ELSEVIER

Journal of Hydrology 251 (2001) 151–162

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Description of the unsaturated soil hydraulic database UNSODA version 2.0

A. Nemes[a,*], M.G. Schaap[b], F.J. Leij[b], J.H.M. Wösten[c]

[a]*Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Herman Ottó u. 15, H-1022 Budapest, Hungary*
[b]*George E. Brown Jr. Salinity Laboratory, USDA-ARS, 450 W. Big Springs Road, Riverside, CA 92507, USA*
[c]*Alterra Research Institute for Green World Research, P.O. Box 47, NL-6700 AA Wageningen, The Netherlands*

Abstract

Quantifying water flow and chemical transport in the vadose zone typically requires knowledge of the unsaturated soil hydraulic properties. The UNsaturated SOil hydraulic DAtabase (UNSODA) was developed to provide a source of unsaturated hydraulic data and some other soil properties for practitioners and researchers. The current database contains measured soil water retention, hydraulic conductivity and water diffusivity data as well as pedological information of some 790-soil samples from around the world. A first MS-DOS version of the database was released in 1996. It has been applied in numerous studies. In this paper, we describe the second version (UNSODA V2.0) for use with Microsoft Access-97®[1]. The format and structure of the new database have been modified to provide additional and more convenient options for data searches, to provide compatibility with other programs for easy loading and downloading of data, and to allow users to customise the contents and look of graphical output. This paper reviews the structure and contents of the database as well as the operations that can be performed on the different data types in UNSODA V2.0. The use and application of the new database are illustrated with two examples. The retrieval of data is briefly illustrated, followed by a more detailed example regarding the interpolation of soil particle-size distribution data obtained according to different national definitions of particle-size classes. The interpolation procedure, which is based on finding similar particle-size distribution curves from a large European data set, also performed well for soils that originate from other geographical areas. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Soil hydraulic database; Unsaturated soils; Water retention; Hydraulic conductivity; Soil water diffusivity; Particle-size distribution

1. Introduction

Numerical models are increasingly being used to manage agricultural production, predict the behaviour of soil contaminants and to simulate transport processes in the vadose zone. These models typically require information on the relationships between water content (θ), pressure head (h) and hydraulic conductivity (K). The quality of these input relationships will often substantially affect the quality of the simulation results (Leij and van Genuchten, 1999).

---

* Corresponding author. Address: Alterra Research Institute for Green World Research, P.O. Box 47, NL-6700 AA Wageningen, The Netherlands. Tel.: +31-317474700; fax: +31-317419000.

E-mail addresses: attila@alterra.wag-ur.nl (A. Nemes), bloro@spike.fa.gau.hu (A. Nemes), mschaap@ussl.ars.usda.gov (M.G. Schaap), fleij@ussl.ars.usda.gov (F.J. Leij), j.h.m.wosten@alterra.wag-ur.nl (J.H.M. Wösten).

[1] Trade names are provided for the benefit of the reader and do not imply an endorsement by the authors or their organizations.

PII: S0022-1694(01)00465-6

Although advances are being made to measure hydraulic properties (cf. Gee and Ward, 1999), the methodology to determine soil water retention, $\theta(h)$, and, especially, the unsaturated hydraulic conductivity, $K(h)$ or $K(\theta)$, is often perceived as inadequate for many applications, which require a large number of samples (Wösten et al., 2001). An alternative to direct measurement of the unsaturated hydraulic properties is the use and/or generalisation of experimental data that are already available. The alternative approaches require reliable data sets of retention and conductivity information, as well as basic soil properties such as soil texture, bulk density, and organic matter content or other information pertinent to the hydraulic behaviour of soils. Electronic databases with soil information already exist, such as those for the United States (http://www.statlab.iastate.edu/soils/ssl/cdinfo.html), Canada (http://res.agr.ca/CANSIS/NSDB/), Australia (http://www.cbr.clw.csiro.au/acpep/), or the world (FAO, 1993; 1995; IGBP,http://www.meteo.fr/cnrm/igbp/). However, the emphasis of most of these databases is on soil taxonomy and they often have limited unsaturated soil hydraulic data. With this in mind, the international UNsaturated SOil DAtabase (UNSODA) (Leij et al., 1996) and subsequently, the European database of soil hydraulic properties (HYPRES) (Wösten et al., 1999), were developed. Both databases contain a wealth of information about soil hydraulic data, measurement methods and other relevant soil data.

Earlier, soil hydraulic databases often consisted of a number of individual data files according to a strict format as mandated by a specific database management program (e.g. Rosenthal et al., 1986; Leij et al., 1996). Such databases have the advantage of modest computer system requirements. New database software has now become available to handle widely varying data collections. The use of a common language such as Structured Query Language (SQL) makes it convenient to enter, manipulate, retrieve and extract information. HYPRES, developed in ORACLE Database Management Systems, is an example (Wösten et al., 1999). An advantage of current database software is their capability to interface with many other software packages.

Several studies, that relied on data from UNSODA, have recently been published. Leij et al. (1997) evaluated the mathematical expressions to describe water retention and unsaturated conductivity data in UNSODA. Schaap and Leij (1998) used UNSODA as one of the three databases to evaluate the accuracy and uncertainty of neural network based pedotransfer functions. Kravchenko and Zhang (1998) predicted the soil water retention data of 110 soils, from retention data using a fractal approach. Arya et al. (1999a,b) predicted the water retention and unsaturated hydraulic conductivity curves from the particle-size distribution using a physico-empirical approach. Hoffmann-Riem et al. (1999) developed a general pore-size distribution model to predict the hydraulic conductivity from retention data. Schaap et al. (1998) and Schaap and Leij (2000) developed neural network models to predict the water retention and unsaturated hydraulic conductivity from simpler soil properties. Kosugi (1999) used data from UNSODA to predict the conductivity assuming lognormal pore-size distribution. Poulsen et al. (2000) predicted the saturated and unsaturated conductivity from water retention data.

The database management program of the first version of UNSODA was, however, written for an MS-DOS environment, which is being viewed as obsolete. Furthermore, the aforementioned rigid format for data query and output, hampers users to make optimal use of the database. We therefore developed a second version of UNSODA for use with Microsoft Access. This paper describes its structure and provides a summary of the available data. We illustrate the use of UNSODA with two examples pertaining to query procedures and the interpolation of particle-size distribution data.

2. Database

2.1. Structure

UNSODA V2.0 is a compilation of the data of the previous MS-DOS based version V1.0 of UNSODA (Leij et al., 1996) in Microsoft Access-97 format. MS Access has been chosen because, it is widely available and allows management of data on 'standalone' computers as well as through computer networks. UNSODA V2.0 provides more flexibility in data entry, manipulation and retrieval as well as data output and interfacing with other applications than

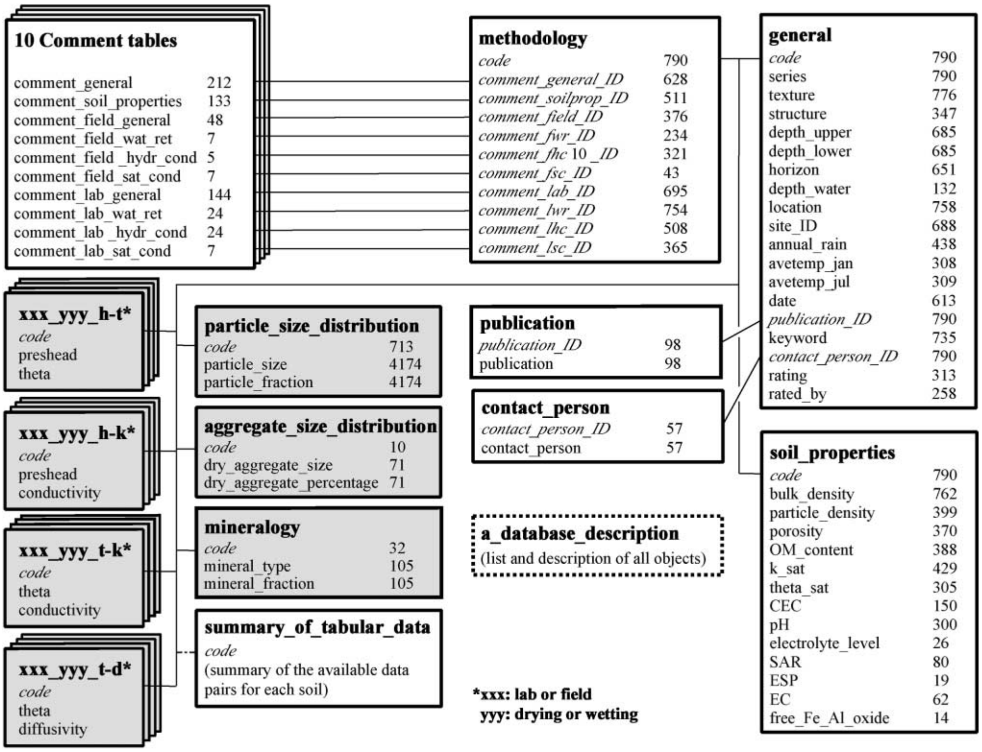

Fig. 1. An overview of the database structure and the data in UNSODA V2.0. The boxes indicate tables (table names are in bold), which include the fields and the number of records (samples) with available data. Indexed fields (in italics) are used to link the tables as depicted with the lines between the tables. The ten comment tables are shown here as one individual table while the 16 hydraulic properties tables are displayed in a generalised form with xxx for lab or field and yyy for drying or wetting branch. The box with dotted borders indicate an auxiliary table.

the first version. Furthermore, MS Access has extensive user-friendly query and graphics capabilities for perusing the database. The database design was kept as general as possible and the user can readily add, delete, edit data or, if desired, modify the database structure.

Data are stored in 36 tables. Tables store data in logical groups of fields containing related information. Each of the 790 soil samples (core sample or horizon) in UNSODA received a unique 4-digit identification number stored in a separate field named ‘code’. The numbering system is based on increments of 10 for records unrelated to other records, while an increment of 1 is used for related records (i.e. the same experiment or location but for a different soil horizon or treatment). The tables are usually linked (‘indexed’) through this field. This ID system is more efficient and error free while defining relations between tables than via text fields.

By opening a table in ‘design view’, the structure of each individual table, the name, data type and description of each field in the table can be inspected and modified. Through the design view, one can also specify what the default value should be, if there are no data for a particular field. Missing data are usually indicated with a ‘No data’ entry.

The structure of the database, names of tables and links between tables are outlined in Fig. 1. The main table of UNSODA is called ‘general’. It holds basic information about the soils such as their geographic location, classification and environment. The ‘soil-properties’ table contains physical and chemical characteristics for each soil.

Altogether 19 tables contain data with a functional relation between an independent and a dependent variable. Hydraulic data are stored as h–θ, h–K, θ–K and θ–D curves. The absolute value of the soil matric or pressure head, h, is given in cm (hPa); the water content, θ, is expressed in cm^3/cm^3; the hydraulic

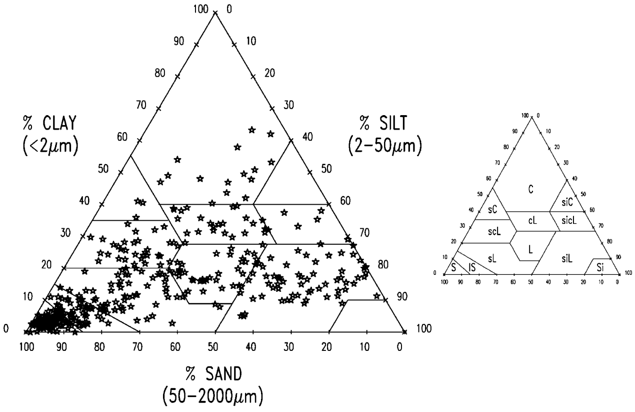

Fig. 2. Distribution of 431 soil codes in UNSODA V2.0 across the USDA-SCS soil textural triangle.

conductivity, K, in cm/d and the soil–water diffusivity, D, in cm^2/d. Distinctions between the wetting and drying branch and between laboratory and field determinations, result in a total of 16 hydraulic tables (cf. Fig. 1). Three other tables contain data on the particle-size distribution, aggregate-size distribution and mineralogy. The aggregate-size distribution table stores the cumulative fraction of soil mass as a function of the equivalent dry aggregate size or diameter in a similar manner as the particle-size table does for soil particles smaller than 2 mm. The mineralogy table shows the mass fraction of individual soil or clay minerals. The above 19 tables contain, for each sample, the field 'code' followed by a data pair. In this way, a soil having eight measured points on e.g. the water retention curve, will occupy eight lines in that particular table. The table 'summary-of-tabular-data' lists the number of available data pairs in the 19 tables for each of the 790 codes. This allows quick querying for samples with — for instance — a certain minimum number of retention points.

The 'methodology' table contains ten 'comment-ID' numbers referring to records in ten different 'comment' tables, which contain information on laboratory and field procedures and on specific methodology for hydraulic measurements. The 'contact-person' and 'publication' tables contain source information related to each particular code. They are designed in a similar manner as the 'comment' tables, but their indexes appear directly in the 'general' table.

Some auxiliary tables are included in UNSODA V2.0. A table entitled 'a_database_description' provides a summary of the field names, data types and other details of all objects (tables, queries and reports) provided in the database. The table is provided for the benefit of the user. Two additional hidden tables 'code_filter' and 'only_codes', are used in a predefined query; they are of no direct importance for the user, but their role in data reporting is briefly discussed later in Section 3.

2.2. Data

The present database holds information on 790 soil horizons ('codes') contributed either by individual scientists or obtained by us from the literature. The number of records for each field in the database can be seen in Fig. 1, except for the 16 tables that contain soil hydraulic data, which are discussed separately. The particle-size table should be interpreted as follows: the database contains 713 soil samples with particle-size data, with a total of 4174 size/fraction data pairs. The aggregate size and mineralogy tables should be interpreted likewise. Fig. 2 shows the USDA-SCS texture triangle and the distribution of the 431 samples with original measured data that are compatible with the classification system. Table 1 shows a summary of the geographical distribution and the textural classification, according to the USDA system, of the UNSODA soils. Most of the data of the database came either from Europe or from North America. The textural classification indicates that coarse-textured soils are in a majority although there is a sizeable amount of soils with finer textures. However,

Table 1
Geographical and textural distribution according to the USDA-SCS classification of soils in UNSODA V2.0

	S	lS	sL	L	SiL	Si	scL	cL	sicL	sC	siC	C	N/D	Total
Africa	3	5	2				1							11
Asia	4		1	5	3		1	1			2	1	2	20
Europe	75	19	46	45	93	2	15	10	15		12	16	12	360
North America	88	34	77	17	37	1	33	21	13	3	10	20	1	355
Pacific Region	1		2		1		1	4	1			2		12
No data (N/D)	14	2	5	2	7		1		1					32
Total	185	60	133	69	141	3	52	36	30	3	24	39	15	790

sandy clays and silts are poorly represented.. The textural distribution of the samples is partly due to the natural occurrence of soil textural types and, undoubtedly, due to an experimental bias in favour of coarse-textured soils.

Table 2 provides the number of hydraulic curves and the total number of data points, subdivided into drying and wetting branches, as well as curves determined in the laboratory, and in the field. There are far more hydraulic data, for drying conditions, and more laboratory, than field data. The actual number of data points on the hydraulic curve, as well as the range in θ or h, is important, when optimising mathematical expressions or pedotransfer functions, to describe hydraulic data. Not all contributed data were used for the compilation of these tables to avoid repetition and bias. Some of the hydraulic data are geometric averages for replicate values of the independent variable, or curves for the same sample or experiment. Fig. 3 shows the scatter-plot of all the soil hydraulic data, available in the database. Laboratory methods generally allow the determination of hydraulic characteristics for dryer conditions than field methods. Details about the measurement procedures are included as comments in separate tables (cf. Fig. 1) to enable the user to focus on data, determined with a particular method.

Table 2
Summary of the number of hydraulic curves (and the number of data pairs in parentheses) for lab and field samples and drying and wetting conditions

		Field	Laboratory
Water retention (h–θ)	Drying	137 (2621)	730 (8066)
	Wetting	2 (8)	33 (528)
Hydr. Conductivity (h–K)[a]	Drying	144 (2826)	730 (6187)
	Wetting	0	8 (71)
Hydr. Conductivity (θ–K)	Drying	294 (5391)	293 (5177)
	Wetting	0	20 (216)
Soil–water diffusivity (θ–D)	Drying	56 (1282)	92 (1456)
	Wetting	0	2 (13)

[a] Only points with $K(h) > 0$ cm/day are included.

3. Database operations

The database allows a number of operations such as searching, according to the user-defined criteria, editing or adding data and extracting or reporting all or selected parts of the database. In the following, we will briefly review these operations.

3.1. Retrieval of information with queries

Specific information can be obtained from UNSODA by running a 'select query' to retrieve data from one or more tables according to user-defined criteria. MS Access offers the possibility to enter queries through a graphical interface, which does not require detailed knowledge of SQL or by specifying queries in SQL. Queries are designed by first selecting the relevant tables, after which, the appropriate fields are chosen. These fields are selected either because their values constitute the desired output of the query or because they are used as constraints. Data that conform to the query criteria will appear as output in a table-like structure. Queries can also be run on data sets that were obtained in prior queries (i.e. nested sets of queries). This may be useful, when data need to be selected, according to complex criteria. Query definitions can be saved for later use.

UNSODA V2.0 comes with one predefined query

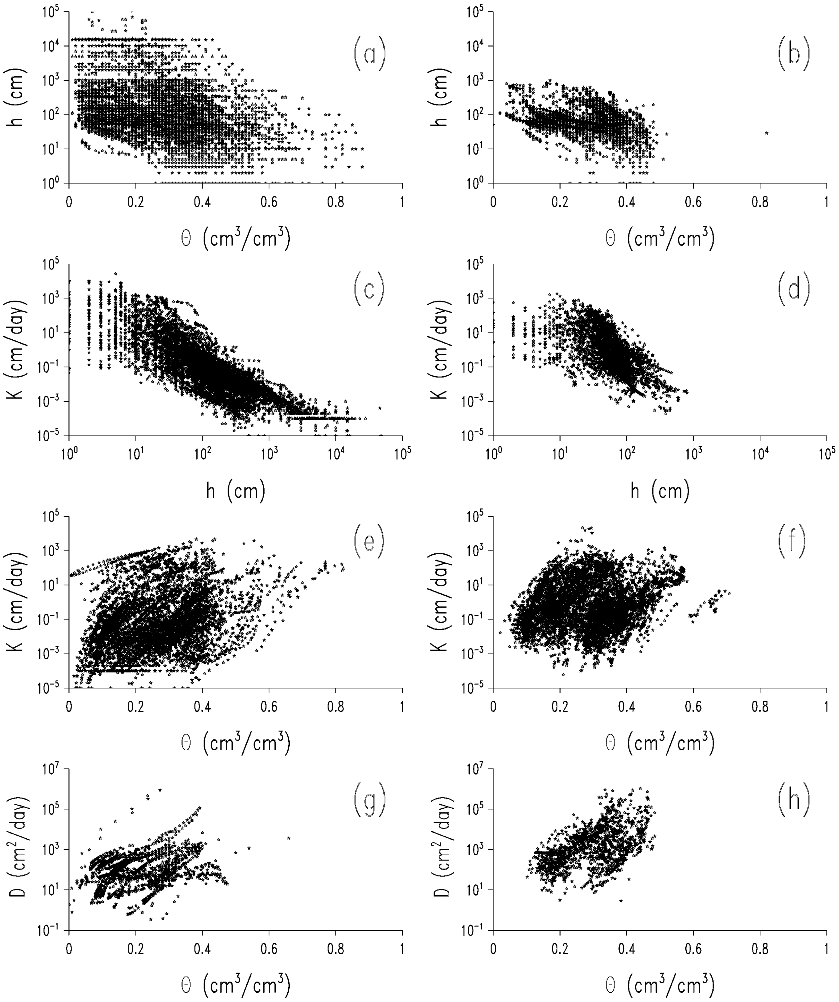

Fig. 3. Scatter-plot of the hydraulic data in UNSODA V2.0. From top to bottom: water retention, $h(\theta)$, hydraulic conductivity, $K(h)$ and $K(\theta)$, and soil water diffusivity $D(\theta)$ data. Laboratory measurements are shown on the left-hand side while field measurements are given on the right-hand side.

called ‘filter_for_reports’ that generates a table called ‘code_filter.’ With this query — and the created table — the user can generate reports that will contain only those ‘codes’ that were selected according to the specified criteria.

3.2. Editing and adding data

Because UNSODA V2.0 is not write-protected, the user can modify the contents of the database. Editing data is simply a matter of overwriting the proper fields. When some, or all of the entries have to be changed, considerable effort can be avoided by defining an ‘update query’ that will search and overwrite the contents of all entries that meet with the preset criteria.

Adding data is straightforward, but the type and format of the added data has to match with those of the targeted field(s), while the structure of the database should not be changed, to ensure that existing queries function properly. Different ways of data entry are possible. One can simply open the targeted table in datasheet view and start typing the new data at the end of the appropriate field. If large amounts of data need to be entered, it may be useful to import the data from ASCII or spreadsheet files.

Alternatively, one can create a form that includes a selection of fields. Forms can display additional information on, for example, the required data format. Using a form, one can simultaneously enter data for fields in two or more tables. Data entered in the fields, of such a form, are automatically included in the original table(s). Customised forms are not provided in UNSODA V2.0, but users can readily create their own that best match their needs.

Each of the columns in the 'summary_of_tabular_-data' table represent one of the 19 tables that hold data with a functional relation between a dependent and an independent variable — as described earlier. Although the long column headers may seem confusing, it is easy to decide which column one should consult, e.g. the field 'count_lab_dry_h-t_CountOfpreshead' shows — for each soil — how many points are available on the drying branch of the water retention (*h*–theta) curve measured in the laboratory. The database comes with 23 hidden queries, which serve to automatically keep the 'summary_of_tabular_data' table up to date once the provided hidden macro 'update_table_summary_of_tabular_data' is run. The macro runs the hidden queries — which are nested — and will automatically update all the counts in the above table. This may be necessary when new data is added to the database. These queries and the macro are not necessary for normal data retrieval, but are more for developers; and hence, are hidden from the view of the user.

3.3. Extraction of data

Results from a query and the contents of the tables can be extracted in ASCII, spreadsheet, or HTML formats, among others. However, preliminary extraction from the database may not always be necessary as Access allows Visual Basic programming to carry out calculations, which are more complex, within the database. Data can also be passed to other programs for further processing using their Visual Basic objects.

Alternatively, users can define reports that generate formatted output, based on data selected by queries. UNSODA V2.0 includes three predefined reports on codes that are listed in the 'code_filter' table, which is generated by the 'filter_for_reports' query — as discussed earlier. The first general reporting option ('general_report') provides information on soil description, physical and chemical properties and comments on methodology used in sampling and measurements. The two other reports present all the available soil hydraulic data, grouped by soil codes in tabular form, or a combination of tables and graphs. Higher-quality graphs may be produced through OLE linkage of MS Access with MS Excel and/or MS Graph.

4. Application examples of UNSODA

In the following, we will give two examples that demonstrate the flexibility of data retrieval and use of UNSODA V2.0. The first example entails a case in which we demonstrate how to retrieve data from UNSODA, with a given set of criteria. The second example is longer and more practical; it describes a test of the particle-size interpolation algorithm of Nemes et al. (1999). Interpolation is needed to represent experimental particle-size data, in terms of standardized size classes for pedotransfer functions, soil databases, and other applications (e.g. Rawls et al., 1991; Schaap et al., 1998; Arya et al., 1999a,b). The examples are deliberately kept simple, because they serve as illustrations for the use of the UNSODA database; examples of more elaborate applications can be found in the references given in Section 1.

4.1. Query example

Let us assume that we are interested in soil samples from the USA having bulk density greater than 1.6 g/cm^3 with more than ten points on the drying branch of the water retention curve measured in the laboratory. In this case, one would choose the 'general', 'soil-properties', and 'summary_of_tabular_data' tables. Note that the tables are linked through the 'code' or 'general_code' fields, essentially allowing the query to consider all the three tables, as one composite table. The fields to be selected from the 'general' table are 'code' and 'location,' for the latter we specify the string 'Like '*USA*'' as search criterion. Further, we select 'bulk_density' from the soil properties table, using the criterion '>1.6', and select the field 'count_lab_dry_h-t_CountOfpreshead' from the 'summary_of_tabular_data' table with the criterion

'>10'. The output of the query includes all code numbers, that match the criteria, but values for constraining fields do not necessarily have to appear in the output. If the measured retention data pairs are needed as output, the query should be extended by adding the table 'lab_drying_h_t', after which, the fields 'preshead' and 'theta' can be selected.

4.2. Interpolation of particle-size data

The second example deals with the problem that soil textural data are not based on a uniform set of particle size because of differences in experimental procedures or in the definition of the silt–sand boundary (e.g. Nemes et al., 1999). As a result, textural classification according to the FAO/USDA system was not always possible because of a missing 50 μm particle-size fraction.

Nemes et al. (1999) developed a so called 'similarity procedure' to estimate unknown particle-size fractions. Using the HYPRES database, they showed that this method was superior to the spline interpolation and the commonly used log-linear interpolation methods. The similarity procedure does not rely on mathematical interpolation, but makes use of a large external reference data set of individual soil particle-size distribution (PSD) curves (reference curves).

We hypothesise that the accuracy of the procedure may therefore depend on the geographical origin of the soils. Because of different conditions for soil genesis, the interpolation for soils from similar regions as the reference data set should yield a more accurate estimation of the 50 μm fraction than for soils from a completely different region. In this study, we tested this hypothesis by estimating the 50 μm fraction for two data sets selected from UNSODA V2.0 each holding samples from different parts of the World.

The reference data set that was used, contains 9607 individual soil PSD curves, originating from the Soil Information System of the Netherlands (Finke, 1995). This reference data set encompasses a wide range of soil textures and is identical to that used by Nemes et al. (1999). The two testing data sets were selected from UNSODA V2.0, using the previously described query feature. A nested set of queries was used to select soils, with at least six points on the PSD curve, including the fraction at 50 μm. First, three separate queries were used to search for codes that have measured 2, 50 and 2000 μm, data respectively. This was achieved by selecting the 'code' and 'particle-size' fields from the particle-size table with the appropriate particle-size specified as a criterion. These queries were subsequently used as input to a new query, along with the 'summary_of_tabular_data' table, which were all manually linked by the code field. In this query, we specified that 'Countofparticle_size' should be greater than 5 and the linked field ensured that only codes that are present in all input tables/queries are selected. The total of 242 soils that remained, were subdivided into 72 samples from Europe (from Belgium and the Netherlands) and 170 samples from outside Europe (mainly the USA) by including the 'general' table in the above query and by using the 'location' field. The selected codes with the related raw particle-size data were then extracted from the database for the subsequent calculations.

The similarity procedure involves searching for a number of soils, in the external reference data set, that have a similar particle-size distribution as the soil for which the missing 50 μm fraction is to be estimated. Similarity between soils can be quantified with the correspondence of fractions at common particle sizes. The selection algorithm compares soils as many matching particle sizes as possible, usually four to six. At first, a soil from the reference data set is considered to have 'similar' PSD to a particular soil, when the difference in their particle fractions is less than 0.1 percent, at each of the common particle sizes, throughout the PSD curve. When there are no or very few soils selected, the 0.1 percent criterion is gradually relaxed until ten soils are selected from the reference data set. The estimate is then calculated as the arithmetic mean of the 50 μm fractions of the selected reference curves.

When the closest matching particle sizes are distant from the 50 μm limit on the size scale, it is possible that the interpolation method leads to inaccurate results. To assess the applicability of this approach, we considered different distances between the matching points that are neighbours to the 50 μm point. To quantify this distance, we followed the same terminology as in Nemes et al. (1999), where ϕ is defined as $-\log_2$[particle-size in mm] and $\Delta\phi$ indicates the distance between the neighbouring points on the ϕ scale. For example, for the distance between 20 and

Table 3
Summary of the number (*n*) and percentage of samples in each USDA-SCS texture group and each $\Delta\phi$ group for the (a) non-European ($N = 294$) and (b) European ($N = 299$) subsets

Texture group													
$\Delta\phi$	S	lS	sL	L	SiL	Si	scL	cL	sicL	sC	siC	C	*n*	% of *N*
Panel A														
2.32	25	13	8	6	8		1		1				62	21.1
3.32	1		1				1						3	1
3.64	25	13	8	6	8		1		1				62	21.1
5.64	35	14	8	8	9		1	4	2				81	27.6
5.71	1	6	4		12		3	1	1			2	30	10.2
5.97	19	9	10	2			13	1		2			56	19
n	106	55	39	22	37	0	20	6	5	2	0	2	**294**	
% of *N*	36.1	18.7	13.3	7.5	12.6	0	6.8	2	1.7	0.7	0	0.7		**100**
Panel B														
2.32	15	6	9	3	18							2	53	17.7
2.71	13	1	1				1		1				17	5.7
3.32	30	12	18	6	36							4	106	35.5
4.32	15	6	9	3	18							2	53	17.7
5.64	15	6	9	3	18							2	53	17.7
5.71	13	1	1				1		1				17	5.7
n	101	32	47	15	90	0	2	0	2	0	0	10	**299**	
% of *N*	33.8	10.7	15.7	5.0	30.1	0	0.7	0	0.7	0	0	3.3		**100**

200 μm, $\Delta\phi$ is 3.322. In some cases, larger $\Delta\phi$ were achieved by intentionally disregarding a measured point, that is close to the 50 μm point from the entire procedure, thus in those cases that point was not used for the comparison of similarity. In this way, some soils could be reused to evaluate multiple $\Delta\phi$ distances.

Estimated values for the 50 μm fraction, were

Table 4
Calculated RMSE values for different USDA-SCS texture groups and $\Delta\phi$ for the non-European (A) and European (B) subsets. Highlighted are figures that rely on the average of at least five samples (see Table 3)

$\Delta\phi$	S	lS	sL	L	SiL	Si	scL	cL	sicL	sC	siC	C	RMSE by $\Delta\phi$
Panel A													
2.32	**4.03**	**4.17**	**7.59**	**4.18**	**6.11**		8.60		3.82				5.06
3.32	0.02		1.40				9.01						5.27
3.64	**3.42**	**2.52**	**3.28**	**6.80**	**6.95**		3.07		2.70				4.29
5.64	**4.10**	**4.80**	**9.78**	**5.29**	**18.39**		18.12	7.51	6.29				8.30
5.71	1.67	**2.87**	3.33		**11.81**		1.39	0.09	10.68			3.54	7.99
5.97	**2.34**	**3.91**	**4.16**	12.85			**1.89**	9.53		3.43			4.02
RMSE by texture	3.63	3.86	6.27	6.52	12.08		5.22	7.26	6.56	3.43		3.54	**6.18**
Panel B													
2.32	**3.43**	**3.00**	**4.09**	0.97	**7.69**							7.68	5.44
2.71	**1.30**	4.00	5.79				0.39		4.98				2.38
3.32	**4.14**	**4.95**	**5.81**	**8.98**	**15.13**							7.55	9.89
4.32	**5.29**	**6.83**	**10.55**	7.31	**24.96**							7.30	15.77
5.64	**5.77**	**3.68**	**9.85**	11.55	**14.63**							6.64	10.46
5.71	**1.14**	4.43	5.74				0.39		4.06				2.25
RMSE by texture	4.04	4.82	7.58	8.35	16.46		0.39		4.55			7.35	**10.20**

Table 5
Original vs. predicted texture groups for the non-European (A) and European (B) subsets. Figures along the diagonals (bold) represent soils with correctly predicted classification

Original textural classification														
		S	lS	sL	L	SiL	Si	scL	cL	sicL	sC	siC	C	**Total**
Panel A														
Textural classification based on the interpolation procedure	S	**85**	6											
	lS	21	**47**	3										
	sL		2	**31**	1									
	L			3	**17**	5		1						
	SiL			2	4	**32**								
	Si													
	scL							**19**						
	cL								**4**					
	sicL								2	**5**				
	sC										**2**			
	siC													
	C												**2**	
Prediction (%)		80	85	79	77	86		95	66	100	100		100	**83%**
Panel B														
		S	lS	sL	L	SiL	Si	scL	cL	sicL	sC	siC	C	**Total**
Textural classification based on the interpolation procedure	S	**90**	6											
	lS	9	**23**											
	sL	2	3	**38**	2	7								
	L			9	**8**	28								
	SiL				5	**55**								
	Si													
	scL							**2**						
	cL													
	sicL									**2**				
	sC													
	siC												7	
	C												**3**	
Prediction (%)		89	72	81	53	61		100		100			30	**74%**

compared with the measured values, which were of course omitted from the UNSODA soils at the selection of the reference curves. The root mean squared error (RMSE) was calculated for each represented USDA-SCS texture group and each $\Delta\phi$ distance separately for the two data sets as:

$$\mathrm{RMSE} = \sqrt{(1/n)\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

where n is the total number of estimated values for a textural group or $\Delta\phi$ value, x and y are the measured and estimated particle-size fraction in mass percentage.

Results of the test are shown in Tables 3–5. Table 3 summarises the available data. An almost equal number of cases were evaluated for the two data sets ($N = 294$ and 299). Most of the soils belonged to coarse- and medium-textured groups. However, some texture groups were, poorly or not represented. The European data set was represented by a considerably larger number of evaluations involving medium-textured — mainly silt loam — soils. Table 4 shows the results of the evaluation, in terms of RMSE values, expressed as percentages. Results that are based on at least five cases (cf. Table 3) are highlighted. In most cases, the coarse textured soils were reliably predicted, which is shown by the lower RMSE values. The group of silt loam soils exhibits much higher RMSE for both data sets, than any of the other texture groups. This reflects that the estimation procedure is least reliable for this texture group. When different $\Delta\phi$ are compared, some larger $\Delta\phi$ show considerably higher RMSE values for both data sets, than others. However, a clear trend of increasing errors with increasing $\Delta\phi$ could not be shown from these data due to the uneven textural composition of the different $\Delta\phi$ groups. Typically, those groups show higher RMSE, in which, there were more silt loam soils. When the corresponding groups for the two data sets are compared, it is clear that the estimation is as good for the non-European test set as that for the European. One expects that, for the European data set, the estimation is not worse than that for the non-European data set. However, the overall RMSE of 10.2 for the European data set is high, because of the much higher representation of medium-textured soils in that data set, for which, the evaluation is much worse. Table 5 shows the original textural classification versus how the soils would be classified if the estimated 50 μm fractions were used. Highlighted numbers in the diagonals show, for each data set, the number of cases when the soil was classified correctly according to the USDA-SCS classification. Numbers outside the diagonals represent the wrongly classified cases. For the non-European soils, the procedure had a better overall ratio to predict in the correct texture group (83%) than for the European soils (74%). The low prediction ratios of silt loam soils in the European data set correspond with the higher RMSE values of Table 4. However, most incorrectly classified soils were classified as a similar group of the textural triangle. However, even small estimation errors could lead to incorrect classification, if the soil was originally close to the texture group boundary.

Altogether, the use of the similarity procedure for estimation of the 50 μm point on the PSD curve shows to be reliable for the coarse-textured soils and much less reliable for medium-textured soils. The number of fine-textured samples in the data sets was quite low; a reliable evaluation could not be made from these data. An increase in reliability with a shorter distance $\Delta\phi$ was not readily apparent. The similarity procedure seems to be applicable for soils outside the source area of the reference data set (i.e. for the non-European soils in this case). The evaluation in this study was somewhat influenced by the uneven distribution of soils in the two testing data sets.

5. Summary and recommendations

Databases with information on soils, often constitute the basis of applications in production agriculture, environmental engineering and remote sensing. UNSODA V1.0 was one of the first public domain databases, with unsaturated soil hydraulic data, from around the world. However, this database was written for an MS-DOS environment, which is becoming obsolete. We have therefore developed UNSODA V2.0 to be used with Microsoft Access. UNSODA V2.0 is compatible with most of the popular software and can be run on a personal computer. The user-friendliness and the wide range of data should make the database a valuable tool. We provided an outline of the structure and data of UNSODA V2.0 and demonstrated the query feature. We also showed an application, which involved a technique to interpolate soil textural data.

Additional data can be easily included in the database. We encourage users to submit relevant laboratory or field data to the authors at the Salinity Laboratory for inclusion in future versions of UNSODA. Candidate data should have, at a minimum, experimental data on the particle-size distribution, and the retention and unsaturated conductivity or diffusivity curves. Currently lacking from the database, are data for tropical soils. Such data are needed badly given the environmental issues in tropical and subtropical areas. Users of the database have to be aware of limitations set by the geographical distribution of the available data and the variability in measurement techniques used to obtain the data.

The development of more accurate pedotransfer functions is just an example of the applications that could benefit from the expansion of the database. Incorporation of data in public domain databases such as UNSODA may furthermore perpetuate their utility beyond individual projects. All too often useful (and expensive!) data are lost because research projects terminate and this type of information can not be disseminated in a meaningful manner in peer-reviewed publications.

6. System requirements, availability of the database

UNSODA V2.0 currently requires 4 Mbytes of disk space to store the database. A Pentium based computer system is recommended, to avoid slow display of reports and graphs. The database is free of charge and can be requested by regular or electronic mail (Mr. Walt Russell, USDA-ARS, George E. Brown Jr. Salinity Laboratory, 450 West Big Springs Road, Riverside, CA 925074617, USA; wrussell@ussl.ars.usda.gov) or it may be downloaded through the Internet (http://www.ussl.ars.usda.gov/). The database management software MS Access is available as part of Microsoft Office (Professional Edition) or as a separate program (Microsoft Corporation, P.O. Box 97017, Redmond WA 98073, USA).

Acknowledgements

The authors wish to thank all individuals who contributed data to the database. Two of us (MGS and FJL) gratefully acknowledge support by NSF and NASA (EAR9804902) and ARO (39153EV) while we all thank support for travel by NATO (CLG 975761).

References

Arya, L.M., Leij, F.J., van Genuchten, M.Th., Shouse, P.J., 1999a. Scaling parameter to predict the soil water characteristic from particle-size distribution data. Soil Sci. Soc. Am. J. 63, 510–519.

Arya, L.M., Leij, F.J., Shouse, P.J., van Genuchten, M.Th., 1999b. Relationship between the hydraulic conductivity function and the particle-size distribution. Soil Sci. Soc. Am. J. 63, 1063–1070.

FAO, 1993. Global and national soils and terrain digital database (SOTER). Procedures manual. FAO World Soil Resour. Rep. 74,. Rome, Italy.

FAO, 1995. Digital soil map of the world and derived soil properties. Version 3.5. FAO, Rome, Italy.

Finke, P.A., 1995. Bodeminventarisatie: van gegevensverzameling tot informatieverstrekking. In: Buurman, P., Sevink, J. (Eds.). Van bodemkaart tot informatiesysteem. Wageningen Press, Wageningen, The Netherlands, pp. 111–125.

Gee, G.W., Ward, A.L., 1999. Innovations in two-phase measurements of soil hydraulic properties. In: van Genuchten, M.Th., Leij, F.J., Wu, L. (Eds.). Proceedings of the International Workshop, Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media. University of California, Riverside, CA, pp. 241–269.

Hoffmann-Riem, H., van Genuchten, M.Th., Flühler, H., 1999. A general model of the hydraulic conductivity of unsaturated soils. In: van Genuchten, M.Th., Leij, F.J., Wu, L. (Eds.). Proceedings of the International Workshop, Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media. University of California, Riverside, CA, pp. 31–42.

Kosugi, K., 1999. General model for unsaturated hydraulic conductivity for soils with lognormal pore-size distribution. Soil Sci. Soc. Am. J. 63, 270–277.

Kravchenko, A., Zhang, R., 1998. Estimating the soil water retention from particle-size distributions: a fractal approach. Soil Sci. 163, 171–179.

Leij, F.J., Alves, W.J., van Genuchten, M.Th., Williams, J.R., 1996. Unsaturated Soil Hydraulic Database, UNSODA 1.0 User's Manual. Report EPA/600/R96/095. US Environmental Protection Agency, Ada, OK 103 p.

Leij, F.J., Russel, W.B., Lesch, S.M., 1997. Closedform expressions for water retention and conductivity data. Ground Water 35, 848–858.

Leij, F.J., van Genuchten, M.Th., 1999. Characterisation and measurement of the hydraulic properties of unsaturated porous media. In: van Genuchten, M.Th., Leij, F.J., Wu, L. (Eds.). Proceedings of the International Workshop, Characterization and Measurements of the Hydraulic Properties of Unsaturated Porous Media. University of California, Riverside, CA, pp. 1–12.

Nemes, A., Wösten, J.H.M., Lilly, A., Oude Voshaar, J.H., 1999. Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. Geoderma 90, 187–202.

Poulsen, T.G., Moldrup, P., Yamaguchi, T., Jacobsen, O.H., 2000. Predicting saturated and unsaturated hydraulic conductivity in undisturbed soils from soil water characteristics. Soil Sci. 164, 877–887.

Rawls, W.J., Gish, T.J., Brakensiek, D.L., 1991. In: Stewart, B.A. (Ed.). Estimating soil water retention from soil physical properties and characteristics. Advances in Soil Science, vol. 16. Springer, New York, pp. 213–234.

Rosenthal, K.M., Ahern, C.R., Cormack, R.S., 1986. WARIS: A computer-based storage and retrieval system for soils and related data. Aust. J. Soil Res. 24, 441–456.

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

Schaap, M.G., Leij, F.J., 2000. Improved prediction of unsaturated hydraulic conductivity with the Mualem-van Genuchten model. Soil Sci. Soc. Am. J. 64, 843–851.

Schaap, M.G., Leij, F.J., Van Genuchten, M.Th., 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Sci. Soc. Am. J. 62, 847–855.

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.

Wösten, J.H.M., Pachepsky, Ya.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251 (3–4), 123–150.