Preliminary investigation to estimate soil NAPL retention using parametric pedotransfer functions**

Hilda Hernádi[1] and András Makó[2]*

[1]University of Pannonia, Georgikon Faculty, 16. Deák, Keszthely H-8360, Hungary
[2]Institute for Soil Sciences and Agricultural Chemistry, Centre for Agricultural Research, Hungarian Academy of Sciences,
15. Herman, Budapest H-1022, Hungary

A b s t r a c t. Organic liquid retention of soils is a primary input variable for modelling the nonaqueous phase liquid transport and behaviour in the subsurface. In environmental and soil physical practice, it is mainly determined by scaling based on the water retention of soils or with charts of average empirical values of organic liquid retention or the fitting parameters of hydraulic functions. Predicting the fitting parameters of organic liquid retention curves with pedotransfer functions might be a promising alternative method, but this topic has only been researched to a limited extent. Thus we investigated the applicability of different hydraulic functions (3- and 4- parameter form of the van Genuchten equation and Brutsaert equation) for fitting organic liquid retention characteristics. Multivariate linear regression was used to build and develop pedotransfer functions, modelling relations between original and transformed values of basic soil properties and organic liquid retention. We attempted to generate parametric pedotransfer functions. According to our results, the applicability of hydraulic functions for fitting nonaqueous phase liquid retention curves to the experimental data was proven. The investigations gave promising results for the possibility to estimate soil nonaqueous phase liquid retention with parametric pedotransfer functions.

K e y w o r d s: pedotransfer function, hydraulic characteristics, van Genuchten equations

INTRODUCTION

Pressure – saturation (P-S) relationships are primary input soil properties for simulating the behaviour and transport of NAPLs (Nonaqueous phase liquids) in the subsurface. Although knowledge of the soil NAPL retention is very important, it is commonly estimated with simplified methods, based on the soil water retention. In environmental practice, P-S is usually determined by charts of average empirical pressure – fluid saturation values or different modified versions of the Leverett function (Leverett, 1941) (Eq. (1)), scaling water retention to NAPL retention using some soil properties (*eg* viscosity, interfacial tension, density) and their combinations (*eg* in the HSSM model – Weaver *et al.,* 1994 or PLSI Simulator – Guarnacchia *et al.,* 1997):

$$\theta_{nw} = \theta_w \frac{\sigma_w \ \rho_{nw}}{\sigma_{nw} \ \rho_w},\qquad(1)$$

where: θ is the fluid retention of soil samples; σ is the interfacial tension of fluids and ρ is the density of the nonwetting (*nw*) and wetting (*w*) phases.

Moreover, Lenhard and Parker (1987) suggested an improved scaling method (Eq. (2)), which was used by Beckett and Joy (2003) to create a dataset of calculated fitting parameters of the van Genuchten equation for NAPL retention curves:

$$S = \left[\frac{1}{1+\left(\beta \ \alpha \ P\right)^n} \right]^m ,\qquad(2)$$

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

Results of recent investigations have demonstrated that the applicability of methods based on determining the NAPL retention from water retention is questionable for natural (well aggregated, higher organic matter or clay contents, *etc*.) soils. Charts of average fluid retention or fitting parameter values might not represent the variability of soil properties that presumably influence the organic liquid

retention of soils. In Leverett type equations, the different interactions between the various fluids and the porous media were assumed to be negligible. Moreover, only a few column experiments had been carried out to validate these estimation methods and these experiments were primarily carried out on glass beads and/or sands (Makó and Hernádi, 2013).

Creating PTFs based on easily measurable soil properties might be a promising alternative method for the prediction of the NAPL retention of soils (Makó, 2005; Makó and Hernádi, 2013).

PTFs have already been used successfully to determine the water retention of soils by several authors (Lamorski *et al.*, 2012; Minasny *et al.*, 1999; Rajkai, *et al.*, 2004; Schaap and Leij, 1998).

The most important topics in building and developing PTFs to estimate the water retention of soils are the determination of the role of additional soil properties on water retention (*eg* more detailed information about PSD, soil structural information, and morphological properties) (Vereeckeen *et al.*, 2010), and the development of new data handling techniques, such as artificial neural networks, regression trees, or inference systems (Botula *et al.*, 2013; Minasny *et al.*, 2004). Even though the promising new statistical techniques might offer higher prediction ability, MLR has the advantage of simple operation with sufficient accuracy to estimate soil water retention characteristic (SWRC) (Minasny *et al.*, 1999; Vereecken *et al.*, 1989).

Notwithstanding the promising results of the applicability of PTFs to estimate water retention of soils, only few researchers have begun to create special PTFs to estimate the NAPL retention of soils (Makó, 2005; Makó and Hernádi, 2013). At the same time, experience in creating PTFs for water retention could be used as the base of generating PTFs to predict the NAPL retentions of soils.

In predicting water retention, basically three main types of PTFs might be distinguished. Point estimation is based on the determination of water retention at a given pressure level. Parametric methods are developed to estimate parameters of the soil water retention characteristic (SWRC). Only few researchers combined the parametric method with the direct fitting of the parametric function to the estimated water retention values (Tomasella *et al.*, 2000). In physico-empirical models, the SWRC was determined using fractal geometry models or by translating the particle size distribution (PSD) into SWRC based on the capillary equation (Haverkamp *et al.*, 2005).

Many authors have found that point PTFs might give more accurate prediction (Minasny *et al.*, 1999; Tomasella *et al.*, 2000) than parametric methods, although parametric methods offer the possibility to estimate soil water retention for a wide range of pressure heads if limited retention data are available. However, the correct application of parametric PTFs requires that the model is correctly specified and the fitting of SWRC and predictions are accurate.

Fitting parameters of hydraulic functions are commonly used to describe unsaturated soil parameters, such as permeability or hydraulic conductivity (van Genuchten, 1980; Lenhard and Parker, 1987). In addition, most of the fluid retention measurement methods are only suitable to determine the main drainage curves, which would be the ground to determine the hysteretic and scanning curves for drainage and imbibition. Thus, the inaccuracy of fitting might cause multiplied errors of the simulated transport parameters in hysteretic multiphase transport models (*eg* the amount of residual organic liquid content, infiltration depth and time) (Guarnaccia *et al.*, 1997).

Unimodal mathematical expressions of van Genuchten equations having 5, 4, or 3 parameters (1980) and the Brutsaert equation (1966) are widely cited and adopted for modelling purposes because of their relative simplicity. The fundamental requirement for hydraulic functions is that they are monotonic and they must reach zero fluid content at finite suction.

Originally, the van Genuchten equation has five parameters. Despite the shape parameters without physical meaning giving flexibility to the model, van Genucthten *et al.* (1991) suggested the use of its four-parameter form in case only a limited range of data (usually in the wet range) is available. Recently, the two most often used constraints for parametric m is $m = 1$ (Brutsaert, 1966), which leads to a symmetrical SWRC plot, and the constraint $m = 1-1/n$ (van Genuchten *et al.*, 1991), which might lead to an asymmetrical SWRC plot. In addition, constraining $m = 1-1/n$ in van Genuchten equation having four parameters might lead to increased prediction potential without losing much of the accuracy and linearity of fitting (Assouline and Or, 2013; Cornelis *et al.*, 2005).

These closed form analytical expressions might give fairly accurate fitting also to determine the organic liquid retention curves (Makó, 2005; Rathfelder and Abriola, 1996). However, no results of investigations of the accuracy of various type hydraulic functions for parametrizing NAPL retention data have been published until now.

In the 1990s, a series of investigations to build PTFs for estimating NAPL retention of soils were initiated in Hungary (Makó, 2005). Now, the basic soil parameters and data of water and NAPL retention measurements with the pressure plate extractor method were collected into a dataset. The aim of this study was to examine the applicability of three different types of hydraulic functions (the van Genuchten equations with three and four parameters and the Brutsaert function) for fitting NAPL retention curves to experimental data with the nonlinear regression method.

The possibility of estimation of the fitting parameters with parametric PTFs was also investigated. The accuracy of parametric PTFs was compared to the accuracy of commonly used scaling methods (Leverett, 1941; Lenhard and Parker, 1987).

MATERIALS AND METHODS

The collected dataset consisted of 369 samples. Particle size distribution of the soil and mineral mixture samples were determined by the conventional pipette method (dispersing agent: $Na_4P_2O_7$) (Gee and Bauder, 1986). Organic matter content measurements were carried out by wet combustion with the Tyurin titrimetric method (oxidation with chromic acid using boiling and $Cr_2O_7^{-2}$ not used in oxidation is titrated with $Fe(NH_4)_2 (SO_4)_2$) (Tyurin, 1931). The calcium carbonate content was measured with a Scheibler calcimeter using 1-10 g samples, adding HCl to the samples and calculating the $CaCO_3$ from the volume of released CO_2 (Nelson, 1982). Bulk density was determined by oven drying 100 cm^3 intact soil cores at 105°C (Blake and Hartge, 1986).

The water retention measurements were performed with distilled water according to the MSZ 08 0205:1978 and ISO 11274:2009 standards. NAPL retention was determined with a non-aromatic nonpolar organic liquid, DUNASOL 180/220 (Hungarian Gas and Oil Company Plc. - MOL Rt.) (Makó and Hernádi, 2013). For the organic liquid retention measurements, two types of porous ceramic plates were used, characterized by 100 and 500 kPa air entry value (for the water-air system). Porous plate cells were the modified version of LAB023 cells (Soilmoisture Equip. Corp.). The original accessories were replaced by Teflon and glass, while porous ceramic plates were completed in pressure cells with oil-resistant silicon rubber. The pressure cells were built into a thermostatic cupboard, which allowed maintenance of isothermal conditions during the measurement (Makó, 2005).

All NAPL retention values used in this investigation were measured at 20°C in the pressure range of 0.0-150 kPa.

Primary regression analysis was performed using basic soil properties (percentage of clay and silt content, the bulk density, and organic matter and $CaCO_3$ content) with the purpose of removing possible outliers. The results of the preliminary data analysis performed using a principal component analysis (SPSS 13.1/Varimax rotation with Kaiser normalization) and outlier detection (casewise diagnostic) were published partly in our former study (Hernádi and Makó, 2014). Here, only the dependence structures between predictor and response variables used in this study were presented by the biplots of the reconstructed original variables and the individual observations on the first two principal factors. In addition, correlation matrixes of predictor and response variables were used to help examine the role of predictor variables on NAPL retention of soils.

Van Genuchten hydraulic function with three (Eq. (3)) and four (Eq. (4)) parameters (1980) and the Brutsaert equations (1966) (Eq. (5)) were fitted to all measured NAPL retention data with the nonlinear optimization method (Nonlinear regression/sequential quadratic programming).

$$\theta_{(h)} = \frac{\theta_s}{\left(1 + \left(\alpha\ P\right)^n\right)^{1-\frac{1}{n}}} \ , \qquad (3)$$

$$\theta_{(h)} = \theta_r + \frac{\theta_s - \theta_r}{\left(1 + \alpha\ P\right)^n)^{1-\frac{1}{n}}} \ , \qquad (4)$$

$$\theta_{(h)} = \frac{\theta_s}{1 + \left(\alpha\ P\right)^n} \ , \qquad (5)$$

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

Differences between water and NAPL retention curves were investigated with the nonparametric test (Kolmogorov-Smirnov test at $p < 0.05$).

The average fitting parameters (proposed by Carsel and Parish, 1988) were applied as the initial values of nonlinear regression. In addition, to avoid calculation of unidentified fitting parameter values, the following boundary conditions were used: $\alpha > 0.001$; $n > 1.01$, $n < 10$; $\theta_s > 0.01$, $\theta_s < 1$ and $\theta_r \geq 0.0$ as suggested by other authors (van Genucten et al., 2001; Vereecken et al., 2010).

To compare the accuracy of fitting, the Pearson R^2 and the root mean squared error (RMSE) (Eq. (6)) were calculated:

$$RMSE = \sqrt{\frac{\sum_1^N (y_i - \hat{y}_i)^2}{N}} \ , \qquad (6)$$

where: y_i is the measured value, $\hat{y}_i$ is the predicted values of y and N is the number of samples.

Fitting van Genuchten equations to the measured NAPL and water retention data offers the possibility to compare experimental retention data measured at different pressure levels. Thus, in our further investigations fluid retention values, calculated to 0.01, 0.2, 1, 2, 5, 10, 15, 40, 100 and 150 kPa pressure levels, were used.

For building PTFs and validation puproses, the dataset was randomly split with the ratio of 90:10 (calibration and validation data). The necessity of transforming response variables (parameter α and n) was investigated with hypothesis testing in mind, based on the examination of their distribution with the Kolmogorov-Smirnov test (as the dataset includes more than 50 observations). Later, the predicted parameters were obtained with reverse transformation.

In most NAPL fate and transport simulation models, the average fitting parameters of texture groups can be commonly used (eg HSSM and STOMP models, NAPL simulator), thus the parametric PTFs were constructed by grouping the data by texture for silty clay, silty clay loam,

silt loam, and sandy loam samples (where n>20) using the basic soil parameters and their transformed values (linear, reciprocal, multiplied, and inverse) as predictor variables (as proposed by Wösten *et al.,* 1995).

PTFs were developed with the multiple linear regression method (SPSS 13.01 - Linear regression). The stepwise iteration method was chosen for the selection of variables to enter the equation.

The accuracy and reliability of the developed parametric PTFs and the parameter estimation were determined by R^2 and *RMSE* and the accuracy of parametric PTFs was compared to the accuracy of commonly used estimation methods proposed by Leverett (1941) and Lenhard and Parker (1987) by the same criteria.

RESULTS AND DISCUSION

Considering the R^2 values, all the three fitting methods showed the same accuracy for fitting NAPL retention curves (Table 1). The similar applicability of van Genuchten and the Brutsaert equations was already observed by other authors on different datasets (Makó, 2005; Rathfelder and Abriola, 1996).

RMSE values showed also the same level of accuracy when van Genuchten or Brutsaert equations were used. In spite of the higher accuracy of the four-parameter form of the van Genuchten equation than the three-parameter form, the latter was used in our further investigations to avoid the error caused by the possible correlation between parameters.

Parameter α and n showed nonlinear behaviour in their distribution, thus logarithmic transformed values of these response variables were calculated (as proposed by Wösten *et al.,* 1995). The transformation resulted in a negatively skewed population of parameter α, but its Kurtosis index decreased remarkably. The transformed parameter n showed more statistically normal distribution (less tailed but rightly skewed) than the original fitted parameter (Table 2).

Significantly higher fluid retention was observed in water-air porous media systems than in air-NAPL-porous media systems at all pressure levels (Fig. 1a), which was proven by the nonparametric test ($p < 0.05$).

The NAPL curve starts at a slightly lower fluid content at saturation than water saturation and then decreases more rapidly as the pressure increases compared to water retention. This could be explained by the differences in fluid properties (density, interfacial tension, *etc*.) and/or soil properties (capillarity, hydrophobicity, *etc*.) or the gravity forces, all of which influence the interactions between the fluids and the porous media (Lenhard and Parker, 1987; Lu *et al*., 2007; Tuller and Or, 1999).

The deviation between NAPL retention and water retention of coarser textured soils (silty loam and sandy loam) in the lower pressure range was significant ($p < 0.05$). The gravitational pores of sandy loam soils filled with NAPL were drained faster than the pores of the finer textured soil samples. As the pressure increased, the difference between NAPL and water retention became smaller and less variant by texture. The maximum deviation was shifted to a higher pressure level in the case of finer textured soils (Fig. 1b).

A similar result was found by Haverkamp *et al*. (2005) investigating the possibility of development of an integral measure of the slope of SWRC with a single shape factor. Finer textured soils had larger shape factors *ie* sandy soils release water more readily than soils with higher clay content.

T a b l e 1. Accuracy of various hydraulic functions to parametrize NAPL retention

Equation	*RMSE* (vol%)	R^2
Three-parameter form van Genucten equation (Eq. (3))	0.017	0.99
Four-parameter form van Genucten equation (Eq. (4))	0.012	0.99
Brutsaert equation (Eq. (5))	0.017	0.99

T a b l e 2. Descriptive statistics of fitting parameters and their transformed values

Parameter	Min.	Max.	Mean	Std. error	Std. dev.	Skewness	Kurtosis
θ_s	0.212	0.789	0.422	0.007	0.116	0.528	0.034
α	0.002	0.999	0.481	0.026	0.434	0.194	-1.823
n	1.028	2.503	1.246	0.013	0.226	2.123	6.951
ln (α)	-6.215	-0.001	-1.582	0.095	1.606	-0.536	-0.913
ln (n)	-3.576	0.407	-1.760	0.050	0.846	0.223	-0.899

a

Mean water and NAPL retention (vol %)

Pressure (kPa)

Distilled
NAPL

b

Mean water and NAPL retention (vol %)

Pressure (kPa)

Texture
Clay
Clay loam
Sandy clay loam
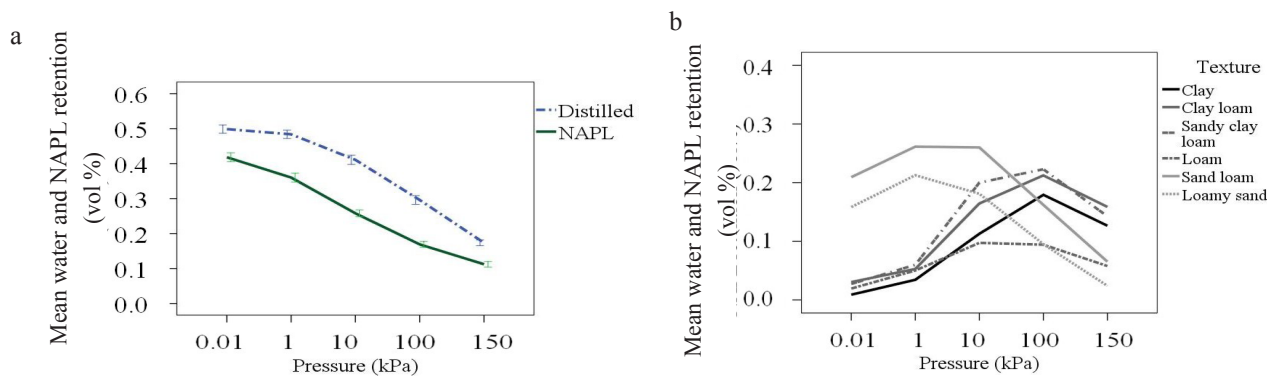Loam
Sand loam
Loamy sand

Fig. 1. Difference between soil water and NAPL retention: a – mean of the fitted fluid retention values for sandy clay loam soils, b – deviation of fluid retentions for selected six texture groups at different pressure levels.

T a b l e 3. Correlation matrix of the original and the reconstructed predictor and response variables

Para-meter	Clay	Silt	Sand	OM	CaCO$_3$	BD	$\theta_{0.01}$	$\theta_{0.2}$	$\theta_{2\,kPa}$	θ_5	θ_{15}	θ_{40}	θ_{100}	θ_{150}
Clay	0.929	0.423	-0.789	0.631	-0.271	-0.486	0.516	0.526	0.546	0.544	0.732	0.879	0.919	0.933
Silt	0.423	0.936	-0.826	0.506	0.607	-0.876	0.893	0.888	0.885	0.936	0.871	0.695	0.641	0.610
Sand	-0.789	-0.826	0.963	-0.673	-0.229	0.824	-0.852	-0.854	-0.863	-0.894	-0.960	-0.932	-0.921	-0.908
OM	0.631	0.506	-0.673	0.493	0.031	-0.523	0.544	0.548	0.558	0.572	0.655	0.684	0.690	0.688
CaCO$_3$	-0.271	0.607	-0.229	0.031	0.797	-0.501	0.496	0.484	0.466	0.517	0.315	0.034	-0.049	-0.090
BD	-0.486	-0.876	0.824	-0.523	-0.501	0.831	-0.850	-0.846	-0.846	-0.891	-0.857	-0.720	-0.678	-0.651
$\theta_{0.01}$	0.516	0.893	-0.852	0.544	0.496	-0.850	0.870	0.866	0.867	0.912	0.883	0.750	0.708	0.682
$\theta_{0.2}$	0.526	0.888	-0.854	0.548	0.484	-0.846	0.866	0.863	0.864	0.908	0.884	0.755	0.715	0.690
θ_2	0.546	0.885	-0.863	0.558	0.466	-0.846	0.867	0.864	0.865	0.909	0.891	0.769	0.731	0.706
θ_5	0.544	0.936	-0.894	0.572	0.517	-0.891	0.912	0.908	0.909	0.956	0.927	0.788	0.745	0.718
θ_{15}	0.732	0.871	-0.960	0.655	0.315	-0.857	0.883	0.884	0.891	0.927	0.967	0.906	0.886	0.869
θ_{40}	0.879	0.695	-0.932	0.684	0.034	-0.720	0.750	0.755	0.769	0.788	0.906	0.948	0.958	0.956
θ_{100}	0.919	0.641	-0.921	0.690	-0.049	-0.678	0.708	0.715	0.731	0.745	0.886	0.958	0.976	0.978
θ_{150}	0.933	0.610	-0.908	0.688	-0.090	-0.651	0.682	0.690	0.706	0.718	0.869	0.956	0.978	0.982

Clay (<2 μm), Silt (50-2 μm), Sand (2000-50 μm), BD is the dry bulk density (g cm^{-3}), OM is the organic matter content and θ is the NAPL retention of the soil samples (vol%) at a given pressure (kPa).

Significant correlations were found between basic soil properties and measured NAPL retention of soil (Table 3).

Communality values (diagonal of the correlation matrix) show that all the variances of each correlated variables might be accounted for by the components, except the communality of organic matter and CaCO$_3$ content, which were somewhat lower than 0.8. The mean communality value was higher than criterion 0.6 (for N>250).

CaCO$_3$ content correlates well with NAPL retention, especially at the lower pressure range (<15 kPa). Organic matter content of soils correlates with NAPL retention at every pressure. NAPL retention is strongly correlated with the bulk density. The negative correlation between bulk density and NAPL retention values shows proportional relationships decreasing inversely with increasing pressure.

T a b l e 4. Correlation matrix of the predictor and response variables (parameter estimation)

Parameter	Clay	Silt	Sand	OM	CaCO$_3$	BD	θ_s	ln(α)	ln(n)
Clay	0.495	0.489	-0.647	0.153	0.288	-0.368	0.463	-0.192	-0.032
Silt	0.489	0.711	-0.784	0.375	0.217	-0.017	0.201	0.206	-0.468
Sand	-0.647	-0.784	0.937	-0.342	-0.333	0.259	-0.441	-0.002	0.320
OM	0.153	0.375	-0.342	0.267	0.022	0.227	-0.109	0.330	-0.439
CaCO$_3$	0.288	0.217	-0.333	0.022	0.187	-0.317	0.345	-0.229	0.111
BD	-0.368	-0.017	0.259	0.227	-0.317	0.800	-0.734	0.744	-0.641
θ_s	0.463	.201	-0.441	-0.109	0.345	-0.734	0.722	-0.625	0.462
ln(α)	-0.192	0.206	-0.002	0.330	-0.229	0.744	-0.625	0.761	-0.746
ln(n)	-0.032	-0.468	0.320	-0.439	0.111	-0.641	0.462	-0.746	0.840

θ_s, ln(α) and ln(n) are fitting parameters of the P-S functions, or their transformed values. Other explanations as in Table 3.

Strong correlations were found between predictor variables as well, *eg* the role of CaCO$_3$ and organic matter content on soils bulk density.

Nonredundant residuals 24 sample (26.0% of data) (computed between observed and reproduced correlations) have absolute values greater than 0.05, which means the model is consistent with the empirical data.

Table 4 shows the same significant correlation between the basic soil properties and fitting parameters of NAPL retention curves or their transformed values.

As could be expected, θ_s, which was constrained at 0.01 kPa, was highly correlated with bulk density and to a lesser extent with finer fractions of soils. In addition, ln(α) was proportional to bulk density, similar to the findings of Cornelis *et al.* (2005) and results of principal component analysis by Vereecken *et al.* (1989) investigating the correlation between soil properties and water retention. The latter was expected, as α is considered to be inversely proportional to the bubbling or air entry pressure of soil (van Genuchten *et al.*, 1991), which is strongly affected by porosity, and thus by bulk density, texture, and structure of soils.

Strong correlations of silt content were obtained with the parameter θ_s and ln(n), but clay content showed a remarkable correlation only with θ_s. Cornelis *et al.* (2005) demonstrated a negative correlation between n and clay or silt content. In addition, as a consequence of this, ln(n) was positively correlated with sand content. Furthermore, it was shown that CaCO$_3$ and organic matter content have significant effects on fitting parameters as well.

However, the reproduced correlation matrices indicate that 19 (52.0%) of the residuals have absolute values greater than 0.05, which might imply that the model does not fit to the empirical correlations really well.

The plot of the reconstructed original variables and the individual observations on the first two principal factors also shows a strong correlation between PSD fractions and NAPL retention (Fig. 2a).

The sand content and bulk density (pointing in the opposite direction) correlate negatively with NAPL retention. After rotation, the first component accounted for 46.11% and the second for 42.29% of the total variability.

The biplot of the PCA of basic soil properties and fitting parameters shows lower correlations; only 63.57% (33.86 and 29.70%) of the variance could be explained by the components. Parameter α is primary and positively influenced by the bulk density and seems not to be strongly correlated with any other parameters. Parameter θ_s and ln(n) were affected negatively by bulk density and positively by the clay, silt, and CaCO$_3$ content (Fig. 2b).

The results of Bartlett test (sphericity <0.5 criterion at p<0.05) and Kaiser-Meier-Olkin test (>0.6) verify the acceptance of sampling adequacy to build proper PTFs for predicting NAPL retention, but suggest that further extension of the dataset is needed for accurate and reliable parameter estimation.

Considering the results of preliminary regression analysis performed for outlier detection, records of pure clay minerals were removed. The randomly split dataset for the selected four texture groups contains the basic soil properties and NAPL retention of 285 soil samples (Table 5).

According to the results of preliminary regression analysis, NAPL might be accurately predicted with PTFs using basic soil properties. Iteration with the stepwise method resulted in PTFs (point estimation) with R^2 ranging from 0.360 to 0.545 and *RMSE* from 1.05 to 1.27 vol%.
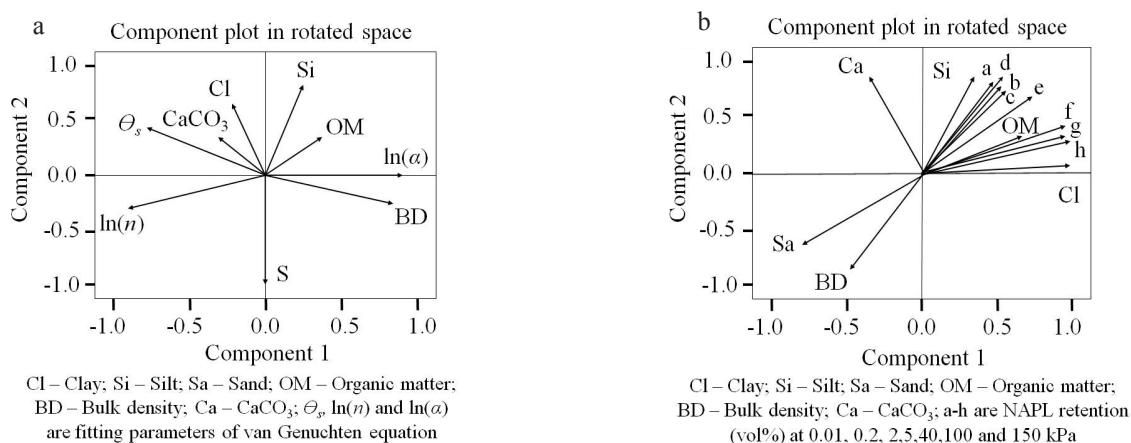
Fig. 2. Component plots in the rotated space: a – component scores of variables used in parametric PTFs, b – component scores of variables parameter θ_s and transformed values of parameter n and α.

T a b l e 5. Descriptive statistics of the basic soil parameters and NAPL retention

Parameter	Unit	Calibration data					
		Mean	Std. error	Std. dev.	Median	Min.	Max.
Clay		25.62	1.02	17.18	22.25	0.00	67.67
Silt		41.69	1.09	18.32	40.64	0.00	78.50
Sand	%	26.72	1.45	24.49	20.00	0.00	98.38
OM		1.33	0.08	1.27	0.94	0.00	5.66
CaCO$_3$		5.85	0.49	8.26	0.70	0.00	30.71
BD	g cm^{-3}	1.40	0.02	0.28	1.43	0.60	1.99
NAPL retention (vol%)							
0.01		45.60	0.82	11.29	46.44	20.83	83.32
0.2		40.54	0.70	11.81	40.02	20.50	79.22
2.0		35.49	0.78	10.67	35.69	2.32	67.05
5.0		32.09	0.73	10.07	31.84	1.86	62.29
10.0		22.99	1.04	10.23	20.35	3.73	60.98
15.0	kPa	24.62	0.60	8.22	24.58	1.25	62.29
20.0		20.98	1.02	10.02	18.68	3.24	59.46
40.0		19.64	0.57	7.82	20.40	0.91	36.84
50.0		19.27	0.86	8.44	16.96	2.84	52.88
100.0		17.67	0.46	7.84	17.53	0.71	50.15
150.0		9.69	0.67	3.49	8.70	1.00	15.30

T a b l e 5. Continuation

Parameter	Unit	Validation data					
		Mean	Std. error	Std. dev.	Median	Min.	Max.
Clay		29.45	3.35	18.66	27.05	0.00	67.67
Silt		44.87	3.63	20.21	51.55	0.00	78.50
Sand	(%)	22.45	5.04	28.06	6.84	0.00	98.62
OM		1.55	0.21	1.18	1.47	0.00	4.00
$CaCO_3$		6.53	1.75	9.74	0.00	0.00	30.00
BD	(g cm^{-3})	1.47	0.03	0.18	1.50	1.14	1.82
NAPL retention (vol%)							
0.01		40.17	2.18	9.99	38.43	25.68	58.20
0.2		34.40	2.20	12.27	32.44	16.63	71.85
2.0		29.81	2.89	12.90	29.96	4.30	50.40
5.0		27.99	2.49	11.40	28.35	3.50	47.80
10.0		26.54	3.95	12.50	24.03	17.60	60.35
15.0	(kPa)	22.86	1.75	8.04	23.80	2.40	32.93
20.0		24.65	3.59	11.36	22.34	15.76	54.74
40.0		18.96	1.79	8.18	20.37	1.50	32.56
50.0		21.54	2.24	7.09	19.67	14.69	37.84
100.0		17.86	1.48	8.24	18.93	1.30	32.98
150.0		3.27	2.37	4.10	0.90	0.90	8.00

Increased accuracy was found when the more detailed fractions of PSD and the transformed values of predictor variables were used in building parametric PTFs for the selected texture groups (primary grouping the data by texture) (Table 6).

RMSE values correspond to the range of *RMSE* values obtained by other authors predicting water retention with PTFs (approximately 2-7 vol% - 0.02-0.07 m³ m⁻³) (Minasny *et al*., 1999; Schaap and Leij, 1998; Shaap *et al*., 1999) concerning the accuracy and reliability as well.

According to R^2, a decrease in accuracy and reliability was observed with increasing pressure, but *RMSE* values did not show this tendency.

Fitting parameters of the van Genuchten equations were accurately predicted by PTFs after grouping the data by texture. Lower reliability of these PTFs was found, which might be the consequence of differences in soil properties between the calibration and validation subsets of data (Table 7).

The lowest predictability of parameter α and θ_s might also be caused by the strong correlation between them.

A similar strong correlation was found by Minasny *et al.* (2004) and Cornelis *et al.* (2005) between fitting parameters, which were explained by overparametrization of the model. Although the strong correlation of the fitting parameters and unique problems leading to poor convergence and ill-defined parameter values might be the consequence of the improperly imposed restrictions on the parameter values or inadequate selection of the initial parameter values.

Vereecken *et al*. (1989) noted that the accuracy of parametric PTFs for estimating water retention was the most sensitive to changes in θ_s amongst the parameters of VG equations.

T a b l e 6. Accuracy and reliability of parametric PTFs for predicting NAPL retention

Parameter (kPa)	Accuracy (n=264)		Reliability (n=26)	
	R^2	*RMSE* (vol%)	R^2	*RMSE* (vol%)
0.001	0.856	0.417	0.565	3.730
1.0	0.898	0.398	0.680	3.310
10.0	0.883	0.363	0.329	2.950
100.0	0.840	0.318	0.353	2.090
150.0	0.800	0.321	0.481	0.526

T a b l e 7. Accuracy and reliability of parameter estimation with PTFs

Parameter	Accuracy (n=264)		Reliability (n=26)	
	R^2	*RMSE* (vol%)	R^2	*RMSE* (vol%)
α	0.987	1.359	0.454	0.966
n	0.900	1.775	0.837	1.775
θ_s	0.988	0.750	0.565	0.750

Fig. 3. Accuracy of different estimation methods compared by their (a) R^2 and (b) *RMSE* values (vol%). PTF_PAR – parametric PTFs for texture groups (predictor variables: PSD fractions (except the coarse sand), organic matter and $CaCO_3$ content, bulk density, and their transformed values. PTF_PO – PTFs – point estimation (predictor variables: clay, silt, organic matter and $CaCO_3$ content, bulk density) SC – is the scaling method proposed by Lenhard and Parker (1987) (Eq. (2)). LEV – is the Leverett method (Leverett, 1941) (Eq. (1)).

In addition, the fitted value of θ_s is usually lower than the measured fluid content at saturation, which might lead to lower accuracy of predicting the saturated end point of the fluid retention curves (van Genuchten *et al.*, 1991).

We must take it into account also the limited applicability of our developed PTFs, which is similarly restricted by our dataset upon which they were based, as formerly noted by several authors (Cornelis *et al.*, 2005; Schaap and Leij, 1998). Larger and more heterogeneous datasets with representative data of various soils are required to deve-

lop accurate and reliable PTFs. Further investigations are needed to examine the effects of other soil properties *eg* structure of soil NAPL retention.

Our dataset contains fluid retention values determined by the same measurement method, which might lead to apparently increased accuracy of PTFs, as reported by Cornelis *et al.* (2005) and Vereecken *et al.* (1989).

In addition, we had an opportunity to measure NAPL retention of soils between 0-150 kPa. Contrarily, PTFs which were built using fluid retention data determined by

pressure plate devices are generally limited to matric suction <= 1 500 kPa. Lu *et al.* (2007) suggest the use of soil SWRC data from 0-1 500 kPa if higher accuracy of prediction is required.

However, the accurate estimation of fluid retention at the end points of the fluid retention curves has not been yet entirely resolved, even in the case of fitting hydraulic functions to water retention data either due to the problems in parameterization itself and the difficulties in precise determination of the soil water content at low pressure, near saturation (Schaap and Leij, 1998; Vereecken *et al.*, 1989). These models might also give relatively poor estimation at low water content, where the soil matric potential is related not only to capillary forces, but also to adsorptive forces. In addition, the accuracy of fitting might be affected by the measurement method of soil water retention and other soil properties, such as the organic matter content and structural properties, *etc.* as well (Cornelis *et al.*, 2005; Minasny *et al.*, 1999).

In our study, better applicability of various PTFs was found for estimating NAPL retention compared to either the commonly used scaling methods proposed by Leverett or the scaling method of Lenhard and Parker (Fig. 3).

In Fig. 3, the *RMSE* values of Leverett-type estimation were not presented as they exceed 25 in all cases.

According to our results, using neither the Leverett equation nor the scaling method gives accurate estimation for NAPL retention of the investigated soil samples.

CONCLUSIONS

1. The dataset containing basic soil properties, water and a nonaqueous phase liquid (namely DUNASOL 180/220) retention data of 285 samples was built to investigate the possibility of development of pedotransfer functions to predict soil nonaqueous phase liquid retention.

2. Significant differences (p<0.05) were found between the measured water and organic liquid retention. The deviations varied with soil texture and with the applied pressure values.

3. The commonly used fitting equations (three and four parameter van Genuchten, Brutsaert) were suitable to describe organic liquid retention curves (R^2 = 0.99, root mean squared error <0.012 vol%).

4. The logarithmic transformed values of response variables (parameters α and n of fitting equations) were suggested for use in building parametric pedotransfer functions.

5. Based on the results of principal component analysis, an important impact of bulk density, and clay, silt, sand, $CaCO_3$, and organic matter content on organic liquid retention was detectable; therefore, the involvement of these soil parameters into the nonaqueous phase liquid retention pedotransfer functions is suggested.

6. The preliminary statistical analysis to build and develop parametric pedotransfer functions gave promising results. Using more detailed fractions of particle size distribution and the transformed values of soil properties resulted in increased accuracy of prediction for the selected texture groups (primary grouping). The reliability of pedotransfer functions was lower to some degree than accuracy.

7. Sufficient accuracy of parameter estimation and lower reliability of predicting fitting parameters were found. The lower predictability of parameter of the fluid content at saturation and ln(α) than ln(n) might be a result of the error of the fitting procedures or the strong correlation between fitting parameters.

7. Our results showed limited applicability of conventional scaling methods compared to prediction of nonaqueous phase liquid retention using pedotransfer functions. Further research is planned to investigate the applicability of the combination of peodtransfer functions and scaling methods (predicting organic liquid retention of arbitrary non-polar liquids) to an extended dataset containing organic liquid retention data measured with different organic liquids.

REFERENCES

Assouline S. and Or, D., 2013. Conceptual and parametric representation of soil hydraulic properties: A review. Vadose Zone J., (12) 1-20, doi:10.2136/vzj2013.07.0121.

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

Blake G.R. and Hartge K.H., 1986. Bulk Density. In: Klute, A. (ed.) Methods of soil analysis. Part 1. Agron. Monogr., 9 ASA and SSSA, Madison, WI, USA.

Botula Y.-D., Nemes A., Mafuka P., Van Ranst E., and Cornelis W., 2013. Prediction of water retention of soils from the humid tropics by the non-parametric k-nearest neighbor approach. Vadose Zone J. 12 (2), doi: 10.2136/vzj2012.0123.

Brutsaert W., 1966. Probably laws for pore size distributions. Soil Sci., 101, 85-92.

Carsel R.F. and Parish R.S., 1988. Developing joint probability distribution of soil water retention characteristics. Water Resour. Res., 24, 755-769.

Cornelis W.M., Khlosi M., Hartmann R., Van Meirvenne, M., and De Vos B., 2005. Comparison of unimodal analytical expressions for the soil-water retention curve. Soil Sci. Soc. Am. J., 69, 1902-1911.

Gee G.W. and Bauder J.W., 1986. Particle-size Analysis. In: Methods of soil analysis (Ed. A. Klute). Part 1. Agron. Monogr. 9. ASA and SSSA, Madison, WI, USA.

Guarnachia J., Pinder G. and Fishman M., 1997. NAPL: simulator documentation. National Risk Management Research Laboratory. Ada, OK 74820. EPA/600/R-97/102.

Haverkamp R., Leij F.J., Fuentes C., Sciortino A., and Ross P.J., 2005. Soil water retention: I. Introduction of a shape index. Soil Sci. Soc. Am. J., 69, 1881-1890.

Hernádi H. and Makó A., 2014. Dataset for creating pedotransfer functions to estimate organic liquid retention of soils. J. Env. Geo., 7 (1-2), 11-22.

Lamorski K., Bieganowski A., Ryżak M., Sochan A., Sławiński C., and Stelmach W., 2012. Assesment of the usefulness of particle size distribution measured by laser diffraction for soil water retention modelling. J. Plant Nutr. Soil Sc., 177(5), 803-813.

Lenhard R.J. and Parker J.C., 1987. Measurement and prediction of saturation-pressure relations in three phase porous media systems. J. Contam. Hydrol., 1, 407-424.

Leverett M.C., 1941. Capillary behavior in porous solids. Trans. of AIME, 142(1), 152-169.

Lu S., Tusheng R. and Yuanshi G., 2007. Evaluation of three models that describe soil water retention curves from saturation to oven dryness. Soil Sci. Soc. Am. J., 72, 1542-1546.

Makó A., 2005. Measuring the two-phase capillary pressure-saturation curves of soil samples saturated with nonpolar liquids. Commun. Soil Sci. Plan., 36, 439-453.

Makó A. and Hernádi H., 2013. Hydrocarbon derivates in soils: Soil physical researches (in Hungarian). Pannon University, Veszprém, Hungary.

Minasny B., Hopmans J.W., Harter T., Eching S.O., Tuli A., and Denton M.A., 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. Soil Sci. Soc. Am. J., 68, 417-429.

Minasny B., McBratney A.B., and Bristow K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma, 93, 225-253.

Nelson R.E., 1982. Carbonate and gypsum. In: Methods of soil analysis (Ed. A.L. Page) . Part 2. Agron. Monogr. 9. ASA and SSSA, Madison, WI, USA.

Rajkai K., Kabos S. and van Genuchten M.T., 2004. Estimating the water retention curve from soil properties: comparison of linear, nonlinear and concomitant variable methods. Soil Till. Res., 79. 145-152.

Rathfelder K. and Abriola L.M., 1996. The influence of capillarity in numerical modelling of organic liquid redistribution in two-phase systems. Adv. Water Res., 21(2), 159-170.

Schaap M.G. and Leij F.J., 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci., 163, 765-779.

Tomasella J., Hodnet M.G., and Rossato L., 2000. Pedotransfer functions for the estimation of soil water retention in Brazilian soils. Soil Sci. Soc. Am. J., 64, 327-338.

Tuller M., Or D., and Dudley L.M., 1999. Adsorption and capillary condensation in porous media: Liquid retention and interfacial configurations in angular pores. Water Resor. Res., 35(7), 1949-1964.

Tyurin I.V., 1931. A new modification of the volumetric method of determining soil 344 organic matter by means of chromic acid. Pochvovedenie, 5-6, 36-47.

van Genuchten M.Th., 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J., 44, 892-989.

van Genuchten M.Th., Leij F.J. and Jates S.R., 1991. The RETZ code for quantifying the hydraulic functions of unsaturated soils. EPA/600/2-91/065, U.S. Environmental Protection Agency, Ada, OK.

van Genuchten M.Th. and Nielsen D.R., 1985. On describing and predicting the hydraulic conductivity of unsaturated soils. Ann. Geophys., 3, 615-627.

Vereecken H., Maes J., Feyen J., and Darius, P., 1989. Estimaing the soil moisture retention characteristic from texture, bulk density and carbon content. Soil Sci., 148, 389-403.

Weaver J.W., Charbeneau R.J., Taux J.D., Lie B.K., and Provost J.B., 1994. The hydrocarbon spill screening model (HSSM). 1. US EPA. EPA/600/R-94/039a.

Wösten J.H.M., Lilly A., Nemes A., and Le Bas C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma, 90, 169-185.