueous Phase Liquid (LNAPL) Parameters Database. Version 2.0. Users Guide. American Petroleum Institute. Publ. 4731. Washington, DC.

Brooks, R. H., Corey, A. T., 1964. Hydraulic properties of porous media. Hydrology Paper No. 3. Colorado State University. Fort Collins, Colorado.

Brunauer, S., Emmett, P. H., Teller, E. 1938. Adsorption of gases in multimolecular layers. *Journal of the American Chemical Society 60 (2) 309–319.* DOI: 10.1021/ja01269a023

Brutsaert, W. 1966. Probably laws for pore size distributions. *Soil Science* 101 85–92. DOI: 10.1097/00010694-196602000-00002

Buzás, I. 1993. Manual for soil and agrochemical analysis. INDA Kiadó, Budapest. (in Hungarian)

Carsel, R.F., Parrish, R.S. 1988. Developing joint probability distribution of soil water retention characteristics. *Water Resources Research* 24 755–769. DOI: 10.1029/wr024i005p00755

Chen, J., Hopmanns, J. W., Grismer, M. E. 1999. Parameter estimation of two-fluid capillary pressure-saturation and permeability functions. *Advances in Water Resources* 22, 479–493. DOI: 10.1016/s0309-1708(98)00025-6

DePastrovitch, T. L., Baradat, Y., Barthel, Y., Chiarelli, A., Fussel, D. R., 1979. Protection of groundwater from oil pollution. CONCAWE Rep. No 3/79. CONCAWE (Conservation of Clean Air and Water – Europe). The Hague. The Netherlands.

Fodor, N., Rajkai, K. 2011. Computer program (SOILarium 1.0) for estimating the physical and hydrophysical properties of soils from other soil characteristics. *Agrokémia és Talajtan* 60, 27–40.

Hernádi, H., Makó, A. 2010. Predicting oil retention of soils polluted with hydrocarbon derivates with pedotransfer functions. Mérnökgeológia-kőzetmechanika. Műegyetem Kiadó. Budapest. (in Hungarian)

Hernádi, H., Makó, A. 2011a. Predicting the oil retention of soils with different methods. In Farsang, A., Ladányi, Zs. (eds.) *Talajvédelem* supplementum (3-4), 363–371 (In Hungarian)

Hernádi, H., Makó, A., Kovács, J., Csatári, T. 2011b. Nonaqueous-phase liquid retention of mineral mixture series containing different clay minerals. *Communications in Soil Science and Plant Analysis* 44, 390–396. DOI: 10.1080/00103624.2013.742314

Lamorsky, K., Pachepsky, Y., Słaviński, C., Walczak, R.T. 2008. Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. *Soil Science America Journal* 72, 1243–1247. DOI: 10.2136/sssaj2007.0280n

Lenhard, R .J., Parker, J. C., 1987. Measurement and prediction of saturation-pressure relations in three phase porous media systems. *Journal of Contaminant Hydrology* 1. 407–424. DOI: 10.1016/0169-7722(87)90017-9

Leverett, M. C. 1941. Capillary behavior in porous solids. *Transactions of the Society of Petroleum Engineers*. *American Institute of Mechanical Engineers* 142, 152–169.

Lilly, A. 2010. A hydrological classification of UK soils based on soil morphological data. 19th World Congress of Soil Science, Soil Solutions for a Changing World 1 – 6 August 2010, Brisbane, Australia. Published on DVD.

Makó, A. 1995. Interactions between the porous phase of soils and organic liquids. PhD dissertation. Keszthely, Hungary. (in Hungarian)

Makó, A. 2002. Measuring and estimating the pressure-saturation curves on undisturbed soil samples using water and NAPL. *Agrokémia és Talajtan* 51, 27–36.

Makó, A. 2005. Measuring the two-phase capillary pressure-saturation curves of soil samples saturated with nonpolar liquids. *Communication in Soil Science and Plant Analysis* 36, 439–453. DOI: 10.1081/css-200043170

Makó A., Elek, B. 2006. Comparison of soil extraction isotherms of soil samples saturated with nonpolar liquids. *Water, Air and Soil Pollution* 6, 331–342. DOI: 10.1007/s11267-005-9026-x

Makó, A., Hernádi H. 2013. Hydrocarbon derivates in soils: Soil physical researches. Monography. (in Hungarian)

Makó, A., Marczali, Zs. 1999b. Laboratory measurement of the soils fluid retention concerning the organic liquid retention. XIII. Országos Környezetvédelmi Konferencia és Szakkiállítás. Siófok. 14-16. szept. 1999. 147-153. (In Hungarian)

Makó, A., Máté, F., Németh, T., Hernádi H. 2004. The temperature-dependence of the NAPL retention of different soils. 12th International Poster Day and Institute of Hydrology Open Day: Transport of water, chemicals and energy in the soil-plant-

atmosphere system. 25th November 2004. Bratislava. Slovak Republik.

Makó, A., Rajkai, K., Tóth, G., Hermann, T. 2005. Estimating soil water retention characteristics from the soil taxonomic classification and mapping informations: consideration of humus categories. *Cereal Research Communications* 34, 199–201. DOI: 10.1556/crc.33.2005.1.27

McBratney, A. B., Minasny B., Cattle S. R., Vervoort R. W. 2002. From pedotransfer functions to soil inference systems. *Geoderma* 109, 41–73. DOI: 10.1016/s0016-7061(02)00139-8

Minasny, B., McBratney, A. B., Bristow, K. L. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma* 93, 225–253. DOI: 10.1016/s0016-7061(99)00061-0

Minasny, B., J., Hopmans, J. W., Harter, T., Eching, S. O., Tuli, A., Denton, M. A. 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Science Society of America Journal* 68, 417–429. DOI: 10.2136/sssaj2004.4170

Nemes, A., Schaap, M. G., Wösten, J. H. M. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. *Soil Science Society of America Journal* 67, 1093–1102. DOI: 10.2136/sssaj2003.1093

Nemes, A., Roberts, R. T., Rawls, W. J., Pachepsky, Y. A., van Genuchten, M. T. 2008. Software to estimate -33 and -1500 kPa soil water retention using the non-parametric k-nearest neighbor technique. *Environmental Modelling & Software* 23 (2), 254–255. DOI: 10.1016/j.envsoft.2007.05.018

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

Parker, J.C., Lenhard, R.J., Kuppusamy,T. 1987. A parametric model for constitutive properties governing multi-phase flow in porous media. *Water Resources Research* 23, 618–624. DOI: 10.1029/wr023i004p00618

Petersen, G. W., Cunningham, R. L., Matelski, R.P. 1968. Moisture characteristics of Pennsylvania soils II. Soil factors affecting moisture retention within a textural class - silt loam. *Soil Science Society America Proceedings* 32, 866–870. DOI: 10.2136/sssaj1968.03615995003200060042x

Rajkai, K. 1988. Relationships between water retention and different soil properties. *Agrokémia és Talajtan* 36–37, 15–30. (In Hungarian)

Rajkai K. 2004. The quantity, distribution and transport of water in soil. RISSAC, Research Institute for Soil Sciences and Agricultural Chemistry. Budapest. (In Hungarian)

Rajkai, K., Kabos, S., van Genuchten, M. T., Jansson, P. E. 1996. Estimation of water-retention characteristics from the bulk density and particle-size distribution of Swedish soils. *Soil Science* 161, 832–845. DOI: 10.1097/00010694-199612000-00003

Rajkai, K., Kabos, S., van Genuchten, M. T., 2004. Estimating the water retention curve from soil properties: Comparison of linear, nonlinear and concomitant variable methods. Soil and Tillage Research 79. (2) 145–152. DOI: 10.1016/j.still.2004.07.003

Rathfelder, K., Abriola, L. M. 1996. The influence of capillarity in numerical modelling of organic liquid redistribution in two-phase systems. *Advances in Water Resources* 21 (2), 159–170. DOI: 10.1016/s0309-1708(96)00039-5

Rawls, W. J., Brakensiek, D. L., 1985. Prediction of soil water properties for hydrologic modeling. In: Watershed Management in the 1980s. Proceeding of Symposium of Irrig. Drainage Div., Denver, CO., April 30–May 1, 1985. American Society pf Civil Engineers. New York. 293-299.

Rawls, W. J., Pachepsky, Y. A., Shen, M. H., 2001. Testing soil water retention estimation with the MUUF pedotransfer model using data from the southern United States. *Journal of Hydrology* 251. 177–185. DOI: 10.1016/s0022-1694(01)00467-x

Rawls, W.J., Pachepsky, Y. A., Ritchie, J. C., Sobecki, T. M., Bloodworth, H. 2003. Effect of Soil Organic Carbon on Soil Water Retention. Geoderma 1 (16), 61–71. DOI: 10.1016/s0016-7061(03)00094-6

Saxton, K. E., Rawls, W. J. 2006. Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society America Journal* 70, 1569–1578. DOI: 10.1016/s0016-7061(03)00094-6

Schaap, M. G., Leij, F. J., van Genuchten, M. T. 1999. A bootstrap-neural network approach to predict soil hydraulic parameters. In:

Proceedings of International Workshop Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media 1237–1250.

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. DOI: 10.1016/s0022-1694(01)00466-8

Tietje, O., Tapkenhinrichs, M. 1993. Evaluation of pedo-transfer functions. *Soil Science Society of America Journal* 57, 1088–1095. DOI: 10.2136/sssaj1993.03615995005700040035x

Tóth, B. 2011. Calculation and characterization of water retention of major Hungarian soil types using soil survey information. PhD dissertation. Keszthely. (in Hungarian)

Tóth, B., Makó, A., Rajkai, K., Kele, Sz. G., Hermann, T., Marth, P. 2006. Use of soil water retention capacity and hydraulic conductivity estimation in the preparation of soil water management maps. *Agrokémia és Talajtan* 55, 49–58. DOI: 10.1556/agrokem.55.2006.1.6

Tóth, B., Makó, A., Tóth, G., Farkas, C., Rajkai, K. 2013. Comparison of pedotransfer functions to estimate the van Genuchten parameters from soil survey information. Agrokémia és Talajtan 62 (1), 5-22. (in Hungarian)

van Genuchten M.T. 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society American Journal* 44, 892–989. DOI: 10.2136/sssaj1980.03615995004400050002x

Vereecken, H., Maes, J., Feyen, J., Darius, P. 1989. Estimaing the soil moisture retention characteristic from texture, bulk density and carbon content. *Soil Science* 148, 389–403. DOI: 10.1097/00010694-198912000-00001

Wösten, J. H. M., Schuren, C. H. J. E. Bouma, J., Stein, A. 1990. Functional sensivity analysis of four methods to generate soil hydraulic functions. *Soil Science Society of America Journal* 54, 832–836. DOI: 10.2136/sssaj1990.03615995005400030036x

Wösten, J.H.M., Finke, P.A., Janses, M.J.W. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma* 66, 227-237. DOI: 10.1016/0016-7061(94)00079-p

Wösten, J. H. M., Lilly, A., Nemes, A., Le Bas, C. 1999. Development and use of a database of hydraulic properties of European soils. *Geoderma* 90, 169–185. DOI: 10.1016/s0016-7061(98)00132-3

Weawer, J. L., Charbeneau, R. J., Tauxe, J. D., Lien, B. K., Provost, J. B. 1994. The hydrocarbon spill screening model (HSSM). 1. US EPA. EPA/600/R-94/039a.