Water retention of salt affected soils: quantitative estimation using soil survey information

Brigitta Tóth[a1], András Makó[a], Alberto Guadagnini[b], Gergely Tóth[c]

[a]University of Pannonia, Georgikon Faculty, Department of Crop Production and Soil Science, Keszthely, Hungary

[b]Dipartimento di Ingegneria Idraulica Ambientale, Infrastrutture Viarie e Rilevamento,Politecnico di Milano, Milan, Italy

[c]European Commission, Joint Research Centre, Institute for Environment and Sustainability, Land Management and Natural Hazards Unit, Ispra, Italy

Running title: Estimation of water retention of salt-affected soils

Acknowledgements

In the framework of Grant No. T048302 provided by the National Scientific Research Fund (OTKA) we had the opportunity to develop the Hungarian Detailed Soil Hydrophysical Database (Hungarian acronym: MARTHA) with the collaboration of the County Offices of the Hungarian Plant and Soil Protection Service. This work was supported by the MyWater project which is co-funded by the European Community's Seventh Framework Programme (FP7/2007- 2013). We would like to

[1] Corresponding author. H-8360 Keszthely, Deák F. u. 16., Hungary Tel: +393472269181, E-mail address: gema@freemail.hu (B. Tóth)

thank Dr. Csilla Farkas, Hilda Hernádi, Marth Péter for their work during the database construction, Alessandro Rossi, for his support to the statistical analysis and Henk Wösten and Attilla Nemes for their valuables comments on the manuscript.

Abstract

Soil water retention (SWR) at -0.1, -33, -1500, and -150000 kPa matric potentials and available water content (AWC) were estimated from information available from 729 horizons of salt-affected soils in the Hungarian Detailed Soil Hydrophysical Database. Soil characteristics of the 1:10,000 scale Hungarian soil maps were used as input parameters. Ordinal and nominal (categorical) variables: texture, organic matter content, calcium carbonate content, soluble salt content, pH, and soil subtype classes of the soil map were used to develop a new prediction method based on the CHAID classification tree. Results of the model development were compared with results using conventional prediction methods (classification tree (CRT) and multiple linear regression (MLR)). Four types of pedotransfer rules were established by classification tree methods. The first rule uses continuous-type input parameters, the second uses soil taxonomical information in addition, the third and fourth one uses categorical-type input parameters. In addition, continuous pedotransfer functions (point estimations) were established as well. Results show that the root mean square error (RMSE) of the developed method is between 1.25 vol% (-150000 kPa) and 6.40 vol% (-33 kPa). With the mentioned available input parameters, for salt-affected soils the prediction reliability is similar with categorical and continuous-type information. To predict SWR from categorical-type information the CHAID method is advisable. In the case of continuous-type input parameters MLR is suggested, based on this study. The established hydropedologic methods can be readily used to prepare available water content maps for the topsoil of salt affected soils based on solely soil survey information.

Supplemental materials are available for this article. Go to the publisher's online edition of Arid Land Research and Management to view the free supplemental file.

Keywords: classification tree, CHAID, CRT, multiple linear regression, pedotransfer function, pedotransfer rule, soil map, solonchak, solonetz, water capacity can be readily used to prepare available water content maps for the topsoil of salt
affected soils based on solely soil survey information.

Introduction

Salt-affected soils always represent a special case when developing pedotransfer functions. This particularity originates from the unique physico-chemical properties of salt-affected soils, which has also a fundamental influence on soil water characteristics (Várallyay, 2002) and related land management options. Literature shows that besides particle size and pore size distribution, chemical properties of the soils are also essential to describe the water retention of these soils (Dane and Klute, 1977; Acharya and Abrol, 1978; Jayawardane and Beattie, 1978; Rajkai, 1988; Lima et al., 1990; Chaudhari and Somawanshi, 2000; Groenevelt et al., 2004). Therefore pedotransfer functions developed for non salt-affected soils can not be applied for salt-affected ones.

Estimation of soil water retention characteristics of Hungarian salt-affected soils was carried out earlier by Rajkai (1988) on a dataset specific for a small region of Hungary (Hortobágy) also showing that the established method is not applicable for salt affected soils from other regions.

The required data on soil hydrological properties to operate water management models are often not available or difficult to measure. Therefore it is attractive to estimate these properties from soil information which is available, or easy to produce. Despite its potential to improve predictions, soil taxonomic information is not commonly used in hydrophysical models. The importance of soil taxonomical information in estimation of soil water retention at different scales was demonstrated by Pachepsky et al. (1982). Rawls et al. (2003) found that information on soil taxonomic order is an important factor for the estimation of soil water content. In our previous studies (Tóth et al., 2005; Tóth et al., 2008) using factor analysis it was also concluded that preliminary grouping based on Hungarian soil subtypes can improve the estimation efficiency of SWR.

In Hungary 1 : 10,000 detailed soil maps are available for more than half of the arable land area (Tóth and Máté, 2006). These maps contain information about basic soil properties such as taxonomic information, organic matter content, texture, calcium carbonate content, pH, soluble salt content but not on hydrological properties.

The aim of the current study was to investigate whether pedotransfer rules based on the CHAID classification tree method using available soil map information (ordinal and nominal soil properties) could predict water retention at -0.1, -33, -1500 and -150000 kPa matric potentials and plant available water content of salt-affected soils with an accuracy similar to that of pedotransfer functions. Because we wanted to make full utilization of our national hydrophysical database we also included prediction of SWR at -150000 kPa into our study, despite its smaller importance for agricultural practice and environmental management. The measurements of hygroscopic water content (SWR at -150000 kPa) were used by soil scientists in the past to predict other soil hydrophysical properties. It is also used to fit water retention curve und - as other measurements of SWR under -1500 kPa - it is also important in applied soil mechanics (Vanapalli et al., 1998).

The importance of this investigation is twofold. On one hand, the currently available estimates of water retention of salt affected soils are derived from a limited number of observations, therefore those might be biased and may need to be improved. Our study aims to broaden the knowledge base and understanding of soil water retention of salt affected soils. On the other hand, no PTFs of salt affected soils are available for categorical-type information, which are presented on soil maps and which form a basis for amelioration measures in practice.

We hypothesised that the special water holding capacity characteristics of salt affected soils can be expressed through pedotransfer functions. Furthermore, we hypothesized that categorical-type soil information carries information for water retention characterisation comparable in their importance to continuous-type soil data. In order to test the validity of these hypotheses a series of statistical analysis was performed.

Four pedotransfer rules were developed based on the classification tree method:

1. Based on measured continuous-type soil properties only.
2. Based on measured continuous-type soil properties (the same properties as in the first method) and also involving taxonomic soil subtype.

3. Applying a method using similar soil properties to the first two methods (including soil subtype) but all classified to categorical-type data using:
    a. regression tree and
    b. CHAID method.

In addition to the four pedotransfer rules a continuous pedotransfer function was also established for the four different matric potentials separately using similar – continuous-type – input parameters as used in the first classification tree.

Plant available water content was predicted and also calculated from the results of the models that estimate water retention at -33 and -1500 matric potentials.

Based on the work of Pachepsky et al. (1996), Børgesen and Schaap (2005) and according to our previous experience (Tóth et al., 2005) water retention at a given matric potential can be better estimated by point estimation methods than by parametric estimation methods. Since our aim is to use the predicted water retention values to prepare maps with the best possible accuracy, therefore we developed point estimation methods for the key water contents instead of parametric estimation methods.

Materials and methods

Database

This study used 729 samples taken from genetic horizons at different depths of 246 soil profiles of salt-affected soils. Data was extracted from the Hungarian Detailed Soil Hydrophysical Database (Hungarian acronym: MARTHA) (Makó et al., 2010). The MARTHA database contains 7035 samples with measured physico-chemical properties and water retention values at all four distinct matric potentials: at -0,1, -33, -1500 and -150000 kPa. All samples from profiles of salt affected soils (Baranyai, 1989) were selected for the study. (Appendix 1. and 2. in the on-line publication of this paper shows the spatial and taxonomic representativity of the samples.) 90 % of the dataset was used as training sample (653 soil horizons), 10 % as test sample (76 soil horizons). Salt affected soils of the dataset belong to 22 soil subtype units of the Hungarian classification system which cover six WRB Reference Soil Groups (WRB, 2006) including Solonetz, Solonchak, Phaeozem, Chernozem, Cambisol and Calcisol. Focus of this analysis was on soil classes according to the Hungarian taxonomic classification, because these are available on national soil map cartograms. The database contains information on both chemical and physical soil properties in continuous-type form. From these soil properties use is made of sand (0.05 – 2 mm) (mass %), silt (0.002-0.05 mm) (mass %), clay (<0.002 mm) (mass %), organic matter content (mass %), calcium carbonate content (mass %), pH_{H2O} and soluble salt content (mass %). The soil analyses were performed in compliance with standard methods (particle size distribution: Gee and Bauder, 1986; organic matter content: Tyurin, 1931; calcium carbonate content: Nelson, 1982; pH_{H2O}: McLean, 1982; soluble salt content: MSZ, 1978). The soil water retention values at different matric potentials were measured following a Hungarian standard (Várallyay, 1973) by using sand box, kaolinite box and pressure membrane extractor. Statistics on basic properties of soil samples in the training and test datasets (Table 1) show the variability of the samples.

One of our aims was to establish a prediction method which can be applied for categorical-type (ordinal and nominal) information of soil maps. Therefore the continuous data of our database was classified according to classes used on Hungarian soil maps (Baranyai, 1989; for details see Appendix 3-7 of the on-line publication of this paper). Analyzed soil properties and the number of their categories were soil texture (N=7), organic matter content (N=14), calcium carbonate content (N=5), pH_{H2O} (N=7) and soluble salt content (N=4).

The proportion of exchangeable sodium content within the sum of exchangeable cations – on which information is available from the detailed soil maps – has a well-known effect on the soil water retention capacity (Dane and Klute, 1977; Acharya and Abrol, 1978; Lima et al., 1990; Chaudhari and Somawanshi, 2000; Várallyay, 2002). However, no data is available on this soil property in the MARTHA database. Therefore sodium content could not be directly considered in the model development and therefore was left out from the numerical analysis. However, some Hungarian soil taxonomic names provide indirect information on the sodium content (i.e. sodium content of solonetz soils exceeds 25% of the total exchangeable cations) and this information was utilized during the evaluation of the results.

Statistical methods

Pedotransfer rules with classification trees

In the first phase of the model development we used continuous-type variables to predict the soil water retention (SWR) at -0.1, -33, -1500 and -150000 kPa matric potentials and plant available water content using the Classification and Regression Tree (CRT; SPSS, 2001) method (CRT1). Plant available water content (AWC) was calculated as the difference in the amount of water content held at -33 kPa and -1500 kPa matric potentials. In the prediction those soil properties are taken into account, which are indicated (in their categorical form) on the Hungarian soil maps (Table 1): sand (0.05 – 2 mm), silt (0.002-0.05 mm), clay (<0.002 mm), organic matter content, calcium carbonate content, pH_{H2O} and soluble salt content. In the second step an analysis is done to explore the benefit of adding soil taxonomical information to the available continuous-type soil properties, again, using the CRT method (CRT2). In the third step CRT (CRT_kat) and the Chi-squared Automatic Interaction Detection (CHAID; Kass, 1980) method of the Classification Tree Analysis (SPSS, 2001) were applied using categorical information, including soil taxonomical classes as available from the detailed Hungarian soil maps. Regression tree method was used to predict soil hydrophysical properties by McKenzie and Jacquier (1999), Rawls et al. (2003), Lilly et al. (2008). Contrary to CRT, the CHAID method provides multi-way subdivisions. Those classes of predictors which are not significantly different regarding their effect on the dependent variable are merged. Variables which do not contribute to the final model significantly will be automatically excluded. Most important predictors are placed at the highest node level. The F tests are applied to determine the least significantly different predictor-pairs. The tree structure also indicates the inter-reliance of independent variables throughout the child nodes. In the CRT method, based on the independent variables, possible binary splits of the dataset are analysed to increase the homogeneity within the child node (Breiman et al., 1998). It uses the least squares deviation impurity measures to chose the best split on the independent variable and best parameter for the splitting (for more details see Breiman et al. (1998) and IBM (2011)).

In order to determine the minimal number of samples in the nodes of a tree (both CHAID and CRT) and avoid overfitting, while having the best overall prediction, an optimization process by tenfold cross-validation has been applied based on the guidelines by Breimann et al (1998). First, prediction models with different node size settings are established on the training sample (N = 653). Then during the tenfold cross-validation, the database is randomly split into ten parts and the developed models are tested on those. The tenfold cross-validation was performed ten times. Model uncertainties are characterized by a risk estimate using the mean square error (MSE) value (SPSS, 2001). That model setting was chosen to be considered optimal in which the average MSE value of the method was the smallest for the ten times tenfold cross-validated data (Hill and Lewicki, 2006).

The 0.01, 0.05 and 0.1 significance levels for splitting nodes and subsequent re-merging of nodes during pruning were tested. The 0.05 significance level was chosen to be used in this study as it was found to result in the best balance between tree size and estimation performance of the model.

In regression tree models (CRT) input parameters' contribution in prediction of the SWR can be characterized by the variable importance:

$$M(x_m) = \sum_{t \in T} \Delta R(\tilde{s}_m, t)$$

where M(x_m) is the measure of importance of variable x_m, $\Delta R(\tilde{s}_m, t)$ is the relative risk reduction at node t for the $\tilde{s}_m$ surrogate split in the T optimal subtree selected by cross-validation (Breiman et al., 1998).

The variable importance is computed by the sum of impurity decreases attributed to the variable at each node as calculated for the best split (Breiman et al., 1998). In case of the CHAID method the developed tree model clearly indicates the hierarchy of the soil properties in the prediction of SWR at given matric potential.

The optimal CRT and CHAID models were applied on the test sample (N = 76) to assess the reliability of the developed methods.

Continuous pedotransfer function with multiple linear regression

Multiple linear regression (MLR) was used to examine the difference in accuracy between pedotransfer rules based on the classification tree method (CRT1) and continuous pedotransfer functions using the same soil properties in continuous-type form. MLR models with stepwise forward technique were established separately for the four SWR at different matric potentials for the training dataset (point PTFs). In the current analysis only those soil properties were used as input parameters which are available from the Hungarian soil maps - however in their raw, continuous form. Transformation is applied on the input parameters analogue to procedures followed by Rajkai (1988), Wösten et al. (1999) and Hodnett and Tomasella (2002): variables without transformation and their linear, reciprocal, exponential relationships and interactions between sand, silt and clay fractions were used in the analysis. Natural logarithm and reciprocal values of calcium carbonate and soluble salt contents were not evaluated since the database contained samples with no $CaCO_3$ and soluble salt contents. To estimate the plant available water capacity (AWC) the same methods were applied as used for the prediction of water retention at distinct matric potentials (CRT1, CRT2, CHAID, MLR).

The seemingly logical option to test the CHAID method with continuous variables combined with the categorical data of soil type was not used. We declined this option because during the performance of the CHAID operation continuous variables are classified into distinct categories, with similar sample sizes for the categories that have no added value in comparison to CART, if full scientific exploitation of physical relationships is not a priority.

Accuracy and reliability analyses of the models

The developed water retention models were compared using the following indicators:

I. Accuracy assessment: root mean square error (RMSE) (for details see Appendix 8. of the on-line publication of this paper), the cross-validated RMSE value (except for multiple linear regression) and Pearson correlation coefficient (SPSS, 2001).

II. Reliability assessment: RMSE, mean absolute relative error (MARE) (for details see Appendix 8. of the on-line publication of this paper) and the Pearson correlation coefficient.

Average accuracy of the models was assessed on the training sample, average reliability of those were evaluated on the test sample. Statistical significance of differences was tested at the 0.05 significance level.

Results and discussion

The reliability of the developed prediction methods are shown in Table 2. Based on the RMSE values all of the developed methods show good performance during the reliability testing. RMSE values of the developed pedotransfer rules are between 1.249 and 6.403 vol% and for continuous pedotransfer functions those are between 1.350 and 5.940 vol% depending on the analyzed matric potentials. However, the correlations between the predicted and measured values of the models as indicated by Pearson coefficient are low (Table 2) especially in the case of AWC and SWR at -0.1 kPa. The highest correlation (r¼0.837) was observed in the case of the prediction of SWR at □150000 kPa with the regression tree (CRT2) model. As a result, the studied soil properties' soil water retention of the Hungarian salt-affected soils can be estimated with a reasonable RMSE value but cannot be fully explained at -0.1 and -33 kPa matric potentials, due to the well-known lack of information regarding the soil structure (Hillel, 1982). Knowledge of bulk density would possibly improve the prediction of SWR at □0.1 kPa (Nemes, 2003; Børgesen & Schaap, 2005; Twarakavi et al., 2009), and sodicity would possibly improve the estimation of SWR at both matric potentials (Rajkai, 1988), however to confirm this hypothesis further analyses are needed. Such information is not available from detailed national soil maps; therefore, we left out these analysis from our current study. SWR at -1500 and -150000 kPa matric potential can be predicted with higher reliability (Table 2) because the adsorption forces can be better explained by organic matter content and clay content at low matric potential values (Hillel, 1982). Prediction of AWC is usually weaker than that of SWR values (Minasny et al., 1999; Lipsius, 2002; Rawls et al., 2003). Lipsius (2002) analyzed eleven different methods published in literature to predict AWC. He found that the PTFs overestimated it in case of soils having low AWC and underestimated it for soils with high AWC. Minasny et al. (1999) described the prediction of AWC by its relationship with clay content. Below 25% clay content the AWC increases with increasing clay content and after that point it slowly decreases.

Although pedotransfer functions developed for certain soils (Rawls et al., 1982; Saxton et al., 1986; Rawls et al., 2003; Saxton & Rawls, 2006; Nemes & Rawls, 2006; Al Majou et al., 2007) show that soil particle size distribution with or without organic matter content is satisfactory to predict soil water retention, our study showed that in case of the Hungarian salt-affected soils, other soil properties are needed as well.

Pedotransfer rules

The summary of the developed regression (CRT1, CRT2) and classification tree (CHAID) prediction methods is presented in Table 3. Soil properties included in the models for the prediction of SWR at different matric potentials and AWC are shown.

Regression trees (CRT1, CRT2, CRT_kat)

Continuous-type input parameters at -0.1 and -33 kPa matric potentials sand, and at -1500 and -150000 kPa clay is the most important variable to predict SWR (Figure 1). In the case of category-type independent variables texture is the most important variable in the estimation. In case of having information about soil texture (CRT_kat) instead of particle size distribution (CRT2) the importance of Hungarian soil subtype increases and becomes the second (at -0.1, -33 and -1500 kPa) or third (at -150000 kPa) most important input parameter.

For the prediction of SWR at -0.1 kPa, besides sand, silt and organic matter, pH is also important in case of continuous input variables (CRT1 and CRT2).

In the case of -33 KPa, texture and Hungarian soil subtype has far the highest splitting power (CRT_kat). In case of information about particle size distribution (CRT1, CRT2) the importance of soil subtype decreases (CRT2) but it is still the fourth most important variable after sand, clay and silt content, prior to organic matter content.

At -1500 kPa Hungarian soil subtype, organic matter content and calcium carbonate content have similar importance being more or less equally at the third place in the variable importance ranking – after clay and sand content.

At -150000 kPa calcium carbonate content is the second or third in the variable importance ranking depending on the type of input parameters.

Organic matter content is ranked between third and fifth, among the studied matric potentials and it has the highest normalized importance at -0.1 kPa. Importance of soluble salt content is the lowest among the studied soil properties.

CHAID classification tree

In CHAID classification trees soil texture is the first splitting variable (Fig. 2) which divides the soils to predict the SWR at all four matric potentials, showing that it is the most important input parameter. At the next level the importance of the variable depends on the matric potential and soil texture. At -0.1 kPa pH, organic matter content and soluble salt content are at the second level, Hungarian soil subtype, calcium-carbonate content and pH are at the third level of the tree (Fig.2). For the prediction of SWR at -33 kPa besides soil texture, soil subtype becomes an important as well in case of silt or finer textures. At -1500 kPa soils finer than silt are further divided based on calcium carbonate content or soil subtype. In addition to soil texture calcium carbonate content was used as input variable to determine SWR at -150000 kPa. For the prediction of AWC soil subtype is the first splitting variable and texture forms the second level of the classification tree.

Continuous pedotransfer functions (point PTFs)

Using the stepwise method, models with significant statistical reliability were developed for predicting the soil water retention at the different matric potential. Table 4 presents pedotransfer functions (MLRs) and their statistical values, respectively.

Clay, sand and calcium carbonate content are included in the pedotransfer functions to predict SWR at all four matric potentials. Organic matter content and pH are important to predict the SWR at -0.1 kPa. Soluble salt content is used as input parameter in the model of SWR at -33 kPa.

Comparison of the different methods

Continuous pedotransfer functions (MLR) established for the studied salt-affected soils provide slightly better predictions than regression trees using the same input parameters (CRT1) during reliability assessment. The RMSE values of MLRs were lower in every case with a maximum difference of 0.4 vol% and Pearson correlation coefficient was generally higher with 10%. However, prediction of the SWR considering the soil subtype (CRT2, CRT_kat and CHAID) performed slightly better both in terms of accuracy and reliability in case of SWR at -0.1, -33 kPa and calculated AWC (Table 2).

Comparing the CRT2 method with CRT1 at -33 kPa, in case of containing Hungarian soil subtype in the model, RMSE decreased by 1.139 vol% and the Pearson correlation coefficient increased by 26% during the test of reliability (Table 2). Rawls et al. (2003) also analysed the importance of soil taxonomical information for the prediction of SWR. They found that including soil taxonomic order besides textural information, improves the estimation of SWR at -33 and -1500 kPa. In case of using organic carbon as well, additional soil taxonomic information improves less the estimation.

The mean absolute relative error (MARE) gives information about the relative size of the estimation error, which we think is useful especially for scientist working in other disciplines. For the comparison of the accuracy of pedotransfer functions developed on different datasets this type of error related performance metrics can be useful.

Using continuous data (CRT2) instead of categorical data (CRT_kat) gave no significantly different results in predicting SWR. However for the test sample among all developed methods, the prediction based on categorical soil properties and soil subtype (CRT_kat and CHAID) showed the highest reliability (Table 2), but the difference between the methods was not significant. The RMSE values of models using category-type input parameters varied between 1.249 and 5.163 vol%, the Pearson correlation coefficients at the different matric potential were between 0.643 and 0.782. The findings that continuous-type input parameters can result in similar estimation reliability as continuous-type one is in accordance with the results of Lilly et al. (2008). They showed that for the estimation of soil hydraulic conductivity regression trees using field based information (soil horizon, ped size and soil texture – category-type variables) gives similar results to regression trees based on information from laboratory measurements (particle size distribution, bulk density, organic matter – continuous-type variables). There is no significant difference in reliability – based on the mean squared error and absolute relative error of the samples – between the regression tree (CRT_kat) and CHAID method at any of the studied matric potentials. The biggest difference between the two methods can be observed at -0.1 kPa, where the RMSE value of the CHAID method is smaller with 0.27 vol%. However, in case of applying exactly the same optimization process in both methods, the structure of the CHAID model is simpler than that of regression tree (CRT_kat), having less terminal nodes (Table 3).

Considering that the reliability of the developed pedotransfer rules and functions are similar, the applicability of both methods primarily is based on the type of available soil information and then on the purpose of the SWR prediction. In the case of continuous-type input parameters, pedotransfer function developed by linear regression is advisable because of its reliability. If the aim is to study the relationship between the easily available soil characteristics and SWR, regression tree is more favorable. Having categorical (ordinal and nominal) type available information we suggest the use of the CHAID method rather than the regression tree if the goal is reliability and also if the interpretation of the results are important.

Relationship between SWR and soil characteristics

Apart from the particle size distribution calcium carbonate, soluble salt content and pH were found to be important predicting properties. Although Khodaverdiloo and Homaee (2004), Rajkai (1988) and Hodnett and Tomasella (2002) showed that in case of “problematic soils” they are important influencing properties of soil water retention, these soil properties are rarely used to predict it. Soluble salts have an effect on the structure of the soil which determines the quantity of macropores. It is an important influencing variable at the highest matric potentials of the SWR curve where macropores have the biggest role in determining soil water retention capacity. Higher salt content (> 0.25%) resulted in lower mean SWR in case of clayey soils at -0.1 kPa (Fig. 2) matric potential. This can be explained by the peptisation effect of the salts which leads to the decrease of macropore volume especially when the texture of the soil is clayey (Várallyay, 2002; Lima et al., 1990).

At the lower matric potential range (at -1500 and -150000 kPa) soils with higher calcium carbonate content always have lower SWR. At these matric potentials adhesion forces dominate over surface tension and the quantity of micropores (<0.2 μm in diameter) and the surface area of the soil particles determines the quantity of the retained water content (Moore, 2004). With increasing calcium carbonate content the quantity of mineral soil particles decreases in a unit volume (Rajkai personal communication, 2011) therefore SWR decreases. When the soil contains calcaric material which is not colloidal, but present in crystallized form (e.g. lime concentration), it further lowers the SWR (Marshall et al., 1996) even at -0.1 kPa (Fig. 2).

Hodnett and Tomasella (2002) hypothesized pH being the indicator of the degree of weathering which has affect on the clay mineralogy and soil structure. In our study it can be assumed that through pH, some indirect information can be obtained about the exchangeable sodium content and the structure of the soil which influences the soil water retention. High pH might be caused by higher sodium content, which increases the amount of water that can be hold by the soil in the lower matric potential range, probably due to the greater adhesion forces of sodium rich earth fraction and changes of pore size distribution. High sodium saturation increases hydration and dispersion of soil colloids, therefore micropore volume increases, which increases water retention at -1500 kPa and lower matric potentials (Várallyay, 2002) where particle surface properties dominate over capillary forces. At -0.1 kPa pH higher than 9 might indicate less favourable hard blocky structure with low gravitational pore volume, which decreases the SWR near saturation (Fig. 2).

Introducing soil taxonomical information into the pedotransfer rules can improve both their estimation accuracy and reliability, especially at -33, -1500 kPa and AWC (Table 2) as it was discussed above. The name of the soil - independently from the type of soil taxonomy - always contains important complex information about the soil physical, hydrological, chemical and biological properties. In contrast to pedotransfer functions, classification methods like CHAID can handle such valuable information too.

The importance of organic matter content is mainly due to its effect on bulk density which is important in the high matric potential range. Organic matter content determines the extent of the organic colloid surface (Rajkai, 1988) and has effect on the structure and the adsorptive properties of the soil (Rawls et al., 2003; Lal and Shukla, 2004). Higher organic matter content results in higher water retention (Fig. 2) except for a few samples which must be further investigated.

Conclusions

The prediction method by the classification tree based on categorical data (CHAID) gives similar results compared to that of prediction methods based on continuous variables. Based on the categorical-type soil information available from Hungarian detailed soil maps SWR at -0.1, -33, -1500 and -150000 kPa could be predicted with reasonable RMSE values for the test sample. The developed pedotransfer rules are readily applicable to predict the SWR based on the available large scale Hungarian soil maps which contain these categorical-type information. The classification tree methods (regression tree and CHAID) were found to be helpful to model the complex relationship between soil water retention and other soil properties of salt affected soils. However, if continuous input parameters are available, for practical applications we suggest using multiple linear regressions, because they need less computation time and provide similar reliability as the regression trees. If the available information is category-type (nominal and=or ordinal), the CHAID method results in a simpler prediction model than regression tree does.

Analyses using both the classification trees and the multiple linear regression method showed that, in addition to soil texture and organic matter content, soil pH and calcium carbonate content influence the water retention of salt-affected soils considerably. In our study, soil genetic information (expressed through taxonomic classes) improved the prediction reliability of the SWR the most at -33 kPa (RMSE reduced by 1.139 vol%, r increased by 26%); however, the difference was not significant. Nevertheless, this information may be utilized in the hydropedological assessment of salt affected soils.

References

Acharya, C.L. and Abrol, I.P. 1978. Exchangeable sodium and soil-water behaviour under field conditions. Soil Sci. 125. 310-319.

Al Majou, H., Bruand A., Duval O. and Cousin I. 2007. Variation of the water-retention properties of soils: Validity of class-pedotransfer functions. CR Geoscience, 339. (9) 632-639.

Baranyai, F. (ed.) 1989. Guide to Large Scale Soil Mapping. (Útmutató a nagyméretarányú talajtérképezés végrehajtásához.) (in Hungarian), Agroinform. Budapest. pp147.

Børgesen, C.D. and Schaap, M.G. 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. Geoderma. 127. 154–167.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J., 1998. Classification and regression trees. (reprint) CRC Press. Florida. pp. 358.

Chaudhari S. K. and Somawanshi R. B. 2000. Effect of water quality on moisture retention characteristics of different texture soils. Journal of Maharashtra Agricultural Universities. 25. (1-3) 128-133.

Dane, J.H. and Klute, A. 1977. Salt effects on the hydraulic conductivity of a swelling soil. Soil Science Society of America Journal. 41. 1043-1049.

Gee, G. W. and Bauder, J. W. 1986. Particle-size analysis. In: Klute, A. (Eds.) Methods of Soil AnalysisAmerican Society of Agronomy, Inc. Soil Science Society of America, Inc. Madison, Wisconsin. USA. Part 1, p383–412.

Groenevelt, P.H., Grant, C.D. and Murray, R.S. 2004. On water availability in saline soils. Australian Journal of Soil Research, 42. 833-840.

Hill, T., and Lewicki, P. 2006. Statistics: methods and applications. A comprehensive reference for science, industry and data mining. Statsoft. Tulsa, OK. pp. 832.

Hodnett, M.G. and Tomasella, J. 2002. Marked differences between van Genuchten soil water retention parameters for temperate and tropical soils: a new pedotransfer functions developed for tropical soils. Geoderma 108. 155-180.

IBM 2011. IBM SPSS Statistics Information Center. http://publib.boulder.ibm.com/infocenter/spssstat/v20r0m0/index.jsp last accessed: 2011 November

Jayawardane, N.S. and Beattie, J.A. 1978. Effect of salt solution composition on moisture release curves of soil. Australian Journal of Soil Research. 17. 89-99.

Kass, G. V., 1980. An exploratory technique for investigating large quantities of categorical data. Applied Statistics. 29. (2) 119-127.

Khodaverdiloo, H,. and M. Homaee., 2004. Pedotransfer Functions of some Calcareous Soils. In: Wöhrle, N. and Scheurer, M.: EUROSOIL 2004. Abstracts and Full Papers. September, 4–12 Freiburg,Germany. 10 (27) 1-11.

Lal, R., Shukla, M.K., 2004. Principles of soil physics. Marcel Dekker, Inc. New York, Basel. pp716.

Lilly, A., Nemes, A., Rawls, W. J. and Pachevsky, Ya. A. 2008. Probabilistic Approach to the Identification of Input Variables to Estimate Hydraulic Conductivity. Soil Sci. Soc. Am. J. 72. 16-24

Lima, L. A., M. E. Grismer, and D. R. Nielsen. 1990. Salinity effects on Yolo Loam hydraulic properties. Soil Science. 150. 451-458.

Lipsius, K. 2002. Estimating available water capacity from basic soil physical properties. A comparison of common pedotransfer functions. Studienarbeit. pp. 38.

Makó, A., Tóth, B., Hernádi, H., Farkas, Cs., and Marth, P., 2010. Introduction of the Hungarian Detailed Soil Hydrophysical Database (MARTHA) and its use to test external pedotransfer functions., Agrokémia és Talajtan. 59. 29-39.

Marshall, T.J., Holmes, J.W. and Rose, C.W., 1996. Soil physics. Cambridge Univ. Press. pp. 453.

McKenzie NJ, Jacquier DW (1997). Improving the field estimation of saturated hydraulic conductivity in soil survey. Australian Journal of Soil Research 35. 4. 803-827.

McLean, E. O., 1982. Soil pH and lime requirement. In: Methods of soil analysis. Part 2. (ed. by A. L. Page) American Society of Agronomy, Inc. Soil Science Society of America, Inc. Madison, Wisconsin. USA. 199–224.

Moore, G. (Ed.) 2004. Soil Guide - A Handbook for Understanding and Managing Agricultural Soils. Chapter 3. Physical factors affecting water infiltration and redistribution. Department of Agriculture, Western Australia. 53-108.

MSZ 1978. Determination of total water-soluble salt content. (Vízben oldható összes sótartalom meghatározása). Hungarian Standard no. MSZ 08-0206-2:1978. Hungarian Standards Institution. Budapest (in Hungarian)

Minasny, B., McBratney, A., Bristow, K.L. 1999. Comparision 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: Page, A. L., Miller, R.H., Keeny, D.R. (Eds.) Methods of soil analysis. Part 2., American Society of Agronomy, Inc. Soil Science Society of America, Inc. Madison, Wisconsin. USA. 181–197.

Nemes, A. 2003. Multi-scale hydraulic pedotransfer functions for Hungarian soils. Ph.D. diss. Wageningen Agric. Univ., Wageningen, the Netherlands.

Nemes, A., and Rawls, W. J. 2006. Evaluation of different representations of the particle-size distribution to predict soil water retention. Geoderma. 132. 1-2. 47-58.

Pachepsky, Y.A., Shcherbakov, R. A, Várallyay, Gy., and Rajkai, K. 1982. Soil water retention as related to other soil physical properties. Pochvovedenie. 2. 42–52.

Pachepsky, Y. A., Timlin, D. and Várallyay, Gy., 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60. 727–733.

Rajkai, K., 1988. The Relationship Between Water Retention and Different Soil Properties. (A talaj víztartó képessége és különböző talajtulajdonságok összefüggésének vizsgálata.) (in Hungarian) Agrokémia és Talajtan. 36–37. 15–30.

Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M., and Bloodworth, H. 2003. Effect of soil organic carbon on soil water retention. Geoderma. 116. 61–76.

Rawls, W. J., Brakensiek, C. L., and Saxton, K. E. 1982. Estimation of soil water properties. Transactions American Society of Agricultural Engineers. 25. 1316–1320.

Saxton, K.E., Rawls, W.J., Romberger, J.S. and Papendick, R.I. 1986. Estimating generalized soil-water characteristics from texture. Soil Science Society of America Journal. 50. 1031-1036.

Saxton, K.E., and Rawls, W.J. 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. Soil Science Society of America Journal. 70. 1569-1578.

SPSS for Windows, Rel. 11.0.1. 2001. SPSS Inc. Chicago.

Tóth, B., Makó, A., Rajkai, K., and Marth, P. 2005. The estimation possibilities of the soil water retention capacity on the basis of available soil map information. 13th International Poster Day: Transport of Water, Chemicals and Energy in the System Soil-Crop Canopy-Atmosphere. Bratislava. 10.11.2005. 557-562.

Tóth, B., Makó, A., Guadagnini, L., Guadagnini, A. 2008. Factor analysis of Hungarian hydrophysical data to predict soil water retention characteristics. Cereal Research Communication. 36. 411-414.

Tóth, G., and Máté, F. 2006. Notes on the development of a national soil information system. (Megjegyzések egy országos, átnézetes, térbeli talajinformációs rendszer kiépítéséhez.) (in Hungarian) Agrokémia és Talajtan. 55 2. 473-478.

Twarakavi, N.K.C., Šimůnek, J. and Schaap, M. G. 2009.Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machine. Soil Science Society of America Journal. 73. 5. 1443-1452.

Tyurin, I. V., 1931. A new modification of the volumetric method of determining soil organic matter by means of chromic acid. Pochvovedenie (in Russian). 26. 36–47.

Vanapalli S.K., Sillers W.S., and Fredlund M.D. 1998. The meaning and relevance of residual state to unsaturated soils. 51st Canadian geotechnical conference. Edmonton, Albert, October 4-7. 1-8.

Várallyay, G. 1973. A new apparatus for the determination of soil moisture potential in the low suction range. (A talaj nedvességpotenciálja és új berendezés annak meghatározására az alacsony (atmoszféra alatti) tenziótartományban.) (in Hungarian) Agrokémia és Talajtan. 22. 1-2. 1-22.

Várallyay, Gy. 2002. Environmental stresses induced by salinity/alkalinity in the Carpathian basin (Central Europe). Agrokémia és Talajtan. 51. 1-2. 233-242.

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.

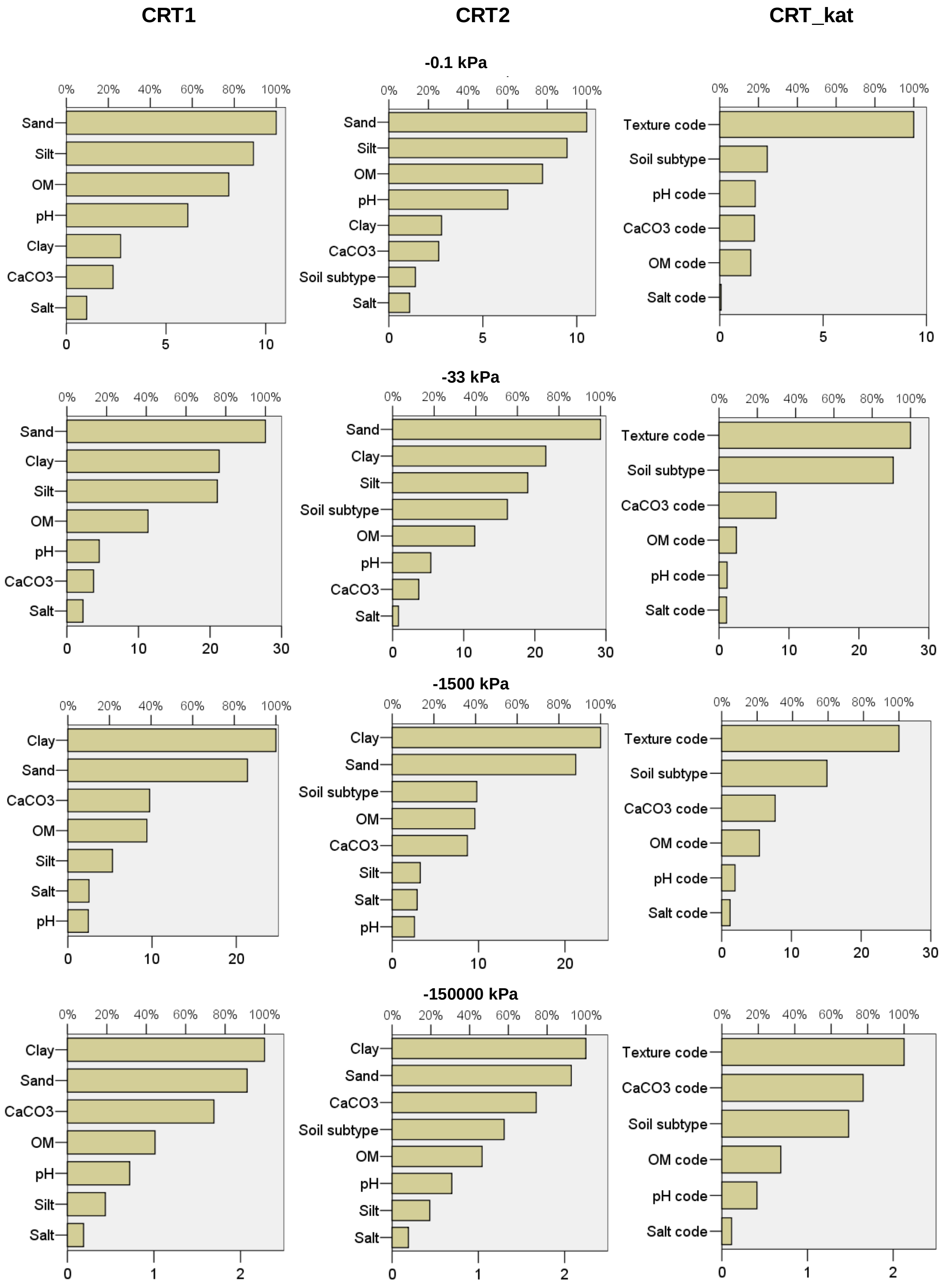

Fig. 1. Variable importance – also in normalized form (%) – calculated for each input parameters to predict SWR at -0.1, -33, -1500 and -150000 kPa matric potential with regression tree methods (CRT1, CRT2, CRT_kat). Description of the input variables: clay (<0.002 mm) (mass %), silt (0.002-0.05 mm) (mass %), sand (0.05 – 2 mm) (mass %), OM: organic matter (mass %), calcium carbonate content (mass %), pH_{H2O}, soluble salt content (mass %). Intervals of the soil properties' codes and list of the studied salt-affected Hungarian soil subtypes can be seen in the appendices (App. 2-7.) of the on-line publication of this paper.

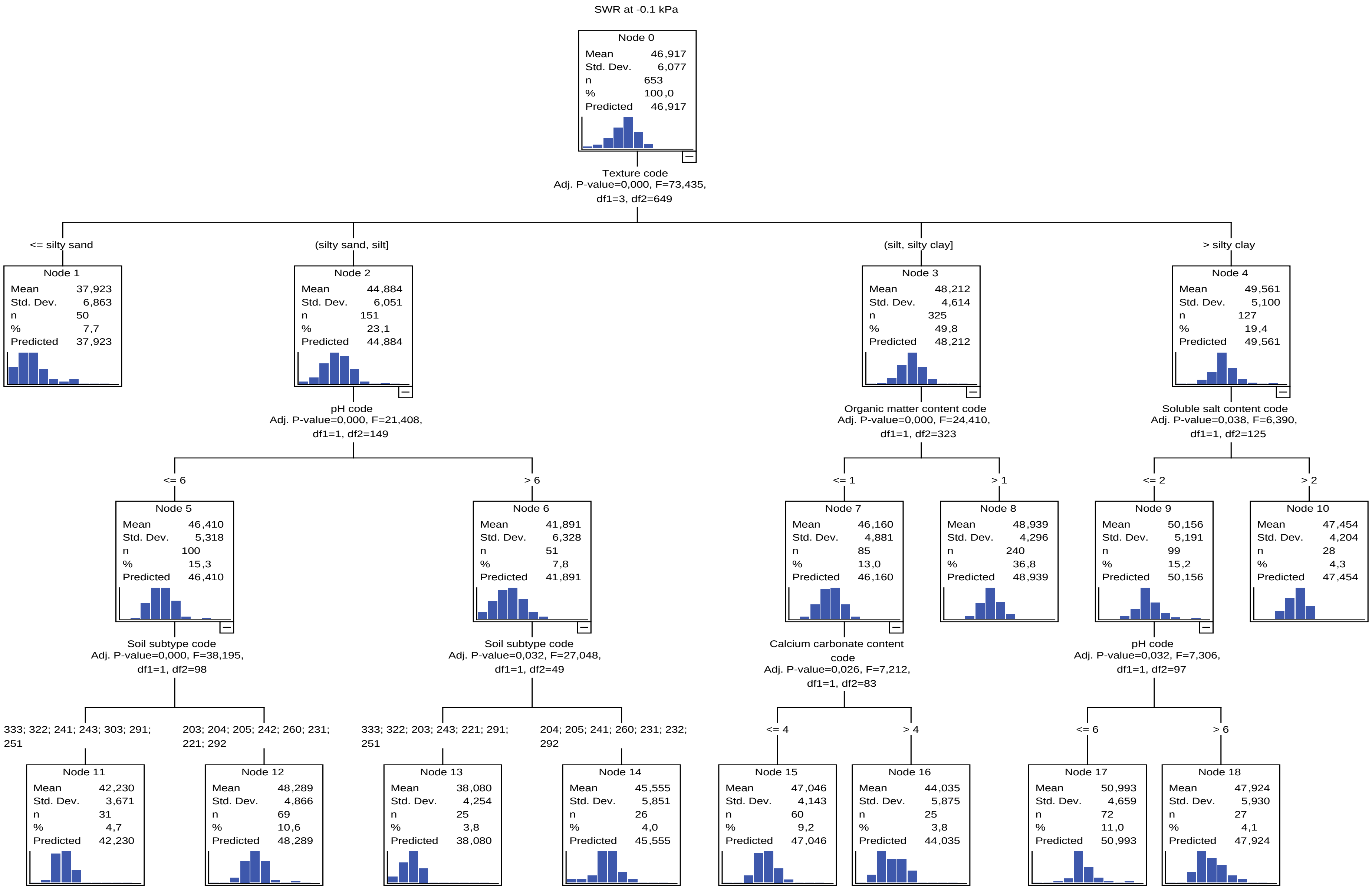

Fig. 2. Tree diagram resulting from the CHAID classification tree method to predict SWR at -0.1 kPa. Intervals of the soil properties' codes and meaning of soil subtype codes can be seen in the appendixes (App. 2-7.). Round bracket means that the subsequent category is not part of the group, category having a squared bracket behind its name is part of the group.

Table 1. Descriptive statistic of the training and test datasets.

Soil parameters	Type of dataset	N	Minimum	Maximum	Mean	SD[a]
Clay content (<0.002 mm) (mass %)	training	653	2.1	67.8	32.4	14.4
	test	76	2.9	60.7	31.6	13.1
Silt content (0.002-0.05 mm) (mass %)	training	653	1.1	74.4	43.4	14.1
	test	76	6.9	74.4	42.7	14.5
Sand content (0.05 – 2 mm) (mass %)	training	653	0.3	95.8	24.3	20.4
	test	76	1.4	90.2	25.7	21.0
Organic matter content (mass %)	training	653	0.1	4.7	1.6	1.2
	test	76	0.1	4.2	1.6	1.1
Calcium carbonate content (mass %)	training	653	0.0	80.0	15.6	13.6
	test	76	0.0	63.0	16.6	12.6
pH_{H2O}	training	653	6.2	10.6	8.6	0.8
	test	76	7.4	10.4	8.6	0.7
Soluble salt content (mass %)	training	639	0.0	1.2	0.1	0.2
	test	75	0.0	0.5	0.1	0.1
Soil water content at -0.1 kpa (vol %)	training	653	26.4	75.6	46.9	6.1
	test	76	33.5	60.3	46.4	5.5
Soil water content at -33 kpa (vol %)	training	653	4.6	63.7	33.1	8.1
	test	76	14.6	53.1	33.6	7.6
Soil water content at -1500 kpa (vol %)	training	653	0.4	37.5	19.7	7.1
	test	76	2.5	38.0	20.8	7.5
Soil water content at -150000 kpa (vol %)	training	653	0.2	9.9	3.7	2.2
	test	76	0.3	8.1	3.7	2.0

[a]Standard deviation

Table 2. Reliability of the estimation of the soil water retention with CRT (regression tree), CHAID classification tree method and multiple linear regression (MLR) at -0.1, -33, -1500 and -150000 kPa matric potential, in case of categorical and continuous independent variables.

		Test dataset			
Method	Predicted soil water content	RMSE (vol%)	MARE (%)	Pearson correlation coefficient	Number of samples
CRT1 (continuous soil properties)	$\theta_{-0.1kPa}$	4.690	8.34	0.541**	76
	θ_{-33kPa}	6.403	14.73	0.578**	76
	$\theta_{-1500kPa}$	5.433	19.99	0.711**	76
	$\theta_{-150000kPa}$	1.369	34.34	0.730**	76
	AWC predicted	4.503	38.51	0.186	76
	AWC calculated[b]	4.826	37.65	0.284**	76
CRT2 (continuous soil properties + soil subtype)	$\theta_{-0.1kPa}$	4.767	8.41	0.522**	76
	θ_{-33kPa}	5.264	11.82	0.731**	76
	$\theta_{-1500kPa}$	5.266	21.65	0.719**	76
	$\theta_{-150000kPa}$	1.509	36.94	0.671**	76
	AWC predicted	4.442	34.79	0.236*	76
	AWC calculated[b]	4.604	33.66	0.427**	76
CRT_kat (categorical soil properties + soil subtype input)	$\theta_{-0.1kPa}$	4.488	7.61	0.588**	76
	θ_{-33kPa}	4.935	11.81	0.772**	76
	$\theta_{-1500kPa}$	5.117	22.90	0.738**	76
	$\theta_{-150000kPa}$	1.249	29.32	0.787**	76
	AWC predicted	4.327	35.19	0.228*	76
	AWC calculated[b]	4.734	37.46	0.181	76
CHAID (categorical soil properties + soil subtype input)	$\theta_{-0.1kPa}$	4.222	7.10	0.643**	76
	θ_{-33kPa}	5.040	12.33	0.760**	76
	$\theta_{-1500kPa}$	5.163	23.37	0.740**	76
	$\theta_{-150000kPa}$	1.265	31.03	0.782**	76
	AWC predicted	4.493	36.03	0.091	76
	AWC calculated[b]	4.580	37.42	0.185	76
MLR (continuous soil properties)	$\theta_{-0.1kPa}$	4.684	7.56	0.531**	76
	θ_{-33kPa}	5.940	14.32	0.641**	75[c]
	$\theta_{-1500kPa}$	5.274	23.43	0.737**	76
	$\theta_{-150000kPa}$	1.350	29.25	0.764**	76
	AWC predicted	4.056	35.27	0.368**	75[c]
	AWC calculated[b]	4.402	37.74	0.211	75[c]

*Correlation is significant at 0.05 level.

** Correlation is significant at 0.01 level.

[a] Average RMSE of the 10 times tenfold crossvalidated test samples.

[b] The plant available water content calculated from predicted SWR values

[c] In case of SWR at -33 kPa and AWC predicted the PTF was worked out for 639 and tested on 75 samples since the soluble salt content of the soil was included as input variable in the multiple linear regression and those samples had measured soluble salt content information.

Table 3. Summary of pedotransfer rules developed by regression tree (CRT) and CHAID classification tree methods.

Type of prediction	Predicted SWR	Soil properties which split the data in CRT and CHAID models	Number of terminal nodes
CRT1	-0.1 kPa	Organic matter, silt, sand, soluble salt, calcium carbonate content, pH_{H2O}	10
	-33 kPa	Sand, clay, calcium carbonate, silt content	10
	-1500 kPa	Clay, sand, calcium carbonate, silt, organic matter content	12
	-150000 kPa	Clay, organic matter, calcium carbonate, silt, sand content, pH_{H2O}	11
	AWC	Silt content, pH_{H2O}, organic matter, clay content	7
CRT2	-0.1 kPa	Organic matter, silt content, soil subtype, soluble salt content	8
	-33 kPa	Sand content, soil subtype, pH_{H2O}, clay, calcium carbonate, silt content	14
	-1500 kPa	Clay, sand content, soil subtype, calcium carbonate, organic matter content	12
	-150000 kPa	Clay, organic matter, calcium carbonate content, soil subtype, silt, soluble salt, sand content	16
	AWC	Soil subtype, soluble salt, silt, sand, organic matter, clay content	9
CRT_kat	-0.1 kPa	Texture, soil subtype, pH_{H2O}, organic matter, calcium carbonate content code	15
	-33 kPa	Texture, soil subtype, pH_{H2O}, organic matter content code	13
	-1500 kPa	Texture, soil subtype, organic matter, calcium carbonate content, soluble salt content, pH_{H2O} code	22
	-150000 kPa	Texture, soil subtype, organic matter, calcium carbonate content code	20
	AWC	Soil subtype, pH_{H2O}, organic matter content, texture code	7
CHAID	-0.1 kPa	Texture, pH_{H2O}, organic matter, soluble salt content, soil subtype, calcium carbonate content code	11
	-33 kPa	Texture, soil subtype code	10
	-1500 kPa	Texture, calcium carbonate content, soil subtype code	9
	-150000 kPa	Texture, calcium carbonate content code	11
	AWC	Soil subtype, texture code	5

CHAID, CRT_kat: classification tree with category type independent variables and Hungarian soil subtype. The intervals of the soil properties' categories can be found in the appendixes (App. 1-5.). CRT1: regression tree based on soil properties similar to the input variables of the first prediction method (CHAID) but in continuous form and without Hungarian soil subtype: clay (<0.002 mm) (mass %), silt (0.002-0.05 mm) (mass %), sand (0.05 – 2 mm) (mass %), organic matter (mass %), calcium carbonate (mass %), pH_{H2O}, soluble salt content (mass %). CRT2: regression tree with the input parameters of the CRT1 method and the Hungarian soil subtype.

Table 4. Continuous pedotransfer functions to predict soil water content at -0.1, -33, -1500 and -150000 kPa matric potentials and plant available water content (with statistical significance).

Pedotransfer function based on multiple linear regression models	Significance of the model predictions		
	F	p	Adjusted R^2
$\Theta_{-0.1} = 11.559 - 0.319 \cdot sand + 1.103 \cdot \ln(OM) - 0.753 \cdot pH^2 + 24.057 \cdot clay^{-1} + 12.410 \cdot pH - 0.001 \cdot CaCO_3^2 - 0.005 \cdot clay \cdot silt - 0.001 \cdot silt^2$	$F_{8,630} = 39.601$	$p < 0.0005$	0.326
$\Theta_{-33} = 44.479 - 0.302 \cdot sand + 0.003 \cdot clay^2 - 0.002 \cdot CaCO_3^2 - 0.005 \cdot clay \cdot silt + 4.401 \cdot salt$	$F_{5,633} = 102.293$	$p < 0.0005$	0.443
$\Theta_{-1500} = 18.738 + 0.346 \cdot clay - 0.002 \cdot CaCO_3^2 - 0.001 \cdot sand^2 - 2.513 \cdot \ln(clay)$	$F_{4,634} = 179.892$	$p < 0.0005$	0.529
$\Theta_{-150000} = 2.539 + 0.081 \cdot clay - 0.049 \cdot CaCO_3 - 0.011 \cdot sand - 0.875 \cdot sand^{-1} - 0.001 \cdot clay \cdot sand$	$F_{5,633} = 193.896$	$p < 0.0005$	0.602
$AWC = 15.584 + 0.080 \cdot silt - 1.352 \cdot \ln(OM) - 0.077 \cdot pH^2 + 4.109 \cdot salt^2$	$F_{4,634} = 13.453$	$p < 0.0005$	0.078

Input variables correspond to soil properties available from soil maps. Description of the input variables: clay (<0.002 mm) (mass %), silt (0.002-0.05 mm) (mass %), sand (0.05 – 2 mm) (mass %), organic matter content (mass %), calcium carbonate content (mass %), pH_{H2O}, soluble salt content (mass %).