USING THE NRCS NATIONAL SOILS INFORMATION SYSTEM (NASIS) TO PROVIDE SOIL HYDRAULIC PROPERTIES FOR ENGINEERING APPLICATIONS

W. J. Rawls, A. Nemes, Y. A. Pachepsky, K. E. Saxton

ABSTRACT. *Modern agricultural, biological, and environmental engineers have a multitude of uses for soil hydraulic parameters that quantify the ability of soils and sediments to retain and transmit water. These parameters are difficult and costly to obtain, especially if large areas of land need to be characterized. An active search for the relationships of soil hydraulic parameters with readily available soil properties began in the 1970s based on compilations of data from various sources. Although substantial progress was made, further developments were hampered by the inhomogeneity of the data compendiums in terms of soil variables included, methods of their measurements, ranges of parameters, regional representation, and uncertain data quality. New opportunities to supply soil hydraulic parameters to the end users have been created by the public domain availability of soils information provided in the USDA-NRCS National Soils Information System (NASIS). These data coupled with analytical advances have enhanced the development of new relationships describing soil hydraulic properties. The database currently contains analytical data for more than 50,000 pedons describing U.S. soils. The data set has provided the opportunity to study the effects of qualitative information such as soil structure and topography properties, which improves our ability to estimate hydraulic soil properties. The size of the database also allowed experimentation with new data analysis methods that were not previously usable. A summary of methods that have used the NASIS dataset to predict the soil hydraulic properties for a range of scales is presented along with examples of engineering applications that use such estimates. Opportunities for future research based on the NASIS dataset are given.*

Keywords. *Pedotransfer functions, Soil data bases, Soil hydraulic properties, Soil properties.*

Modern engineering has multiple uses for hydraulic soil parameters that quantify the ability of soils and sediments to hold and transmit water and solutes. Soil hydraulic properties are cumbersome and costly to measure, especially if large areas of land need to be characterized. Combining and interpreting soil hydraulic properties for engineering purposes is a demanding and sometimes nebulous task, with millions of dollars often resting on the result. Historically, engineering data have been handbooks, general rules, and published soil surveys. These were often quite general, not site specific, and cumbersome. Modern engineering analyses require faster and more complete definitions of soil hydraulic properties.

Active research on the relationships of soil hydraulic parameters with readily available soil properties began in the 1970s. Most, if not all, of the early large-scale studies used data that were contributed by many individuals or extracted from publications. For example, Rawls et al. (1982) used 26 sources of compiled soil water retention data and 35 different sources of hydraulic conductivity data. Such data sets served research purposes well at the time and are still referenced, such as Timlin et al. (1999) or Schaap et al. (2001).

The 2007 version of the USDA-NRCS NASIS database (NRCS, 2007) contains analytical data for over 50,000 pedons of the U.S. Data analysis procedures have been standardized, and the measurements were performed almost exclusively by the National Soil Survey Laboratory in Lincoln, Nebraska. These two factors provide significant homogeneity to the data. Additional data are continually being added for an extremely wide range of soils. The NRCS NASIS database increasingly serves as an important data source for studies on soil hydraulic properties.

RESEARCH RESULTS FROM NASIS DATA

Early pedotransfer functions (PTFs) typically used basic soil physical properties, such as soil texture class information, particle size distribution (PSD), organic matter (OM) content, and bulk density (D_b) as input (see summary in Nemes and Rawls, 2006). More recent PTF studies that took advantage of the availability of the NRCS NASIS database focused on using additional input variables to such models, such as soil consistence, soil structure (Rawls and Pachepsky, 2002a), or topography (Rawls and Pachepsky, 2002b; Pachepsky et al., 2001) related variables.

Submitted for review in February 2007 as manuscript number SW 6913; approved for publication by the Soil & Water Division of ASABE in July 2007 as a contribution to the ASABE 100th Anniversary Soil and Water Invited Review Series.

The authors are **Walter J. Rawls, ASABE Fellow,** Hydrologist (retired), USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland; **A. Nemes,** Soil Scientist, USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, and Department of Environmental Sciences, University of California Riverside, Riverside, California; **Yakov A. Pachepsky,** Soil Scientist, USDA-ARS Environmental Microbial Safety Laboratory, Beltsville, Maryland; and **Keith E. Saxton, ASABE Fellow,** Hydraulic Engineer, Saxton Engineering and Associates, Pullman, Washington. **Corresponding author:** Walter J. Rawls, USDA-ARS Hydrology and Remote Sensing Laboratory, 10300 Baltimore Ave., Bldg. 007, BARC-West, Beltsville, MD 20705; phone: 540-895-9712; fax: 301-504-8931; e-mail: Walterrawls@hughes.net.

Some studies experimented with the grouping of soils prior to developing such relationships. Availability of common descriptive factors in the source database for a large number of soils was a prerequisite for such studies. Rawls et al. (1999) evaluated the usefulness of soil structural variables (grade, size, and shape) and the grouping of soils into the USDA soil texture classes, which can easily be estimated in the field, in evaluating soil water retention at four matric potential values (−6, −10, −33, and −1500 kPa). These authors used regression tree analysis to facilitate the inclusion of the above qualitative-type variables in the analyses, and compared the estimations to the use of more conventional, quantitative variables as input. They concluded that classification of the soils into textural classes was the primary determinant in determining soil water retention, and all listed structure-related variables appeared as delineators of homogeneous groups of samples within the texture classes. The general order of importance of those structural variables was shape, grade, and size; however, their relative importance varied somewhat from matric potential to matric potential, and between topsoils and subsoils.

Pachepsky and Rawls (1999) used a subset of the NASIS database, representing the state of Oklahoma, to examine the effect of grouping soils on PTF performance. The authors grouped the soils according to four criteria (soil great group, soil moisture regime, soil temperature regime, and soil textural class) and used clay, sand, and coarse fragment content, OM content, D_b at −33 kPa, and the cation exchange capacity (CEC)/clay ratio as predictors to estimate soil water content at −33 and −1500 kPa matric potential. Group method of data handling (GMDH) was used to develop regression equations. They found that preliminary grouping improved the accuracy of PTFs in most cases, but none of the examined grouping criteria could be identified as superior to the others. However, the reliability, i.e., the capability to make estimations for independent samples, of such group-specific PTFs did not prove to be significantly better than that without grouping, showing that there is no direct link between PTF accuracy and reliability.

Rawls and Pachepsky (2002b) evaluated the use of topographic variables in estimating soil water retention. Data on 216 soil pedons from NASIS were used, and field descriptors like genetic horizon number, slope, position on the slope classes, and land surface shape classes were used, along with soil textural classes, to estimate soil water retention at −33 and −1500 kPa using regression trees. Resulting tree structures were different for the two matric potentials, but the inclusion of topographic variables and soil horizon notation seemed to make up for errors made by using field-determined soil texture. For the A horizons, using field texture and categorical topographic variables was more accurate than using laboratory-determined soil texture. This study showed potential for the use of topographic descriptors in the estimation of soil water retention for large-scale applications.

Subsequently, Rawls and Nemes (2007, personal communication) used a subset of NASIS coupled with regression tree analysis to evaluate the usefulness of topographic variables, such as slope, shape of slope longitudinally and parallel to elevation contours, and hillslope profile class (the two-dimensional slope segments of a typical hillslope, with similar characteristics), in estimating −33 kPa and −1500 kPa soil water retention. They also experimented with adding geomorphic slope segment classes as input (representing the position of the pedon site within the segment of the slope) or replacing hillslope profile classes with it. Using the latter variable did not yield significant improvement primarily because of its correlation with the variable representing hillslope profile. It also restricted the availability of samples for the analyses; therefore, it was not further used in the analyses. In all cases, the primary and secondary grouping variable was soil texture class, delineating more homogeneous groups by texture. Hillslope profile class and the actual slope appeared to be important for loam, silt, and silt loam textures (−33 kPa) and sandy loam texture (−1500 kPa) in estimating soil water retention. If hillslope profile classification was not used, its place was taken by the longitudinal shape of the slope, while essentially leaving the tree structure unchanged for −33 kPa. For −1500 kPa, soils with fine texture were further divided by slope and the longitudinal shape of the slope, while the coarser-textured soils (loamy sand, sand, and sandy loam) were not subdivided by any of the topographic variables.

Rawls and Pachepsky (2002a) used estimators that describe soil consistence (i.e., dry consistency, stickiness, and plasticity) in addition to structural variables (shape, size, and grade class) and soil textural classification in an attempt to improve soil water retention estimates at −33 and −1500 kPa matric potentials. The rationale was that such estimators are widely available as they are routinely collected in field soil surveys, and their connection to soil hydraulic properties is easy to infer. Regression tree analysis showed that plasticity class, grade class, and dry consistency class were leading estimators of soil water retention at both examined matric potentials. Increase in plasticity, stronger grade for non-plastic soils, and harder dry consistency led to greater soil water retention. Adding consistence and structure-related variables to textural classification improved the accuracy of estimations to a small but significant degree. The above three studies were among the first ones that highlighted the value of qualitative-type data, which were mostly overlooked in previous soil hydraulic PTFs.

Various authors reported the relationship between organic carbon or organic matter (later OM) content and soil water retention differently. Rawls et al. (2003) used a subset of about 12,000 samples from the NRCS NASIS database as well as data from pilot studies on soil quality to examine this relationship. They used regression tree analysis and GMDH to show the benefit of using OM content and information on taxonomic order as input in addition to textural classification or using PSD data. They also used the resulting GMDH equations to display isolines of water content at −33 and −1500 kPa matric potentials. It was shown that the sign and degree of relationship between OM content and soil water retention is dependent on the amount of OM, but it is also dependent on soil texture, with clayey soils displaying negative relationship between OM and soil water retention.

Rawls et al. (2004) revisited the OM-soil water retention relationship. After reviewing past results, the authors performed new analyses using a subset of A horizons from NASIS and used GMDH to investigate the importance of OM on estimating soil water retention at −33 and −1500 kPa for the whole data set and after grouping by taxonomic order and texture classes. While significant improvement in the estimations was not achieved by the inclusion of OM content in the models, OM content appeared as selected input in all taxonomic orders except Vertisols. It was most significant in Mollisols, Alfisols, and Ultisols.

When textural classification was performed prior to developing the regression equations, improvements were also marginal, but OM content had great relative importance in the models for the coarsest-textured soils (sand, loamy sand,

and sandy loam) and for silts. Schaap et al. (2004) used over 47,000 samples from the NASIS database to test a number of PTFs and to show that alternative objective functions in PTFs can lead to reduced bias in the estimations. The authors did not find a unanimously dominant PTF.

Nemes and Rawls (2006) used data from three databases, among them NASIS, to evaluate the usefulness of PSD data determined in accordance with standards of different classification systems in estimating soil water retention. No classification system had clear advantage in estimating soil water retention, and the continuous representation of PSD (i.e., by geometric mean diameter and its standard deviation) was not superior to the pointwise representation of PSD. No evidence was found that interpolated PSD data would be less useful or accurate in estimating soil water retention; however, using the incorrect definition of sand, silt, and clay fractions was reported to carry significant risks.

Nemes et al (2006a) introduced a novel application to an existing nonparametric estimation/classification technique in estimating soil water retention using data from NASIS. Their approach consists of finding the *k* number of nearest neighbors (hence the name "*k*-nearest neighbor" (*k*-NN) technique) from a reference data set to each sample in the application/test data set in terms of their selected input properties. Once the *k* neighbors are identified, the weighted average of the values of their output variables will serve as the estimate. Nemes et al. (2006a) characterized this technique as a robust, competitive alternative to other, parametric PTF techniques, with a number of advantages over more complex parametric techniques. Its main advantage is that no redevelopment of equations is required if new data become available. Such a characteristic is particularly beneficial if usage of the technique is coupled with data of a continuously developing database, such as NASIS. The *k*-nearest neighbor software can be found at: www.ars.usda.gov/ba/anri/hrsl/computer_models.

Nemes et al. (2006b) further advanced the testing of the *k*-NN technique by completing a sensitivity analysis on seven different aspects that are relevant while using a *k*-NN model. In most cases, the authors took advantage of the same data that were used by Nemes et al. (2006a), but they also took advantage of data from two other data sets to test the validity of their assumptions on independent data sets.

Saturated hydraulic conductivity (K_s) is somewhat less analyzed in PTF research. This is primarily due to the more limited availability of measured data, which is partly due to the cost and complexity of reliable measurements. However, for the same reason, reliable PTFs that estimate K_s are probably even more in demand. NASIS contains a limited amount of measured K_s data, and a limited number of studies exist that estimate K_s using those data. Rawls et al. (1998) used soil texture, D_b, and the slope of the soil water retention curve to estimate K_s. This work was probably the earliest published work that utilized NRCS NASIS data for the purposes of estimating any soil hydraulic properties. Soils were first grouped by USDA texture classes, after which a redefined version of the Kozeny-Carman equation, based on effective porosity and the slope of the soil water retention curve, was presented and parameterized. This study is also seen as an extension of the study by Rawls et al. (1982), since data became available for more texture classes and a low/high D_b distinction is made within most of the texture classes. More recently, Nemes et al. (2005) examined the influence of OM on the estimation of K_s using data from three sources, including NASIS. They examined the performance of existing PTFs and used GMDH to develop regression equations for the direct estimation of K_s, as well as to indirectly estimate expected changes to K_s using a generalized Kozeny-Carman approach. It was concluded that estimations negatively correlated changes in OM to changes in K_s for some, if not all soils, independently of the development data and estimation technique that was used in existing and newly developed PTFs. The range of such soils appeared data set dependent, but was extensive within the valid input range of each PTF.

Applications in Engineering

Water management engineering projects such as watershed hydrology, domestic and agricultural water supplies, and soil water drainage are some of the most common applications of soil hydraulic properties. Because a large percentage of precipitation is infiltrated into the surface soil profiles, virtually all engineering hydrologic methods involve assessing the soil intake, water holding capacities, and transient antecedent water status. Surface runoff subsequently plays a large role in water erosion, floods, hydropower, and river management.

Infiltrated water is the water source for crop production over large expanses of rainfed agriculture. The soil profile is the temporary water storage reservoir for this important resource determined by the annual climatic and production cycles. With incoming precipitation highly variable both spatially and temporally, the resulting production is equally variable, ranging from excessively wet and requiring drainage to moderate to drought status. Major production supplies and markets depend on timely and accurate engineering assessments of these cyclical patterns, since soil water availability on a daily basis can have large impacts on the production results.

Irrigation system designs and operations for supplemental crop production water are highly dependent on local soil water characteristics. Compared to rainfed agriculture, irrigated crop water demand and supply is much more manageable with proper knowledge of the soil water and associated soil chemical characteristics. The success or failure of projects of all sizes rests with knowing the water characteristics of the soil profiles. Initial engineering assessments are often accomplished using existing data sets such as NASIS, and then supplemented with site-specific data as deemed necessary, although these analyses are expensive and relatively slow to accomplish.

Water management of irrigated systems rests with the design and operation of the application methods coupled with knowledge of the soil water capacity and plant demands. Failures of soil water data or application mechanics at any crop growth stage can result in catastrophic production losses. Management tools have rapidly become more sophisticated than the "rules of thumb" and "experience" based decisions used only a few years ago. Soil water profile measurements of water content and/or tension have become common; however, essentially all of these tools still require an understanding of local soil water characteristics, which determine plant-available water. These management tools are essential for virtually all systems and become particularly critical for high-value and water-susceptible crops. Recent emphasis on parallel management of soil chemicals, such as

minimizing nitrogen leaching, required leaching of soil salinity, and reduced turf grass chemical loss, has placed further demand on knowing the relationships with soil water.

Regions of excess water often require surface and subsurface drainage for efficient crop production or other land uses. Drainage systems are expensive and require efficient engineering designs based on the hydrology, topography, and soil water characteristics of the local fields. As with irrigation designs, preliminary drainage designs are often based on applicable archived soil data. Too few or improperly installed drainage lines provide poor results and poor crop production, while excessive lines add to the expense. Recent designs involving dual-purpose systems, which provide management options for both subsurface drainage and supplemental irrigation, place further demands on the soil water characteristic assessments.

Environmental engineering for projects such as wetland assessment, wildlife habitat, or flood protection all involve soil characteristics. Wetland determinations have become increasingly complex as the criteria involved have gone beyond traditional hydrologic definitions to include temporal and spatial indicators of plant and soil factors. Increased understanding of soil properties has significantly improved these important decisions.

Other engineering applications that benefit from accessing the NASIS data bank are not as obvious. Examples include geological engineers who quantify subsurface water for groundwater supplies or drainage requirements for slope stability over long-term hydrologic regimes to avoid costly and dangerous embankment failures. Structural engineers require both soil physical and water characteristics as they design building footings and subsurface structures.

A new set of soil water relationships was recently developed from the NASIS soil data base based on the readily available variables of soil texture and organic matter. Included are new relationships for water tensions and conductivities plus previously developed effects of soil density, gravel, and salinity to form a comprehensive predictive system. These equations form an interactive model of hydraulic soil properties for agricultural water management and hydrologic analyses (Saxton and Rawls, 2006). The predictive system includes a graphical user interface to provide easy application and rapid solutions and is available at: http://hydrolab.arsusda.gov/soilwater/Index.htm.

NRCS has developed rapid soils data access to the NASIS data archive on the web, which allowsthe soil series and associated laboratory data for a specific area of interest to be obtained for the U.S. (http://soils.usda.gov). Detailed instructions for engineering applications are available at: http://hydrolab.arsusda.gov/soilwater/Index.htm.

Summary

New opportunities to supply soil hydraulic parameters to end users have been created by the public domain availability of soils information provided in the NRCS National Soils Information System (NASIS) (NRCS, 2007). These data coupled with analytical advances has enhanced the development of new relationships describing soil hydraulic properties. The database currently contains analytical data for more than 50,000 pedons describing U.S. soils. The data set has provided the opportunity to study the effects of qualitative information, such as soil structure and topography properties, which improves our ability to estimate hydraulic soil properties. The size of the database also allows experimentation with new data analysis methods that were not previously usable.

References

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

Nemes, A., W. J. Rawls, and Y. A. Pachepsky. 2005. Influence of organic matter on the estimation of saturated hydraulic conductivity. *SSSA J.* 69(4): 1330-1337.

Nemes, A., W. J. Rawls, and Y. A. Pachepsky. 2006a. Use of a nonparametric nearest-neighbor approach to estimate soil hydraulic properties. *SSSA J.* 70(2): 327-336.

Nemes, A., W. J. Rawls, Y. A. Pachepsky, and M. Th. van Genuchten. 2006b. Sensitivity analysis for the nonparametric nearest-neighbor technique to estimate soil water retention. *Vadose Zone J.* 5(4): 1222-1235.

NRCS. 2007. National Soil Survey Characterization Data. Lincoln, Neb.: USDA-NRCS National Soil Survey Laboratory.

Pachepsky, Y. A., and W. J. Rawls. 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. *SSSA J.* 63(6): 1748-1757.

Pachepsky, Y. A., D. J. Timlin, and W. J. Rawls. 2001 Soil water retention as related to topographic variables. *SSSA J.* 65(6): 1787-1795.

Rawls, W. J., and Y. A. Pachepsky. 2002a. Soil consistence and structure as predictors of water retention. *SSSA J.* 66(4): 1115-1126.

Rawls, W. J., and Y. A. Pachepsky. 2002b. Using field topographic descriptors to estimate soil water retention. *Soil Sci.* 167(7): 423-435.

Rawls, W. J., D. L. Brakensiek, and K. E. Saxton. 1982. Estimation of soil water properties. *Trans. ASAE* 25(5): 1316-1320, 1328.

Rawls, W. J., D. Gimenez, and R. Grossman. 1998. Use of soil texture, bulk density, and the slope of the water retention curve to predict saturated hydraulic conductivity. *Trans. ASAE* 41(4): 983-988.

Rawls, W. J., Y. A. Pachepsky, T. M. Sobecki, and H. Lin. 1999. Effects of soil structure on soil water retention. In *Proc. 2nd Approximation Intl. Conference on Soil Resources: Their Inventory, Analysis, and Interpretation for Use in the 21st Century.* St. Paul, Minn.: University of Minnesota.

Rawls, W. J., Y. A. Pachepsky, J. C. Ritchie, T. M. Sobecki, and H. Bloodworth. 2003. Effect of soil organic carbon on soil water retention. *Geoderma* 116(1-2): 61-76.

Rawls, W. J., A. Nemes, and Y. A. Pachepsky. 2004. Effect of soil organic matter on soil hydraulic properties. In *Development of Pedotransfer Functions in Soil Hydrology*, 95-114. Y. A. Pachepsky and W. J. Rawls, eds. Developments in Soil Science, vol. 30. Amsterdam, The Netherlands: Elsevier.

Saxton, K. E., and W. J. Rawls. 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *SSSA J.* 70(5): 1569-1578.

Schaap, M. G., F. J. Leij, and M. Th. van Genuchten. 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.* 251: 163-176.

Schaap, M. G., A. Nemes, and M. Th. van Genuchten. 2004. Comparison of models for indirect estimation of water retention and available water in surface soils. *Vadose Zone J.* 3(4): 1455-1463.

Timlin, D. J., L. R. Ahuja, Y. A. Pachepsky, R. D. Williams, D. Gimenez, and W. J. Rawls. 1999. Use of Brooks-Corey parameters to improve estimates of saturated conductivity from effective porosity. *SSSA J.* 63(5): 1086-1092.