Hydropedology and pedotransfer functions

Y.A. Pachepsky, W.J. Rawls, H.S. Lin

Available online 5 May 2005

Abstract

The emerging interdisciplinary research field of hydropedology attracts a substantial attention because of its promise to bridging pedology and hydrology. Pedotransfer functions (PTFs) emerged as relationships between soil hydraulic parameters and the easier measurable properties usually available from soil survey. One hypothetical explanation of current PTF shortcomings is that PTF inputs do not describe the structure of pore space per se and, therefore, do not represent relationships between structure and function of soil pore space. A possible direction for improvement is to look for PTF predictors that are better related to the structure of water-bearing pathways, in particular using the pedological soil structure description. The objective of this work was to develop and discuss an example of pedotransfer function relating soil structure and soil hydrologic parameters. We used the subset of 2149 samples from the US National Soil Characterization database that had values of water contents at $\theta_33$ kPa and bulk densities on clods, structure characterized with grade, size and shape, textural class determined in the field and from lab textural analysis. Classification and regression trees were used to group soil samples according to their water contents at $\theta_33$ kPa. The clay class was the best grouping parameter in all but loamy sand textural classes. The structural parameters served as important grouping variables to define groups of soil samples with distinctly different average water retention for the groups. Defining and quantifying soil structure at various scales, including pedon, hillslope and watershed scales, may contribute for the development scale-relevant PTFs at those scales.

© 2005 Elsevier B.V. All rights reserved.

1. Introduction

The emerging interdisciplinary research field of hydropedology attracts a substantial attention because of its promise to bridging pedology and hydrology. Such interaction is desirable because the wealth of pedological information can advance understanding and predicting water distribution in soils and landscapes, whereas advances of hydrology can enrich interpretation of soil properties.

One possible approach to the hydropedology agenda is to consider it from the standpoint of
relations between structure and function. Hydrologic functioning of soils and landscapes is defined by the structure of pathways and voids available for water to move and to be stored. In turn, structure of soil pore space is substantially affected by the functioning of soils and landscapes in hydrological cycles. This relationship has a multitude of feedbacks that modify the function according to changes in structure, and vice versa. In particular, both ecological changes and changes in management are known to alter both soil structure and its hydrologic functioning. Pedology is strong in providing information about structure of soil and soil cover whereas hydrology renders rich information about soil hydraulic functioning.

Relationships between structure and function are revealed and studied in many disciplines, e.g. plant science, molecular biology, sociology, just to name a few. A general trend of such research is to quantify the relation between structure and function by expressing this relation in form of an empirical or mechanistic model.

Pedotransfer functions emerged as relationships between soil hydraulic parameters and the easier measurable properties usually available from soil survey (Bouma, 1989). Utility of pedotransfer functions was recognized immediately because of multiple uses of soil hydraulic properties. For example, soil water retention and transport parameters are used in hydrology to partition precipitation into runoff and infiltration and to assess evapotranspiration. In agronomy, the same data are used to schedule management practices, especially irrigation and chemical application. In meteorology, surface soil moisture is needed to establish components of the heat balance. In contaminant hydrology and geochemistry, estimates of hydraulic properties in vadose zone provide an essential precondition of estimating contaminant transport (Rawls et al., 1991). Measurements of soil hydraulic properties are relatively time-consuming and become impractical when hydrologic estimates are needed for large areas. Estimating water retention from basic soil data available from soil survey becomes an alternative to measurements in many applications (Van Genuchten and Leij, 1992; Timlin et al., 1996; Pachepsky et al., 1999). Comprehensive reviews of the status of pedotransfer functions have been published recently (Wösten et al., 2001; McBratney et al., 2002; Pachepsky and Rawls, 2004).

As the use and development of pedotransfer functions (PTFs) progressed, several problems became obvious and were articulated. First, the PTF accuracy remained limited in spite of adding potentially useful predictors and using sophisticated tools of data mining with artificial intelligence and machine learning. Second, the portability of PTFs remained limited; PTFs developed in one region or from one database had limited applicability in other conditions (e.g. Williams et al., 1992; Tietje and Tapkenhinrichs, 1993; Kern, 1995; Wösten et al., 2001).

One hypothetical explanation of PTF shortcomings is that PTF does not describe the structure of pore space per se and therefore, does not represent relationships between structure and function well enough. Typical PTF inputs, such as soil texture, bulk density, or organic carbon content, are related to the pore structure in a broad sense, but are not sufficient to characterize the pore structure of a specific soil. There are indirect confirmations of this hypothesis. For example, excellent estimates of soil hydraulic conductivity were obtained when void sizes have been measured directly (Anderson and Bouma, 1973). Estimation of water retention has been substantially improved when one or more points on soil water retention curve have been added to the list of PTF predictors (Ahuja et al., 1985). The latter happened probably because water retention curve provides more information about soil pore structure than texture and bulk density.

Measurement and characterization of soil pore space remains limited in its capabilities, although some progress based on tomography has been achieved (i.e., Mooney, 2002). Therefore, one of possible directions is to look for PTF predictors that are better related to the structure of water-bearing pathways than traditionally used texture and bulk density. One of possibilities is using the pedological soil structure description. This also may have drawbacks because (a) soil structure is described in qualitative rather than quantitative terms, and (b) structure characterization is usually done at the scale that is too coarse to reveal arrangement of fine pores that retain water at low soil matric potential. An attempt to use the soil structure descriptors in
the water retention PTFs has shown some improvement in the PTF accuracy (Rawls and Pachepsky, 2002a).

Soil structure is characterized with categorical variables. Classes or categories, like weak, moderate, and strong for the grade, are set and the class or category for each soil sample is recorded. Categorical data on structure cannot be directly used in statistical regressions or neural networks to estimate water retention from other soil properties. Recently the method of classification and regression trees (CART) was recognized as a suitable statistical technique for using categorical variables as predictors (Clark and Pregibon, 1992). Regression trees were successfully used to explore databases in natural sciences (Fielding, 1999), and, in particular, in soil science (McKenzie and Jacquier, 1997; O’Connell and Ryan, 2002; Park and Vlek, 2002).

The objective of this work was to develop and discuss an example of pedotransfer function relating soil structure and soil hydrologic parameters.

2. Materials and methods

2.1. Soil dataset

We used the subset of 2149 samples from the US National Soil Characterization database (Soil Survey Staff, 1997) that had (a) values of water contents at \( \theta_{33} \) and \( \theta_{1500} \) kPa on clods and bulk densities at 33 kPa and of the air dry soil, (b) structure characterized with grade, size and shape, and (c) textural class determined in the field and from lab textural analysis, all measured and described in the same pedon. Thirty percent of all samples in that data set belonged to pedons that did not have a taxonomic family phrase. Mollisols, Aridisols, Alfisols, and Entisols were the most numerous among soils with known taxonomy in the data set, and constituted 24%, 14%, 11%, and 6%, respectively. About half of all samples came from California, Colorado, Idaho, Kansas, New Mexico, Texas, and Washington. The major field-determined textural class in the data set was silt loam found in about 24% of all samples (Table 1). Sandy loam, loam, clay, and silty clay loam were represented with 15%, 12%, 12%, and 10% of all samples, respectively. Silt and sandy clay were each represented with less than 0.5% of all samples, sands and loamy sands were each about 3% of all samples. Values of volumetric water contents at \( \theta_{33} \), \( \theta_{33} \), and \( \theta_{1500} \), were obtained as products of gravimetric water contents on the corresponding bulk density.

Field structure was defined according the USDA Soil Survey Manual (Soil Survey Staff, 1997). In brief, soil structural units are defined as repetitive soil bodies that are commonly bounded by planes or zones of weakness that are not apparent consequences of compositional differences.” Shapes of compositional units are classified into (a) the units

<table>
<thead>
<tr>
<th>Textural class from laboratory data</th>
<th>Sand</th>
<th>Loamy sand</th>
<th>Sandy loam</th>
<th>Loam</th>
<th>Silt loam</th>
<th>Silt</th>
<th>Sandy clay loam</th>
<th>Clay loam</th>
<th>Silty clay loam</th>
<th>Sandy clay</th>
<th>Silty clay</th>
<th>Clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count: 70</td>
<td>67</td>
<td>321</td>
<td>259</td>
<td>511</td>
<td>6</td>
<td>76</td>
<td>200</td>
<td>221</td>
<td>8</td>
<td>149</td>
<td>261</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Textural class from field data</th>
<th>Sand</th>
<th>Loamy sand</th>
<th>Sandy loam</th>
<th>Loam</th>
<th>Silt loam</th>
<th>Silt</th>
<th>Sandy clay loam</th>
<th>Clay loam</th>
<th>Silty clay loam</th>
<th>Sandy clay</th>
<th>Silty clay</th>
<th>Clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count: 70</td>
<td>67</td>
<td>321</td>
<td>259</td>
<td>511</td>
<td>6</td>
<td>76</td>
<td>200</td>
<td>221</td>
<td>8</td>
<td>149</td>
<td>261</td>
<td></td>
</tr>
</tbody>
</table>
are flat, platelike and are generally oriented horizontally; lenticular structure, is recognized for plates that are thickest in the middle and thin toward the edges; (b) prismatic, that are usually longer vertically, (c) columnar that similar to prisms but have tops very distinct and normally rounded, (d) blocky that are nearly equidimensional but grade to plates, (e) granular, or crumb, where units are approximately spherical or polyhedral and are bounded by curved or very irregular faces. Size of structural units is divided into five classes (very fine, fine, medium, coarse, and very coarse); size class boundaries depend on the shape class. For the purposes of this work, three size classes were defined by combining very fine and fine, and coarse and very coarse. Grade describes the distinctness of units. Criteria are the ease of separation into discrete units and the proportion of units that hold together when the soil is handled. Four classes are used: (a) structureless, where no observable aggregation or no definite and orderly arrangement of natural lines of weaknesses occurs, (b) weak, when the units are barely observable in place, and, when gently disturbed, soil material does not exhibit no planes of weakness (c) moderate when units are well formed and evident in the undisturbed soil; when soil material is disturbed, peds part from adjoining peds to reveal nearly entire faces, (d) strong when units are distinct in undisturbed soil and, when disturbed, the soil material separates mainly in the whole units.

Fig. 1 shows distributions of structural properties among samples in the data set. The weak and moderate grades are the most common in the data set, whereas samples with the strong grade constitute only about 10%. Medium and fine sizes dominate in the data set. Angular blocky, blocky, and subangular blocky shapes were by far over-represented in the dataset. No columnar shapes and structureless soil were found; 14 samples had the wedge ped shape.

2.2. Data analysis

Optimum partitioning of databases with regression trees was used to find both the best predictors and best grouping of samples. The general functioning of the algorithm is recursive. Suppose that a database is organized as a table with columns $x_1$, $x_2$, $x_3$, ..., $x_N$ representing predictor variables and the column $y$ representing the response variable. First, the database table is sorted by column $x_1$, if $x_1$ is numerical, or subdivided into groups having the same $x_1$ if $x_1$ is categorical. All possible splits of this column into two parts are used to compute the measure of non-homogeneity among the values $y$ in these two parts. The same is done for all other columns. Results of all splits columns are compared and the best grouping variable is found which provides the split with the smallest overall non-homogeneity in two parts of database. This variable is used to create the first branching of the tree: the part of the database table with values of the grouping variable above (below) the split constitutes the left (right) branch. The first node is formed by the split, and the first binary partitioning is accomplished. The data subsets in branches are further partitioned in the same way.

Fig. 1. Distributions of structural parameters in soils in the data set of this work.
The non-homogeneity after a split is measured by computing deviances which are defined for an observation values $y$ as

$$D(\mu, y_i) = \sum(y_i - \mu)^2$$

Here $\mu$ is the mean value across all observations $y_i$. Each possible split generates left $\sum_L D(\mu_L, y_L)$ and right $\sum_R D(\mu_R, y_R)$ deviance values where subscripts ‘$L$’ and ‘$R$’ indicate subsets of $y$ values obtained after the split. The split deviance is the sum of right and left deviances:

$$D_{\text{split}}(\bar{\mu}_L, \bar{\mu}_R, y) = \sum_L D(\bar{\mu}_L, y_L) + \sum_R D(\bar{\mu}_R, y_R)$$

The split that maximizes the change in deviance

$$\Delta D = \sum D(\mu, y_i) - D_{\text{split}}(\bar{\mu}_L, \bar{\mu}_R, y)$$

is the split to choose.

2.3. Results and discussion

Correspondence between laboratory and field determination of textural class is presented in Table 1. The correspondence is a typical one (Rawls and Pachepsky, 2002b). The weighted average percentage of cases when field and lab determination coincide is 52%. These data show that the field-determined textural class may not be a reliable input in a pedotransfer function. However, when broad clay classes are defined, the accuracy of field determination may increase markedly (personal communication of the US Natural Resource Conservation Service staff). Clay classes were defined as (1) 0–14%, (2) 14–28%, (3) 28–42%, and (4) >42%. Clay class was added as an input in the regression trees along with field-determined textural class and soil structure parameters.

Regression trees water retention at 33 kPa water potential are shown in Fig. 2 for the 9 field-determined textural classes that have been well represented in the database. The clay class was the best grouping parameter in all but loamy sand textural classes.

The grade class was the grouping parameter for all soils with intermediate texture. Only loamy sands, silty clay loam, silty clay, and clay samples did not have grade in the list of the grouping parameters. A stronger grade increased water retention at $-33$ kPa by 2–4 vol.%. The observed effect of grade on the average $\theta_{33}$ is similar to the one reported for water retention at 10 kPa $\theta_{10}$ by several authors. Bouma (1992) observed differences in water retention between weak and strong grade in arable and grassed Haplaquent, respectively, both having subangular blocky structure. The average water retention at $-10$ kPa was larger in samples with strong grade (Bouma, 1992), although this difference was not statistically significant. Soil with a weaker grade also had smaller water retention at $-10$ kPa in the study of Anderson and Bouma (1973) who compared water retention of two fine silty mesic Argiudolls both having medium prismatic parting to subangular blocky structural units. Yet another insight in the importance of the grade give data from the study of Shaw et al. (1997) where pore-size distributions have been compared for $B_{tv}$ and $B_h$ horizons in 18 pedons of fine loamy, siliceous, thermic Kandiudults with various contents of plinthite. Image analysis showed much larger percentage of pores with the equivalent diameter between 0.05 and 0.005 cm in horizons with weak grade as compared with horizons with the moderate grade. That range of equivalent diameters corresponds to the range of matric potentials between 0.6 and 6 kPa which means that soil in the horizon with a weak grade loses much more water as the suction is applied as compared to the soil in horizons with the moderate grade. Southard and Buol (1988) observed that in Ultisols that they had studied, grade of blocky structure gradually became stronger with depth, whereas the amount of pores emptying at 10 kPa decreased with depth. This meant an increase in water retention since bulk density did not show depth-related trends. Grade appears to be a relatively strong grouping variable to distinguish between soils with different water retention.

The shape class was not in the list of grouping parameters only in sandy clay loam, clay loam, and silty clay loam samples. In loamy sand and loam, crumb, granular, and blocky structure leads to lower water retention as compared to platy, lenticular and prismatic structure. Blocky structure causes an increase in water retention, i.e. sandy loam, loam, silty clay and clay loam soil with strong grade and fine size being an exception. We hypothesized that, in the latter case, blocky shape of soil structural units
reflects the presence of smectite minerals that enhance water retention of soils. Another reason may be that soil textural differences create differences in role of shape of structural units in water retention. There have been reports on the relationship between aggregate shape and water retention (Holden, 1995).

The size class was the grouping parameter in loam, loamy sand, silty clay loam, and silty clay. The finer the size the larger was water retention. Soils with fine structural units had the $-33$ kPa water content 2–5% larger than soils with medium and coarse structural units. The absence of large structural units might mean absence of large pores and a wide pore-size distribution that should provide relatively large water retention near field capacity.

Field-determined structural categorical parameters provide enough information to be used in regression trees for partition soil samples by their water retention. All those parameters are defined only by three or four broad classes, and are observer-specific (i.e. Post et al., 1986) to the same extent as soil textural class is. The structural parameters can serve as grouping variables to define groups of soil samples with distinctly different average water retention for the groups. Nevertheless, the clay class was the leading grouping parameter. One possible explanation for that may be that arrangement of pore space that remains water-filled at $-33$ kPa (maximum effective pore radius is 0.0045 mm) does not have strong direct relationship with structural parameters that are visually recognized for structural units of much larger size.

An example of pedotransfer function in this work was developed for the water content at $-33$ kPa.
which is often selected to approximate soil water holding capacity, to estimate the available water content (Soil Survey Staff, 1997), and to estimate saturated soil hydraulic conductivity (Ahuja et al., 1984). This indicates particular hydropedological relevance of this water content, and also explains selection of this water content as a standard soil parameter in US soil survey. The pedotransfer relationships between water contents at other soil water potential and soil structure should be explored as the suitable databases become available.

Although the classification and regression tree technique has provided an interpretable partitioning and has shown internal relationships for the database in this work, it has limitations that preclude addressing several issues that might be of interest in some studies. Other multidimensional classification techniques should be used to find out whether a holistic representation of soil structure with the triplet of size, grade, and shape categories may have more predictive power as compared with using each of those structural parameters as independent predictors. Errors in the tree classification include the effect of field misclassification errors in categories texture and structure. To decompose regression errors and separate effects of misclassification, other regression techniques, e.g. "dummy coding" (McCullagh and Nelder, 1989), could be used provided the misclassification errors are known. A version of the dummy coding was successfully used by Lin et al. (1999) to estimate Ksat from morphometric indices.

We stress that the results of the CART application in this work do not imply any causal relationship between soil structural parameters, on one hand, and water retention, on another hand. Water retention is affected by the structure of pore space which is probably related to the properties of visible structural units. However, structural units are observed at the coarser spatial scale as compared with the pore size distribution measured with water retention. There may be a similarity in spatial arrangement of structural units at the scales of soil horizon and finer, and that may explain the suitability of structural data for grouping soils by their water retention (Pachepsky and Rawls, 2003). Applicability of fractal models to pore space scaling (Pachepsky et al., 2001) raise the hope that such similarity may indeed exist. Finally, the classification tree may merely reflect the fact that structural parameters and water retention are affected by the same basic soil properties, i.e. content and type of clay minerals, organic matter content and quality, etc. This is undoubtedly interesting issue to explore, because it seems to be remarkable that qualitative observations of soil morphology can be translated in quantitative soil hydraulic parameters.

Multiplicity and site-specificity of hydrologic models gain evidence and acceptance in hydrology (Beven, 2000) and it is therefore best (NRC, 2001) to consider a broad range of reasonable alternative hypotheses and models and base the models on a variety of different types of data. Armed with advances in categorizing soil–landscape relationships and cataloging existing structures, pedology has a potential to substantially contribute to the building the range of hypotheses that should be considered in hydrologic modeling. Needs of hydrologic modeling, in turn, may catalyze effort on organization of available soil information in a form(at) relevant to modeling needs. With any hydrological model, satisfactory estimates of model parameters and their ranges are important for satisfactory calibration of models, for simulations to assess the model behavior with realistic scenarios, and for assessment of the calibrated model performance based on prior and posterior probability distribution functions of the parameters (Neuman and Wierenga, 2003). This implies that pedotransfer functions can become an important component of pedotransfer functions to be used in model parameter-

![Fig. 3. Interaction between scales in PTF development and application.](image-url)
ization is already developed (Pachepsky and Rawls, 2005). More can be expected as information on soil structure is being incorporated in pedotransfer functions. Development of pedotransfer functions also begins to benefit using an indirect information about soil properties which can be obtained in spatially dense measurements, such as topographic attributes, soil color, soil penetration resistance, etc., and becomes readily available.

Hydraulic properties are known to be scale-dependent, their values change if the same model is used at the core and at coarser scales. Currently PTFs are developed at the soil core scale and yet they are meant to be used at coarser scales of pedon, hillslope, or watershed. Therefore, an upscaling procedure is needed to make the PTF predictions usable at coarser scales (Fig. 3). Such upscaling is far from trivial because in most cases the process description in a hydrologic model changes with the change in scale, and parameters of coarse-scale models differ from the parameters of the fine-scale models. One example is using soil water holding capacity and water balance model rather than soil water retention curve and Richards equation at pedon and hillslope scales. A single value of soil water potential to read field capacity from the water retention curve has never been found (i.e., Haise et al., 1955).

Attempts to develop PTFs for parameters of hydrologic models at scales coarser than soil core scale are scant. One may expect that characterization of soil structure at coarser scales can be useful to develop such scale-relevant PTFs. One possible approach is to develop a quantitative interpretation of the structure of soil cover. The fast progress in pattern recognition techniques creates the technical background for using soil maps (in conjunction with topographic and vegetation maps) for characterization of soil structure at coarser scales. This will help to eliminate the existing mismatch of scales in pedotransfer function development and applications (Fig. 3) by including in PTF the structural pedological information quantified at relevant scales.

Research on pedotransfer functions represents one of many opportunities of bridging hydrology and pedology. More can be and should be done in interfacing knowledge about water and soil.

**Acknowledgements**

We are grateful to Dr. D. Giménez and Dr. François Bartoli for their insightful comments and suggestions for the improvement.

**References**


