Multi-scale hydraulic pedotransfer functions for Hungarian soils

Promotor:
Prof. Dr. Ir. J. Bouma
Professor of Soil Inventarisation and Land Evaluation
Wageningen University
The Netherlands

Co-Promotors:
Dr. Ir. J.H.M. Wösten
Senior Scientist
Alterra
The Netherlands

Prof. Dr. Gy. Várallyay
Research Professor
Research Institute for Soil Science and Agricultural Chemistry
Budapest, Hungary

Examination Committee:

Dr. Ya. A. Pachepsky,	USDA-ARS, Beltsville, MD, USA
Prof. Dr. Ir. A.K. Bregt,	Wageningen University, The Netherlands
Prof. Dr. Ir. R.A. Feddes,	Wageningen University, The Netherlands
Prof. Dr. Ir. A. Veldkamp,	Wageningen University, The Netherlands

Multi-scale hydraulic pedotransfer functions for Hungarian soils

Attila Nemes

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. Dr. Ir. L. Speelman,
in het openbaar te verdedigen
op dinsdag 1 april 2003
des namiddags te 13.30 uur in de Aula

CIP-data Koninklijke Bibliotheek, Den Haag

Nemes, Attila, 2003

Multi-scale hydraulic pedotransfer functions for Hungarian soils
Doctoral Thesis Wageningen University – with summaries in English and Dutch

ISBN 90-5808-804-9

Acknowledgements

I wish to express my gratitude to many individuals as well as organizations without whom all this work, and the promotion – let's put it this way - would just have been an April 1 joke.

First of all, I would like to thank - with all my heart - one of my co-promotors, Dr. J.H.M. (Henk) Wösten, for his guidance throughout all my studies. Henk was always there; I could knock on his door any time – making no appointments (!!) - with all kinds of questions or problems. His supervision, guidance, suggestions and remarks have helped so much to shape this work. We know each other for ten years (almost by the day at the promotion). What I learned from him over these years helped me to find out about the research topic I was interested in and also guided me to its completion. Thank you Henk!!!

I respectfully acknowledge my promotor Prof. Johan Bouma and co-promotor Prof. George Várallyay who recognized the potential of this work to become a PhD thesis, and who encouraged me all the way to keep heads up, the "end of the tunnel" will be seen after a while. All my respects go to Prof. Várallyay for seeing fantasy in this project, and accepting me first to the PhD program of the University of Gödöllő (currently Szt. István University) and then giving me the opportunity to enjoy the benefits of belonging to the Research Institute for Soil Science and Agricultural Chemistry (RISSAC) of the Hungarian Academy of Sciences. Although the days that I physically spent at the institute could probably be counted quite easily, it did mean a lot to me, more than it probably looked to many people.

Ties between Professors Bouma and Várallyay date back to decades. Prof. Bouma was very aware of research priorities and interests in Hungarian soil and water research. I am most grateful to him for accepting me in the Sandwich PhD Program of Wageningen University. I thank him for doing so much to make this work fit the scientific standards of Wageningen University. I sincerely wish it to everyone to have such a promotor with such permanently positive spirit! No matter what problem I encountered, he always shared a laugh, - said: "Oh, boy!" - and quickly came up with a solution.

If we go back in time, I certainly have to express my gratitude to Prof. György Szalai and most of all to Dr. Béla Nováky, who awarded me with my first TEMPUS fellowship to go to Wageningen, and encouraged me to complete that project to become my MSc thesis, which was later supervised by Dr. Nováky. Although we hardly had any contact later, I want to let him know that he played a crucial role to put me on this track! Thank you very much for the encouragement and motivation!

I of course have to acknowledge more people, when I think of the past. Dr. Erika Micheli of Gödöllő (today Szt. István) University helped me get acquainted with Prof. Eileen Kladivko, Dr. Gregory Willoughby and Prof. William McFee of Purdue University. The periods I spent at Purdue helped me to become internationally oriented.

I thank those who contributed - in one way or another - to receive the almost endless number of fellowships/grants for my studies. In particular, Prof. György Füleky, Dr. Erika Micheli and Prof. George Várallyay had to write so many support letters that it did not take much to convince them that they should sign them in bulk. It went almost the same way with the Invitation Letters and my co-promotor, Henk. Organizations that provided me with fellowships/grants were: TEMPUS, The Hungarian Fellowship Board, the SOROS Foundation, The Ministry of Agriculture, Nature Management and Fisheries of the Netherlands (via the International Agricultural Centre), NUFFIC, Wageningen University, the ANKA Foundation, The European Union and NATO.

Of course I often needed external resources, so I kept on harassing people for data or to teach/help me to take advantage of some methods/programs. Contributors and managers of the UNSODA, HYPRES and BIS (Netherlands Soil Survey Database) databases are greatly acknowledged. Data collected/managed by the research groups of Dr. Walter Rawls and Dr. Lajpat Ahuja have also been used and the access is very appreciated.

While a considerable part of the international data was available with relative ease, I had (I should say we had) a hard time accessing all the necessary data from Hungary. Employees of

RISSAC spent excessive hours to access old, archived information. I first of all offer my thanks to Dr. Kálmán Rajkai and Gyöngyvér Motsai for looking up so much data, sometimes surpassing mountaineers by reaching file cabinets at unusual heights. Dr. Tibor Tóth, Dr. Csilla Farkas (Randriamampianina) and Dr. Erik Czinege provided me with lots of soil information and – most importantly – shared their field data with me; Erik even went out to do extra sampling to help me, for which I really feel indebted. It was my luck, that they did not know at that time how much of their precious time it is going to cost to answer all my questions about those data collections. Many more people contributed – by the original measurements or by finding old files – but unfortunately I don't know them all or I'm simply not aware of the role of many people. Nevertheless, please all of you receive my whole-hearted thanks! I also thank for getting the nickname 'Nessy' within RISSAC. Why? It refers to the Monster of Loch Ness, whom everybody heard of, but who is yet to be seen. I guess I indeed spent short days in there, but getting a nickname often means that they do care about you and what you do. Thank you all for that!

I thank the Hungarian Meteorological Service and especially Dr. Ildikó Wantuchné Dobi for providing all so necessary meteorological data, and the weather generator software.

I have to start the list of people who helped me in terms of methodology with Dr. Marcel Schaap from Riverside. Marcel, although I list you here, we both know – and now others will do too – that this was about more than just neural networks, or data. If one cannot imagine two colleagues become true friends, well, then ask us about it! I express my thanks to the other members of the Dutch "Forces" in Riverside, Dr. M.Th. (Rien) van Genuchten and Dr. Feike Leij for providing me the opportunity to go and work with this group and to learn so much from them.

I thank Dr. Allan Lilly for sharing his knowledge on databases and SQL, Jan Oude Voshaar for his aid with Genstat and statistical problems, Jannes Stolte for his help in getting along with Pascal, and for his laughs almost every time we met on the corridors. The same goes to Dr. Heidi Hamers, Joop Kroes, Erik van den Elsen, Dr. Coen Ritsema and Dr. Louis Dekker, Dr. Jos van Dam, Jan Wesseling and of course the late Jaap Huygen who all helped me in different ways. It may look a small thing to them, but it is all great help when one really needs it! I further thank Dr. Imre Czinkota, Dr. László Tolner and Nándor Fodor for similar help in the 'bloody hard times' as Dr. Tolner would say.

I respectfully acknowledge staff members of the Office for Student Affairs of Wageningen University for all their help in settling residence permit, finance, registration, graduation and other issues during the time I spent in Wageningen. My special thanks go to Ankie Lamberts, Stan van Heijst and Dieuwke Alkema.

Without their names, but with the same gratitude I have to mention here the staff of the Library of Alterra (Bibliotheek de Haaff) and all the managers of the computer system of the Alterra Research Institute (and formerly The Winand Staring Centre) who always helped me when I was in some kind of trouble. And I often was, they can tell! Thank you guys! (A special version goes to Michiel Pieters: Köszönöm!). I also thank the anonymous referees and the responsible editorial staff of all the papers that are presented here, and those 12 colleagues at Alterra who selflessly contributed their computers when a huge load of calculations had to be made. Without your help, it would have lasted for several more months! I also sincerely appreciate the secretarial staff at RISSAC and at the Department of Soil and Land Use at Alterra - the B&L Babes - for their continuous support during the time of my studies. I respectfully thank the management staff of Alterra for providing me with office space, computer facilities and the often-needed help of Alterra employees. I also thank the always-changing community of Hungarian students in Wageningen for their relaxing company!

I further have to thank whoever invented basketball. This was the game that helped me stay fresh mentally by making me tired physically a couple of times a week.

I have to mention Albert Koffeman, who was my roommate at the office during the last months, and kept me alive and amused with his (mis-)belief, that a certain professional soccer player can actually play soccer. Trust me, he cannot! But thanks for the discussions about it, which gave a whole lot of flavor to my days!

Last, but of course most importantly I am most grateful to my parents who encouraged, inspired and supported me to continue my studies – only to give them worries about when I was finally going to finish it - and Anikó, my beloved wife, for being with me, loving me so much, giving her full support and providing such a fantastic atmosphere and environment at home all this time!

Contents

Abstract

Water and nutrient balance are among the main concerns about the sustainability of our soils. Numerous computer models have been developed to simulate soil water and solute transport and plant growth. However, use of these models has often been limited by lack of accurate input parameters. Often, the limiting input parameters are water retention and hydraulic conductivity. For many applications, estimation of soil hydraulic characteristics with pedotransfer functions (PTFs) may offer an alternative to costly and troublesome field or laboratory measurements.

Many environmental problems are not restricted to national boundaries and therefore solutions require international co-operation. Soil particle-size distribution (PSD) is among the key predictors to most soil hydraulic PTFs. However, despite a number of recognised international standards, those data are rarely compatible across national frontiers, which hinders the establishment of international databases. The performance of four different interpolation procedures was evaluated to achieve compatibility of PSD data. Based on the number and distribution of measured points on the cumulative PSD curve, a general rule was formulated as when to fit a spline function or use a novel 'similarity procedure' to estimate missing values. The 'similarity procedure' uses an external source of soil information from which soils are selected with PSD that match the distribution of the soil under investigation. Fitting a non-parametric spline function to the PSD data showed similar accuracy and is independent of any external data sources, however, it is more sensitive to low data density on the PSD curve.

International soil hydraulic databases can be valuable alternatives to smaller (e.g. national) databases when seeking solutions to soil water management related problems with international significance. Focusing on soil hydraulic data, the international UNsaturated SOil hydraulic DAtabase (UNSODA) was developed. The format, structure and operation of the redesigned second version of the database (UNSODA V2.0) were discussed. The 'similarity procedure' for the interpolation of PSD was tested for it's validity for soils of different geographical areas using data of UNSODA V2.0. While applying the procedure, a large European data set was used to help interpolate PSD of non-European soils. It was hypothesized that the procedure would not perform equally well for soils of other geographical areas, however, the study rejected this hypothesis.

It was recognised that many soil hydraulic data have also been collected in Hungary in the framework of various independent projects. However, only part of those were stored in a common database. The HUNSODA database is introduced, which serves as storage for soil physical and hydraulic data as well as basic soil information on 840 soil horizons. The structure, contents and basic operations of the database are discussed. Data were then used to develop class pedotransfer functions for Hungarian soils according to both the FAO and USDA soil texture classes. There is considerable overlap between average curves of different classes if many small classes are distinguished. A less detailed classification system yields more distinction between class functions.

In most PTF comparisons it remains unclear what the main sources of the estimation errors are. National, continental and intercontinental scale data collections were used to derive PTFs to estimate soil water retention. The same methodology (neural network model) and the same sets of predictors were used to allow the source database to be the only variable that is changed. We evaluated the performance of 11 different PTFs developed from each of the data sets, in order to study the influence of different

combinations of predictors. All PTFs were tested using independent Hungarian data. Water retention estimations showed an improving trend as the list of input variables increased. Using a small set of relevant (local) data – when available - is better than using a large but more general data set. Estimated water retention curves (WRCs) were then used to simulate soil moisture time series of seven Hungarian soils. Small differences were found among the PTFs derived from different scale data collections. PTF estimates were only marginally worse than estimates using laboratory measured WRCs. Differences between estimations using different scale data sets (or measured WRCs) were not the main source of error in the simulation model.

Scenario studies can be used in planning and prevention to fine-tune expert knowledge, as those give a quantitative dimension to the outcome of implemented changes. PTFs can provide important soil physical data at relatively low costs and at no risk to the environment. Three scenario studies demonstrate how PTFs may help improve land management planning. Exploratory modeling was used to give estimates of (i) benefits and risks associated with irrigating a heterogeneous field, in terms of soil water balance components and the soil's ability to supply the vegetation with water, (ii) the effect of changes in physical properties of a soil on flux balance and plant water deficit, and (iii) the risk of leaching to deeper horizons in different soil types of Hungary, based on the simulation of 20 years. Using an exploratory research approach, quantitative answers were provided to "what if" type of questions, allowing the distinction of trends in potential problems and successes.

Chapter 1

General Introduction

General Introduction

Rational soil management and land use are important elements of sustainable agricultural development, and environmental management. Soil is the medium, which supports plants with water and essential nutrients. It is known that one of the main concerns about the sustainability of our soils is the proper knowledge and management of their water and nutrient balance (Várallyay, 1989a). Different agricultural management practices have been developed and applied to optimise aeration and water supply capacity of soils throughout the entire growing season. Tillage, irrigation and drainage are major parts of the equation while seeking to avoid (water)stress for our crops, whether it is due to too much or too little water.

In the meantime, increasing agricultural and industrial activities require an intensive supply of nutrient reserves but at the same time threaten soils by being polluted with various types of chemicals. Movement of chemicals and nutrients is basically directed by soil-water in the form of solutes. There is an important role of all those factors, properties and processes, which influence the water- and therefore the nutrient (and pollutant) transport in the soils (Várallyay, 1989a).

Soils are particularly important natural resources and contribute significantly to the national economy in Hungary (Várallyay, 1987). Their rational utilisation, conservation and maintenance of their functionality have particular significance in the economy and in environment protection, as almost 70% of the area of the country is utilisable from an agricultural point of view (Láng, 1994).

Earlier studies showed that while agricultural production, human and industrial use requires an increasing amount of water, water reserves that can be devoted to these don't increase sustainably (Várallyay, 1994). Moreover, due to its basin characteristics Hungary probably faces the problem of decreasing water resources. Until any alternative solution is found, increasing the efficiency of water use – and thus the use of soil water as well - may be the tool to resolve the problem of the difference between *needs* and *resources* of water (Várallyay, 1987). To achieve that, we need to have proper understanding of the spatial and temporal variability of soil properties and processes, which are very complex. Following are some examples that show the complicated situation, soil water management experts and farmers have to face in Hungary during a growing season.

Potential evapotranspiration at the main agricultural area of Hungary – The Great Hungarian Plain – may exceed 700 mm per year, whereas annual rainfall adds up to 450-600 mm only (National Plan for Water Management, 1964; Stefanovits, 1963). Moreover, areas with less precipitation tend to have higher evapotranspiration. This means, that on a major part of the area there is a potential risk of drought each year. However, even that amount of precipitation is not evenly distributed, neither spatially nor temporally. This creates problems of two opposite kinds. In the late summer period, the amount of precipitation is very low, and due to relatively high temperatures and high proportion of sunny days plants are stressed by drought. The irrigation potential is limited by many factors, among others by the relief and the limited amount of good-quality irrigation water (Szalai, 1989). In early spring the problem is the opposite. As winter precipitation is much higher – and it partially falls as snow - it is collected and stored in and on top of the soil. In this case the amount of water that exceeds the infiltration rate and water holding capacity of the soils, appears as inland water after melting. Thus, in the early season, low aeration for the plants, and conditions that make fieldwork difficult - if not impossible - hamper agriculture. In the summer period, dry conditions, heat and drought may cause stress for all the crops. As there are few feasible options – if any at all – to manage this situation, the main focus should be on prevention and preparation for this situation.

Soils of a considerable part of Hungary have high salt contents on the surface and/or in deeper horizons. About one-third of the territory of the Great Hungarian Plain is affected by salinity/sodicity, and another one-third is potentially salt affected (Tóth et al., 2001), meaning, that those are not affected at present, but could become considerably saline or sodic as a consequence of irrigation (Szabolcs 1974). The threat of re-mobilisation and re-distribution of some salts is of major concern and has been extensively studied to have a better understanding and to avoid undesirable changes (e.g. Szabolcs et al., 1969; Tóth and Blaskó, 1998).

Regarding environmental concerns, the change of water table depth may also cause different hazards. This may need special attention when we consider the currently often changing political and/or social situation of the Central and Eastern European region. With these changes, situations occur when major disputes about hydrological issues involving border-rivers or their catchment area are not settled and not concluded in an agreement. Systematic change in the levels of surface and/or subsurface water reservoirs may – in extreme cases – cause complete environmental changes at some locations by changing ground water table levels (Várallyay, 1989a). With dropping groundwater levels, underlying coarse material may not (and will not) provide the possibility of groundwater recharge to the overlaying soils, drastically decreasing the amount of available soil water.

Predicted climatic change may also influence the water balance of several regions, and so may cause similar changes as referred to above (e.g. Nováky et al., 1996). Whether it is only a different distribution of rainfall and/or temperature characteristics, or it is indeed a change in the amount of precipitation or in average temperature, one way or another it will be reflected in water management of soils.

These are all situations we may have to manage. To avoid or to prepare for any of the above situations, we need to understand a very complex system of environmental factors. Computer models are tools, which help to understand and describe such environmental phenomena and underlying mechanisms. There are an increasing number and variety of numerical computer models that are being developed throughout the World to simulate soil water and solute transport and plant growth (Kabat and Hack-ten Broeke, 1989; Tóth and Blaskó, 1998). However, use of these models has often been limited by the lack of accurate input parameters (e.g. Wösten et al., 2001).

Among the key input parameters of these models are the soil hydraulic properties, such as water retention and hydraulic conductivity. Because most models greatly depend on accurate information on these parameters much work has been done in the measurement area (cf. Gee and Ward, 1999). Several methods are now available to obtain measurements under field or laboratory conditions. Despite all the efforts, the fact remains that the hydraulic parameters are still very difficult to measure accurately, especially for undisturbed field soils.

For many applications, prediction of soil hydraulic properties with pedotransfer functions (PTFs) (Bouma and van Lanen, 1987) may offer a competitive alternative to the costly and troublesome measurements. Pedotransfer functions relate hydraulic parameters with more easily measured soil data such as soil texture, organic matter content and/or other data routinely measured or registered in soil surveys. However, the capability to derive pedotransfer functions requires the establishment and use of a comprehensive database of soil hydrological and pedological data.

Problem definition

In a number of countries, different sizes of databases on soil hydraulic properties have been established and their analysis has resulted in a number of different pedotransfer functions (PTFs). Different approaches have been investigated to derive PTFs. Distinct points on the water retention curve have been estimated by some PTFs (e.g. Pachepsky et al., 1996; Rajkai and Kabos, 1999), whereas others predict parameters of the parameterised water retention curve (Vereecken et al., 1989; Wösten et al., 1994; Schaap et al., 1998). Different regression functions have been developed which were later supplemented with other techniques to improve their predictive capabilities. As an alternative to the conventional regression type PTFs, other possibilities like optimisation with artificial neural network (ANN) models are being explored to derive soil hydraulic PTFs (Pachapsky et al., 1996; Schaap and Bouten, 1996; Tamari et al., 1996). Limitations of the PTFs have also been identified by several studies (e.g. Bastet et al., 1997; Minasny et al., 1999). Database dependency was also found, so that the application of these functions was reliable only for the area that the data source database represents (e.g. Schaap and Leij, 1998). The above may suggest, that each interested country or region must establish its own database to be able to develop reliable PTFs.

Many studies have been conducted in Hungary also to answer questions of local soil hydrological phenomena (e.g. Várallyay et al., 1979; Várallyay et al., 1980; Rajkai, 1984; Rajkai et al., 1981; Rajkai and Kabos, 1999). It is recognised that lots of soil hydraulic data have been collected in the framework of various independent projects. However not all of those were stored in a common database. Available Hungarian data were fragmented, of different degree of detail and could only be gathered from various different sources. Therefore those were neither easily accessible for present/future researchers nor were then suitable for many types of applications. So far, in Hungary, soil hydraulic pedotransfer functions have been derived using a smaller dataset (e.g. Pachepsky et al., 1996; Rajkai et al., 1981; Rajkai and Kabos 1999). Establishment of an extensive national database, and derivation of pedotransfer functions would presumably offer an easier source of soil hydraulic information for future research and applications on a national level.

Environmental problems however do not consider manmade national boundaries and therefore often require international co-operation to find solutions. International collaboration increases in solving environmental problems and in preventing new problems to occur (e.g. van Dam et al., 1994; Batjes and van Engelen, 1997). International soil hydraulic databases are seen to be valuable alternatives to separate smaller (e.g. national) databases when answering research questions that are related to soil water management. Recently, data of 12 European countries have been collected and stored in the HYPRES (HYdraulic PRoperties of European Soils) Database (Wösten et al., 1999). Pedotransfer functions derived from the HYPRES database – using regression techniques - are already being used in a number of different studies within the EU. A comparable study is underway in the USA where the UNSODA multi-national database of soil hydraulic properties was created (Leij et al., 1996).

Although pedotransfer functions derived from international databases are being used in various studies, many questions remain regarding the range and scale of studies they can be used for, as well as regarding the prediction accuracy they provide. What is the level of benefits from international databases for individual countries? To what extent can the different scale PTFs be utilised? What are their limitations? How can we take advantage of the different scale PTFs? These are questions that have not been studied extensively so far and they all directly relate to the central question, whether or not the establishment of national pedotransfer functions is justified considering the availability of operational, international databases and derived pedotransfer functions.

Objectives

For a number of applications, soil hydraulic PTFs may replace expensive directly measured soil hydraulic input data, such as water retention and hydraulic conductivity.

The aim of this research is to explore the reliability of using pedotransfer functions for simulating soil hydraulic processes that are associated with various forms of land use. The study concentrates on the information content of international databases which will be tested for national-scale applications. We test the hypothesis that PTFs developed from an international database cannot be used at smaller, national scale, and so country- or region-specific PTFs are necessary. The utility of different scale databases will be demonstrated through the establishment of new databases and the analysis of existing databases.

Pedotransfer functions are derived using the same neural network technique, and using the same sets of input data. This allows the source database to be the only factor that is variable, in terms of its source area, representativity in physical properties and its size. The performance of international scale PTFs is tested in comparison with national scale PTFs and laboratory measurements.

Functional evaluation will be performed that allows insight into the contribution of PTFs to the inaccuracy and uncertainty of simulations. Subsequently, using estimates of some of the best PTFs, ways PTFs can be applied will also be demonstrated. Expected effect of changes in soil or land management or climate will be explored for some major soil types of Hungary, with the objective to demonstrate the possibilities for practical use of PTFs. Another example will demonstrate how PTFs can be used to aid optimal irrigation water usage.

Outline of Thesis

Chapter 2 outlines a procedure to harmonise soil texture data. Countries contributing data to international databases often use different classification systems, among others, in the determination of soil texture. Lack of reliable methods to interpolate the particle-size distribution often hinders progress in international collaboration. A new interpolation procedure has been developed to enable advanced utilisation of the European HYPRES database (Wösten et al., 1999).

Chapter 3 summarizes the development, structure and contents of the novel SQL-based version of the UNSODA international database. Contents of the database have already served as source data to a range of different studies. This new version enables more detailed analysis of the database contents, and features advanced search, selection and extraction capabilities. UNSODA is further used to test whether the novel interpolation method – presented in Chapter 1 - is dependent on the geographical origin of soils.

Chapter 4 outlines a newly developed Hungarian database of soil hydraulic functions. The database holds an extended amount of data in comparison with previous data sets that were used in relevant studies. The database has been developed to suit both Hungarian and frequently used international systems. It is a searchable SQL-based database, that is compatible to many international databases and applications.

Chapter 5 describes an extensive comparison of pedotransfer functions. PTFs were derived – using the same methodology – from two international and the national scale database to study their accuracy and utility at a national scale. Predicted water retention curves are evaluated in comparison with measured curves. Connections between prediction accuracy and the uncertainty in soil representation in the source databases regarding soil physical types are also examined. Predicted functions are further used to simulate soil moisture profiles of a number of Hungarian soils.

Chapter 6 then shows possible examples for future use of soil hydraulic pedotransfer functions. Irrigation, tillage and nutrient/pesticide leaching are all important keywords in today's agriculture and environmental management. Three different exploratory studies show how the use of PTFs may help improve land management practices, thereby justifying their development.

Chapter 2

Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases

Published in: Geoderma 90: 187-202. (1999)

Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases

A. Nemes[1], J.H.M. Wösten[2], A. Lilly[3] and J.H. Oude Voshaar[4]

[1] Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Herman Ottó u. 15, H-1022 Budapest, Hungary

[2] ALTERRA Research Institute for Green World Research [formerly: DLO Winand Staring Centre, (SC-DLO)], P.O. Box 47, 6700 AA Wageningen, The Netherlands

[3] Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen AB15 8QH Scotland

[4] Centre for Biometry Wageningen, P.O. Box 16, 6700 AA Wageningen, The Netherlands

Abstract

Many environmental and agricultural problems are not restricted to national boundaries and therefore require international co-operation if solutions are to be found. Often, these solutions require the ability to use soil data as input in simulation models, however, despite a number of recognised international standards, soil data are rarely compatible across national frontiers. This problem was encountered when creating the HYdraulic PRoperties of European Soils (HYPRES) database. The data, which includes particle-size distributions, were collected from 20 institutions in 12 countries. Only few of these institutions adhered strictly to a recognised international system such as FAO. Therefore, interpolation of the cumulative particle-size distribution was required to achieve compatibility of particle-size distributions within the HYPRES database.

In this study, four different interpolation procedures were evaluated. The accuracy of the different procedures was found to vary with size intervals between measured points of the particle-size distribution. The loglinear interpolation of the cumulative particle-size distribution has previously been used in various studies but was found to give the least accurate estimation of the four procedures. Fitting the Gompertz curve, which is a special non-symmetrical type of curve described by a closed-form equation, showed less sensitivity to size intervals between measured points. However, interpolation within some of the particle-size distributions was not sufficiently accurate and this procedure could not be applied to particle-size distributions where the number of measured size fractions was less than the number of model parameters. Fitting a non-parametric spline function to the particle-size distributions showed a considerable increase in accuracy of the interpolation with decreasing size intervals between measured points. As a novel approach, the similarity procedure was introduced which does not use any mathematical interpolation functions. It uses an external source of soil information from which soils are selected with particle-size distributions that match the distribution of the soil under investigation. This similarity procedure was capable of giving the most accurate interpolations. Once an extensive external reference data set with well-quantified particle-size distributions is available, the similarity procedure becomes a very powerful tool for interpolations. Based on the number and distribution of measured points on the particle-size distributions, a general rule was formulated as when to fit a spline function or use the novel similarity procedure to estimate missing values. Results of this study were used to classify all soils in the HYPRES database into the same soil texture classes used in the 1:1.000.000 scale Soil Geographical Database of Europe.

Keywords: texture description, estimation procedure, spline, soil-texture similarity, standardisation

Introduction

Increasingly, environmental problems cross national frontiers and therefore require international co-operation if solutions are to be found (Eijsackers and Hamers, 1993). A prerequisite to overcome these problems is the existence of international databases with environmental and biophysical data. The HYdraulic PRoperties of European Soils (HYPRES) database is such a database (Wösten et al., 1999). This database drew together existing soil hydraulic properties and more easily measurable soil data such as soil texture, from 20 institutions within 12 European countries. These data were used to derive both continuous and class pedotransfer functions for the estimation of soil hydraulic properties, which are input data for soil water transport models or EU-wide land evaluation systems. HYPRES was designed to link with other existing EU soil databases, in particular with the 1:1.000.000 scale Soil Geographical Database of Europe (Jamagne et al., 1994). As this geographical database uses the FAO system for defining soil texture (Figure 1), the pedotransfer functions derived from the HYPRES database had to adhere to this system.

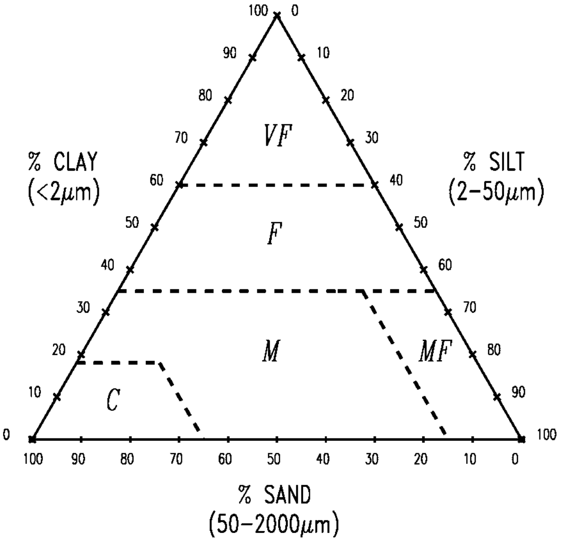


Figure 1 - Definition of soil texture classes according to the Soil Geographical Database of Europe. (C = coarse, M = medium, MF = medium-fine, F = fine, VF = very fine).

FAO (1990) and USDA (1951) define clay as the particle-size fraction < 2µm, silt as the fraction between 2µm and 50µm and sand as the fraction between 50µm and 2000µm. However, only a few countries who contributed data to HYPRES, use these particle-size classes in their national classification systems for soil texture. In this case standardisation of particle-size description is required to achieve compatibility of soil data among different countries and to be able to classify these data in the FAO/USDA system. Lack of reliable methods for particle-size standardisation has resulted in formation of international soil databases that could not make use of data which did not comply with the FAO/USDA particle-size description (Batjes, 1996). An additional reason for particle-size standardisation is the attractiveness of using continuous pedotransfer functions as estimators of soil hydraulic properties. Figure 2 shows the particle-size classes used by the countries that contributed data to HYPRES. In general, there was a need to estimate the fractions of particles at 50µm from data where fractions were measured at particle-size limits of 60µm, 63µm, 200µm or 2000µm succeeding the 20µm limit.

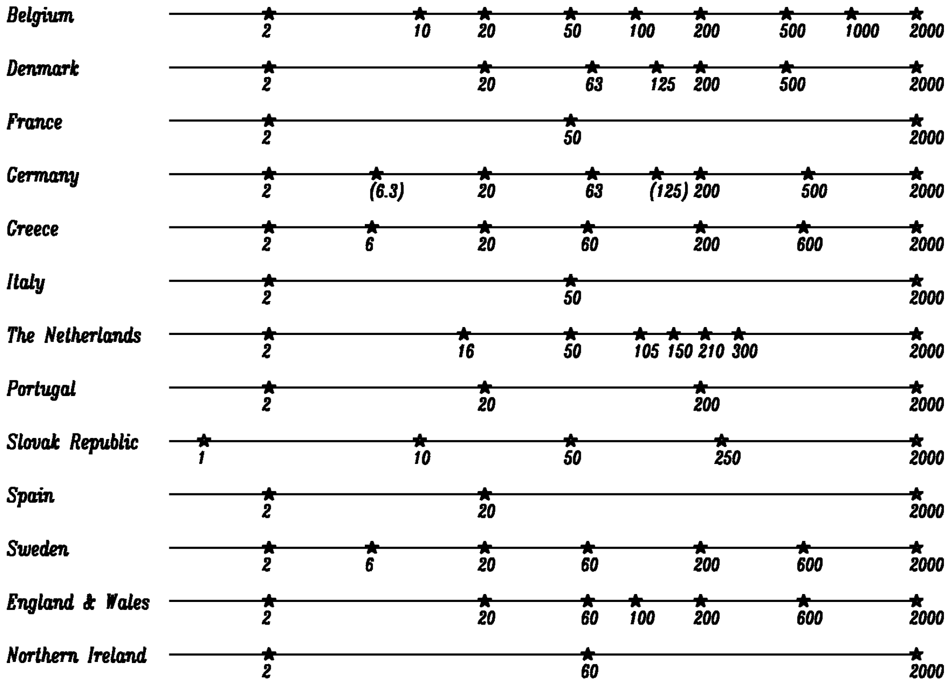


Figure 2 - Particle-size limits used by countries contributing data to the HYPRES database.

Several studies suggest that the particle-size distributions in soils show an approximately lognormal distribution (Campbell, 1985; Shirazi and Boersma 1984). However, soils with bimodal particle-size distributions also do occur (Walker and Chittleborough, 1986). Buchan (1989a) described the applicability of lognormal models for particle-size distributions and found that these are only applicable for about half of the USDA soil texture classes (Soil Survey Staff, 1975). He also discussed the effects of the number of particle-size fractions that are measured on the shape of the cumulative particle-size distribution. The more complex the cumulative distribution is, the greater the number of required model parameters are. Rousseva (1987) applied two different techniques (graph and polynomial fit) to transform particle-size distributions from Katchinski's texture scheme (Katchinski, 1956) to the scheme used by the United States Department of Agriculture (USDA) (Soil Survey Staff, 1975). She concluded that polynomial fits do not convert soil texture data adequately and that use of graphs is better, even though it is time-consuming, laborious and subjective. Rousseva (1997) defined closed-form models of exponential and power law. She investigated the suitability of these models to fit cumulative particle-size distributions of different shapes and with varying numbers of measured points. Suitability of the models appeared to be influenced by texture type (coarse or fine textured soils) rather than by measured size ranges. Shirazi et al. (1988) established connections between texture classifications adopted by the USDA (Soil Survey Staff, 1975), the International Society of Soil Science (ISSS) (International Society of Soil Science, 1929) and the American Society of Civil Engineers (Vanoni, 1980). The work of Shirazi et al. was based on a description of the clay, silt and sand fractions by the geometric mean and the geometric standard deviation of their size ranges. Later, applicability of this method was questioned by among others Buchan (1989b) and Yaalon (1989). Buchan et al. (1993) compared five different lognormal models for soil particle-size distributions. All five models accounted for more than 90% of the variance in the particle-size distribution of most of the examined soils. However, the algorithm did not converge for about 10% of the soils in their study.

In practise, the loglinear interpolation has often been used to estimate missing particle-size classes for the FAO/USDA texture classification (e.g. Tietje and Hennings, 1996). However, alternatives do exist and therefore the specific objectives of this study can be summarised as:

- Identify and describe four different realistic procedures for interpolation of the particle-size distribution.
- Evaluate the accuracy of these procedures using data from the HYPRES database and from the Soil Information System of the Netherlands (Finke, 1995).

Materials and methods

Test and reference data sets

Particle-size data were extracted from both the HYPRES database and the Soil Information System of the Netherlands to develop and test procedures for interpolation of particle-size fractions in the 2-50 μm range. The selected data were stratified into three groups: a reference data set and two test data sets.

- *The Reference data set* was created by randomly extracting data from the Soil Information System of the Netherlands and can be considered representative for the overall soil database of the Netherlands. The reference data set contains 9607 individual soil horizons and was used to develop the similarity procedure, one of the interpolation procedures evaluated in this study. The particle-size distribution of the soils in the reference data set encompasses a wide range of soil textures (Figure 3).

- *Test data set 1* comprises the remaining data from the Soil Information System of the Netherlands and contains 3453 individual soil horizons.

- *Test data set 2* was extracted from the HYPRES database and contains 1524 individual soil horizons all originating from Germany. These horizons were selected because they had well quantified cumulative particle-size distributions based on particle-size fraction measurements for numerous particle-size limits.

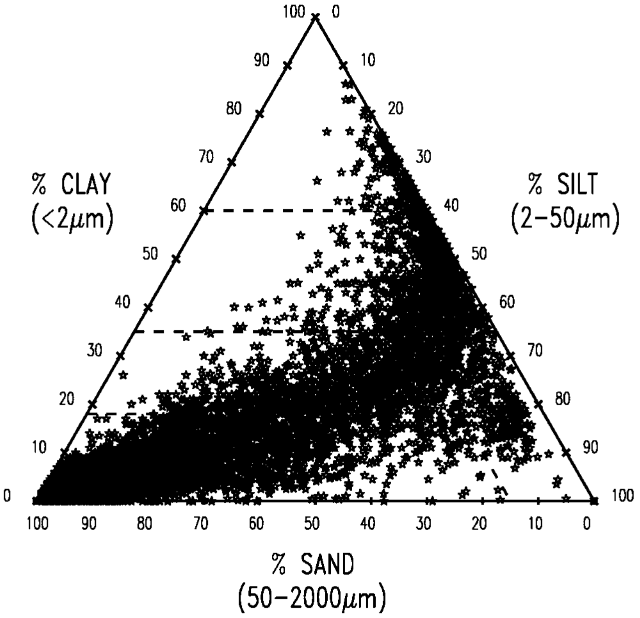


Figure 3 Textural composition of the reference data set.

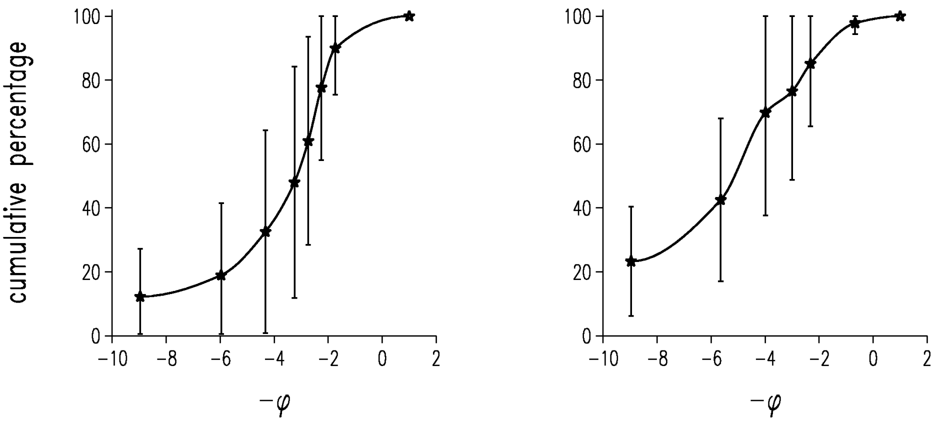


Figure 4 Average cumulative particle-size distributions of the 2 test data sets. Vertical bars indicate ± one standard deviation from the arithmetic mean.

Distribution of soils in test data sets 1 and 2 are considerably different as shown by the two average cumulative particle-size distributions and standard deviations of their measured particle-size fractions (Figure 4). These average cumulative distributions were obtained by simple arithmetic averaging of the individual cumulative distributions. Soils in data set 2 have considerably finer textures with greater average clay and silt contents than soils in data set 1. Large standard deviations indicate the large variety of soils represented in each data set. Some of the soils in data set 2 exhibit a degree of bimodality as indicated by deviation of the particle-size distribution from the expected smooth line. This is most apparent for the 63 μm point ($-\phi = -3.988$). It is not possible to directly compare the soils in the two data sets by plotting them in the FAO texture triangle because the particle-size classes of data set 2 are not compatible.

Estimation procedures

Four interpolation procedures were evaluated using the two test data sets. This required that particle-size distributions were transformed to a logarithmic scale. In accordance with what is common in sedimentology, the ϕ scale is used (e.g. Walker and Chittleborough, 1986; Moran et al., 1988), where ϕ is defined as $-\log_2$(particle-size in mm). We defined variables $\Delta\phi$ and α (Figure 5) which both were used in testing and evaluating the efficiency of the interpolation procedures. $\Delta\phi$ is the distance on the ϕ scale between the two particle-size limits neighbouring the limit for which the fraction has to be estimated and α is the distance between the cumulative percentages of these two particle-size limits.

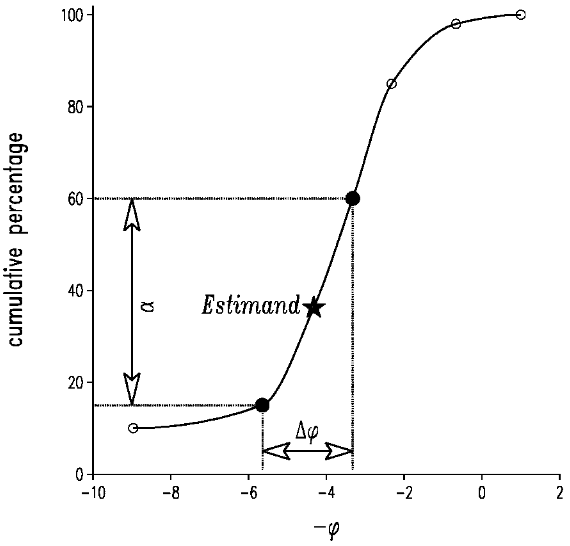


Figure 5 Representation of distances between particle-size fractions (α) and between particle-size limits ($\Delta\phi$). Dots represent measured values and the star represents an estimated value.

1. Loglinear interpolation

A loglinear interpolation of the cumulative particle-size distribution curve has often been used (e.g. Leij et al., 1996; Tietje and Hennings, 1996). In this procedure, the estimation of the cumulative fraction of a particular particle-size limit was based on the assumption of a loglinear relationship between two neighbouring particle-size limits with measured fractions. Mathematical notation of the loglinear interpolation on the ϕ scale is given in equation [1]:

$$CP_n = CP_{n-1} + \frac{(-\phi_n) - (-\phi_{n-1})}{(-\phi_{n+1}) - (-\phi_{n-1})} (CP_{n+1} - CP_{n-1}) \qquad [1]$$

where CP is the cumulative percentage on the particle-size distribution curve, $-\phi$ is the $\log_2$ value of the particle-size limits (in mm) and notations *n*, *n-1*, *n+1* are the missing particle-size limit, the preceding neighbouring limit and the succeeding neighbouring limit respectively.

2. Gompertz procedure

The Gompertz curve (Johnson and Kotz, 1970) which is a special case of the more general logistic curve, was also fitted to the measured cumulative particle-size distribution. Unlike the logistic curve, the Gompertz curve is described by a non-symmetrical closed-form equation. Application in this study of the non-symmetrical Gompertz curve was justified by preliminary investigations. They showed that the Gompertz function was superior to the lognormal distribution function, which is symmetrical on the log scale. The mathematical notation for the Gompertz curve is given in equation [2]:

$$y_i = \alpha + \gamma \exp(-\exp(-\beta(x_i - \mu))) \qquad [2]$$

where α, β, γ and μ are the shape parameters of the curve. If $\beta > 0$ then the curve is called *right sided.* This results in a steeper rise of the monotonous curve close to the lower asymptote. Conversely, if $\beta < 0$ it means a steeper rise close to the upper asymptote which is called *left sided.* Preliminary investigations showed that, in most cases, the *left sided* curve gave the best fit to the measured points. As a consequence, investigations were restricted to the use of the left sided curve.

3. Spline procedure

Fitting a highly flexible non-parametric spline was also tested. Smoothing splines are complicated functions constructed from segments of cubic polynomials between distinct values of a variable which are constrained by being "smooth" at the junctions. Parameterisation of the fitted smooth curve is a complicated process in which the degree of smoothness can be controlled (Hastie and Tibshirani, 1990). Various maximum degrees of smoothing splines were previously examined in this study. It proved that in general, sixth order splines were sufficient accurate. Further increase in the order of splines did not result in any significant increase in accuracy, whereas allowing a smaller maximum degree of smoothness resulted in considerable loss of accuracy.

4. Similarity procedure using a large external reference data set

A new approach called 'similarity procedure' was also tested. Characteristic of this procedure is that it does not use any mathematical interpolation function. In this procedure, use is crucial of an external reference data set containing several (7 or 8) measured particle-size fractions for a wide variety of soils. The procedure involves searching for soils in the external reference data set that have particle-size distributions that are similar to the particle-size distribution of the soil for which a particular particle-size class is missing. Figure 6 gives a schematic representation of the procedure. Suppose that for a soil with 5 known fractions at particle-sizes limits $x_1,\ldots x_5$, the unknown fraction at particle-size limit A has to be estimated. At the same time, a large external reference data set of soils with known fractions at the particle-sizes limits $x_1,\ldots x_5$ and A is available. From this reference data set 10 soils are selected that have fractions at the particle-size limits $x_1,\ldots x_5$ that are similar to the fractions of the soil under investigation. More formally, soils are selected with lowest values for maximum differences in fractions at the common particle-size limits $x_1,\ldots x_5$. The arithmetic mean of the fractions of these 10 soils at particle-size limit A is calculated and assigned as estimation of the fraction at particle-size limit A for the soil under investigation. In fact information on soils in the external reference data set is searched for soils that match the soil under investigation. A prerequisite for this procedure is that the data set is large and covers a wide range of soil types, otherwise not always enough matching soils can be found. In selecting the 10 most similar soils, the maximum differences in fractions at each of the particle-size limits almost never exceeded ±0.5%.

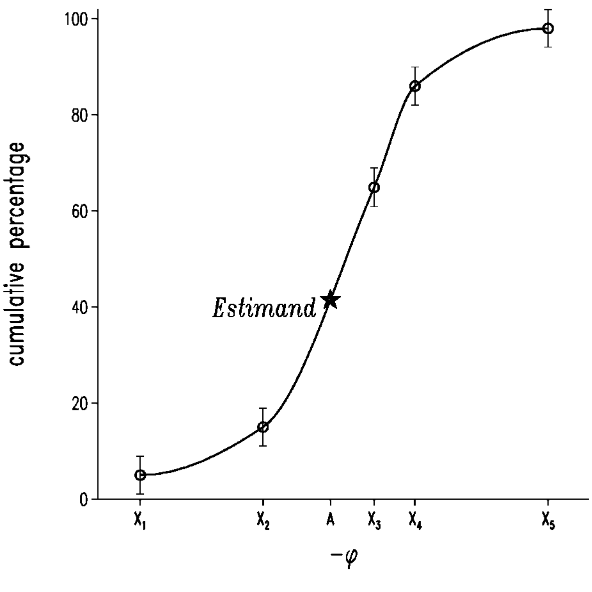


Figure 6 Visualisation of the similarity procedure where $x_1,\ldots x_5$ indicate particle-size limits for which particle-size fractions have been measured. 'A' represents the particle-size limit for which the particle-size fraction has to be estimated.

Evaluation of interpolation procedures

Accuracy of the four interpolation procedures was evaluated by systematically omitting selected points from the particle-size distributions in the test data sets. Each interpolation technique was then used to estimate these missing values (leave-one-out method). Estimated values were compared with omitted measured values and both the mean error (ME) and the root mean squared error (RMSE) were calculated according to equations [3] and [4] respectively:

$$ME = (1/n)\sum_{i=1}^{n}(x_i - y_i) \quad [3]$$

$$RMSE = \sqrt{(1/n)\sum_{i=1}^{n}(x_i - y_i)^2} \quad [4]$$

where n = number of estimated and measured values, x = measured particle-size fraction, y = estimated particle-size fraction.

As the particle-size limits were not uniformly distributed over the cumulative particle-size distribution, $\Delta\phi$ between the two limits neighbouring the estimation varied. This allowed evaluation of accuracy of interpolation procedures in estimating particle-size fractions at any particle-size limits on the curve from closely spaced (i.e. small $\Delta\phi$) to widely spaced (i.e. large $\Delta\phi$) limits.

Results

Accuracy of all procedures depends highly on the spacing of the particle-size limits for which fractions have been measured. Table 1 shows the RMSE and ME values for the two test data sets and the four procedures. In practically all cases, loglinear interpolation had the largest RMSE values for given $\Delta\phi$ values and was therefore the least accurate estimator. Using the non-parametric spline procedure gave better estimations than the loglinear interpolation procedure, however, accuracy of estimations using splines greatly decreased with increasing $\Delta\phi$. If the number of measured points was limited to for example 3 or 4, RMSE became unacceptable high and no further testing was done. The Gompertz procedure showed less variation in accuracy with varying $\Delta\phi$ compared to the spline procedure. For test data set 1, the Gompertz procedure performed better than splines in cases where the particle-size limits around the estimation were widely spaced (large $\Delta\phi$). However, both performed equally well in case particle-size limits were closely spaced. Splines gave better estimations than the less flexible Gompertz curve for small $\Delta\phi$ values in test data set 2. In general, the similarity procedure yielded the lowest RMSE values of all four procedures tested. However, for data set 2 the similarity procedure yielded uncharacteristically high RMSE and ME values in case $\Delta\phi$ was either 2.64 or 3.32. This is attributed to the more bimodal particle-size distribution in data set 2. The mean errors (ME) show a similar pattern to the RMSE. The loglinear procedure introduced the highest bias (consistent under or overestimation), followed by the spline procedure and the Gompertz procedure. In some cases the spline procedure gave considerably higher bias than the Gompertz procedure, while in other cases it was lower. Again, the similarity procedure yielded the lowest bias in most cases.

Table 1 Results of the sensitivity analysis for the 4 interpolation procedures.

Variable $\Delta\phi$	Procedures							
	Loglinear		Gompertz		Spline		Similarity	
	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME
Test data set 1								
1.00	6.21	-1.52	4.07	-0.26	4.44	-1.07	3.69	-0.59
1.59	8.83	-4.41	4.61	-2.02	5.17	-1.11	3.50	0.05
2.07	15.30	-8.08	7.48	-3.52	11.77	-6.15	6.20	-0.08
3.23	14.70	-6.54	8.74	2.20	10.48	0.52	7.31	-0.65
3.71	19.73	-10.57	9.62	0.34	15.20	-5.41	8.27	-0.76
4.63	10.79	-6.71	a	a	a	a	3.28	0.35
5.71	13.21	-7.02	a	a	a	a	8.22	-0.40
6.23	22.34	-14.40	a	a	a	a	10.34	-0.59
6.71	31.54	-22.37	a	a	a	a	11.39	-0.91
Test data set 2.								
1.67	5.69	-1.42	7.91	-0.52	4.74	-1.74	2.79	-0.56
2.64	11.52	4.99	9.15	4.04	8.90	5.02	9.37	4.86
3.32	16.54	5.50	12.34	5.34	14.09	5.18	11.13	5.47
4.98	17.06	-11.32	a	a	a	a	6.15	-1.86
5.97	16.13	-10.64	a	a	a	a	6.62	1.98
6.64	11.10	-1.53	a	a	a	a	6.12	1.88

[a] indicates where either a procedure was not applicable or where RMSE or ME indicated highly inaccurate fits.

Figure 7 shows the relationship between increasing α values and estimation errors for one specific $\Delta\phi$ value using test data set 1. The figure shows increasing estimation errors with increasing distance (α value) between particle-size fractions neighbouring the measured/estimated particle-size fraction. There appears to be little bias in estimation errors although the Gompertz procedure resulted in a slight underestimation (Figure 7). While the Gompertz, spline and similarity procedures produced the lowest errors at small α values, the loglinear procedure was less stable and shows a clear trend of increasing errors with increasing α values over the full range. At larger α values, the similarity procedure yielded the lowest error. A similar trend is observed for other $\Delta\phi$ values.

Both the similarity and spline procedure were used to achieve compatibility of particle-size data of soils in HYPRES with the FAO texture system. For each country that uses a national texture classification system that differs from the FAO texture system, one interpolation procedure was applied. For countries that have only 3 measured points on the particle-size distribution (e.g. Spain, Scotland and Northern Ireland) the spline procedure was highly inaccurate. The Gompertz procedure could not be used either as the Gompertz curve is described by 4 parameters. The similarity procedure was superior in these cases. The similarity procedure was also applied for interpolating data from Portugal and the Slovak Republic. The spline procedure was used for the Danish, German, Greek, Swedish and English data. The standardised particle-size classes were then used to group each soil horizon in the HYPRES database into one of the 5 texture classes used in the 1:1.000.000 Soil Geographical Database of Europe (Jamagne et al., 1994). The final result of this classification is shown in Figure 8.

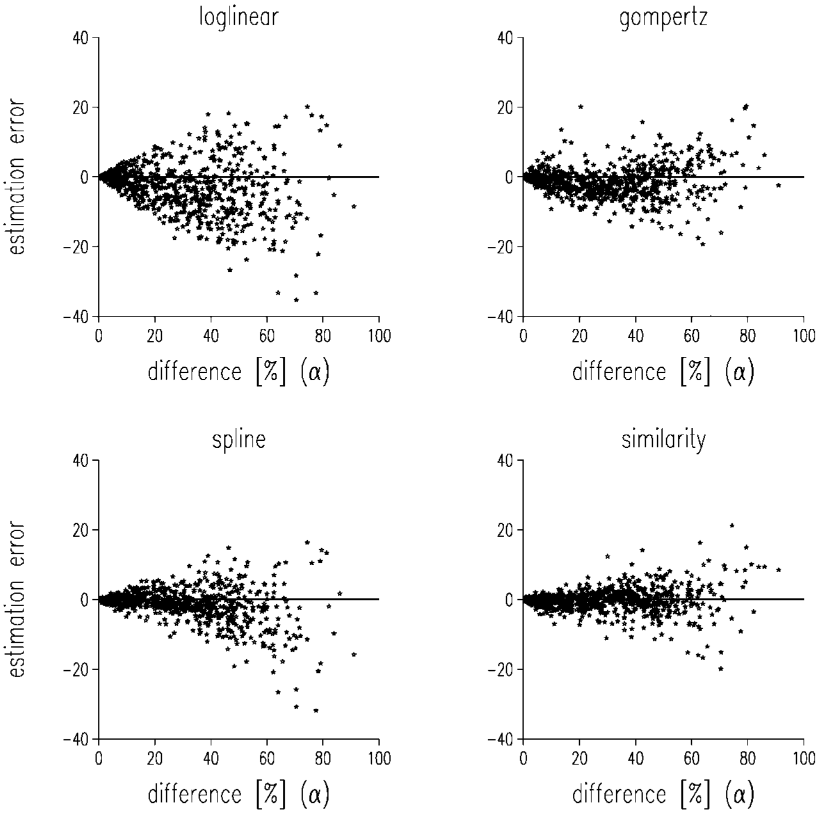


Figure 7 Visual evaluation of the sensitivity of the 4 interpolation procedures to changes in distances between particle-size fractions (α).

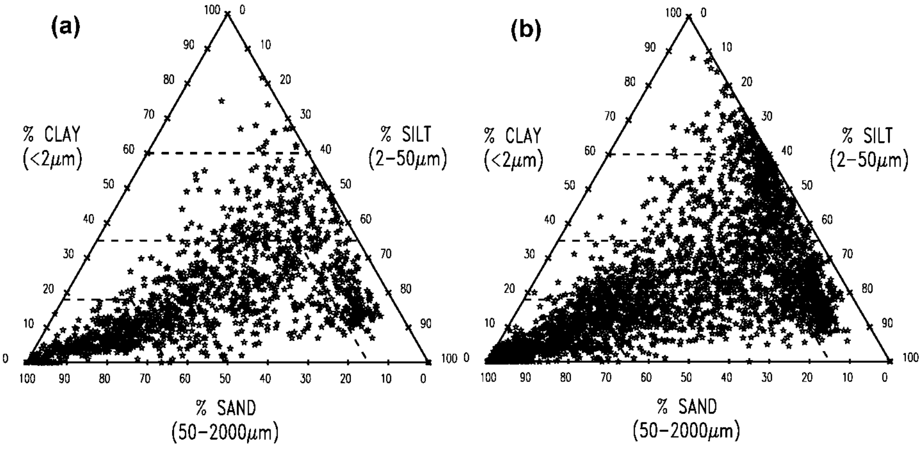


Figure 8 Textural composition of the 1600 topsoil horizons (a) and 3382 subsoil horizons (b) in the HYPRES database.

Discussion

Four different interpolation procedures designed to standardise particle-size distributions of soils in the HYPRES database were tested and evaluated using data from Germany and the Netherlands. Of these four procedures, the novel similarity procedure, which uses a large external reference data set to select 10 similar soils, proved to give the best estimations. It appeared that the splines gave acceptable results only when $\Delta\phi$ is relatively small. Although the Gompertz procedure gave estimations of an accuracy similar to the spline procedure, it proved to be insufficient flexible in case of bimodal particle-size distributions and in case of soils with high sand fractions. In all cases, irrespective of the $\Delta\phi$ value, the loglinear interpolation procedure is outperformed by the other procedures and is therefore, not recommended. A simple criterion for deciding which of the procedures to use, is to look at the distance between the two particle-size limits neighbouring the unknown particle-size limit ($\Delta\phi$ value). When this distance is approximately 2 or higher, than the similarity procedure proves to be the most accurate procedure. When the distance is less than 2, use of splines is sufficient accurate as well. In the latter case, the spline procedure has the advantage that it is simple and needs no external data.

As mentioned, the performance of the similarity procedure depends heavily on the availability of a large, external reference data set. It is demonstrated that such a set of 9607 soil horizons from the Netherlands has a sufficient large variation in cumulative particle-size distributions that it can be used for soils outside the Netherlands. In case of e.g. the Spanish and Portuguese soils the particle-size fractions at the common particle-size limits of the 10 'similar soils' selected from the reference set remained within ±0.5% of the measured fractions. As a consequence, it can be concluded that the reference set, notwithstanding the fact that it originates from the Netherlands, contains soils that have a particle-size distribution very similar to the distribution observed e.g. for Spanish and Portuguese soils.

In case of the German test dataset 2, the spline procedure is used to yield the missing 20μm and 63μm point in the reference set. In the testing procedure the 20μm or the 63 μm point is omitted from test dataset 2 and therefore can serve as measured value. Therefore, in case of using test dataset 2 the estimation error actually is a combination of two errors. The first error results from applying the spline procedure to the reference set and the second one from applying the similarity interpolation procedure to test dataset 2.

Acknowledgements

The generous contribution of data by the collaborating institutions made this study possible. Financial support was provided by the European Commission under the Human Capital and Mobility (DG XII) program as a project entitled '*Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning*' (CHRX-CT94-0639).

Chapter 3

Description of the Unsaturated Soil Hydraulic Database UNSODA Version 2.0

Published in: Journal of Hydrology 251(3-4): 151-162. (2001)

Description of the Unsaturated Soil Hydraulic Database UNSODA Version 2.0

A. Nemes[1], M.G. Schaap[2], F.J. Leij[2] and J.H.M. Wösten[3]

[1] Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Herman Ottó u. 15, H-1022 Budapest, Hungary

[2] George E. Brown Jr. Salinity Laboratory, USDA-ARS, 450 W. Big Springs Road, Riverside, CA 92507, USA

[3] ALTERRA Research Institute for Green World Research, P.O. Box 47, 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.

Keywords: soil hydraulic database, unsaturated soils, water retention, hydraulic conductivity, soil water diffusivity, particle-size distribution

---

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

Introduction

Numerical models are increasingly being used to manage agricultural production, to 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).

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., 2001a). 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 SQL (Structured Query Language) 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 have recently been published that relied on data from UNSODA. Leij et al. (1997) evaluated mathematical expressions to describe water retention and unsaturated conductivity data in UNSODA. Schaap and Leij (1998) used UNSODA as one of 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 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.

The Database

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 'stand-alone' 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 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 Figure 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 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. Figure 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.

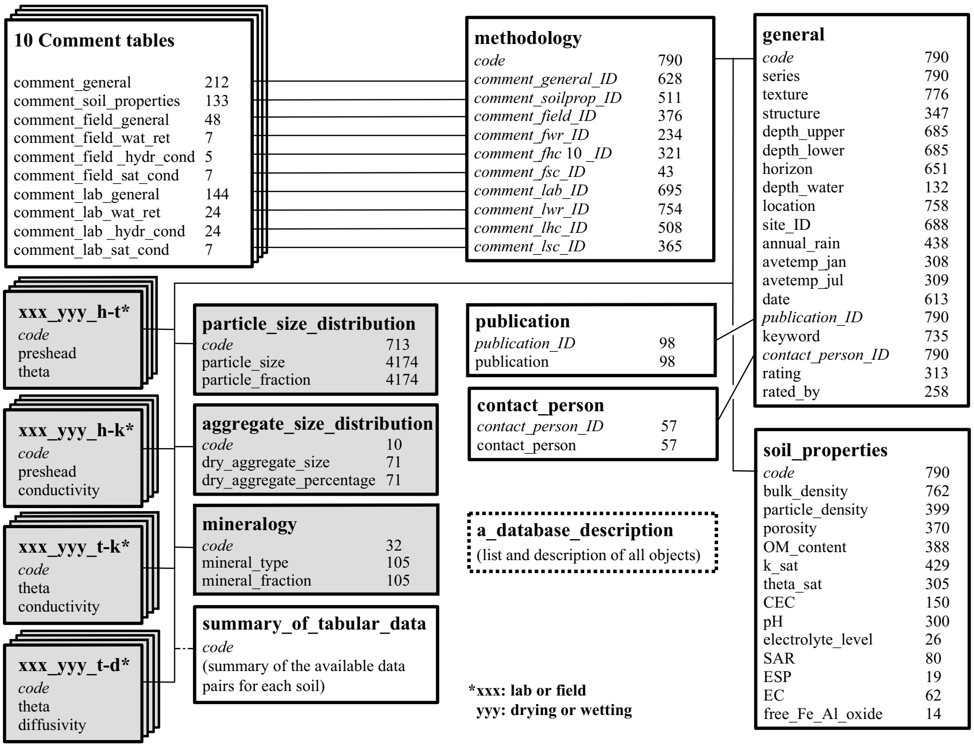


Figure 1 An overview of the database structure and the data in UNSODA V2.0. The boxes indicate tables (table names are in bold), which include the fields and the number of records (samples) with available data. Indexed fields (in italics) are used to link the tables as depicted with the lines between 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 "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 the Operations section.

Data

The present database holds information on 790 soil horizons (“codes”) either contributed by individual scientists or obtained by us from the literature. The number of records for each field in the database can be seen in Figure 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. Figure 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 from either Europe or from North America. The textural classification indicates that coarse-textured soils are in the majority although there is a sizeable amount of soils with finer textures. Sandy clays and silts are poorly represented, however. 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.

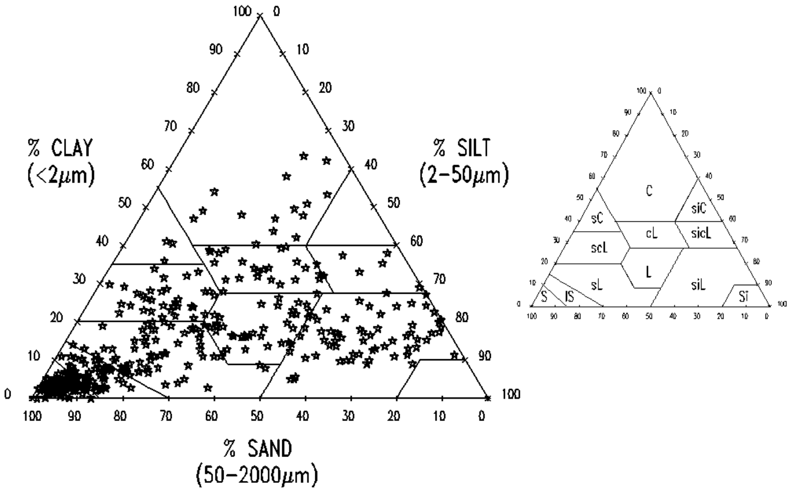


Figure 2 Distribution of 431 soil codes in UNSODA V2.0 across the USDA-SCS soil textural triangle.

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. Figure 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. Figure 1) to enable the user to focus on data determined with a particular method.

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

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

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

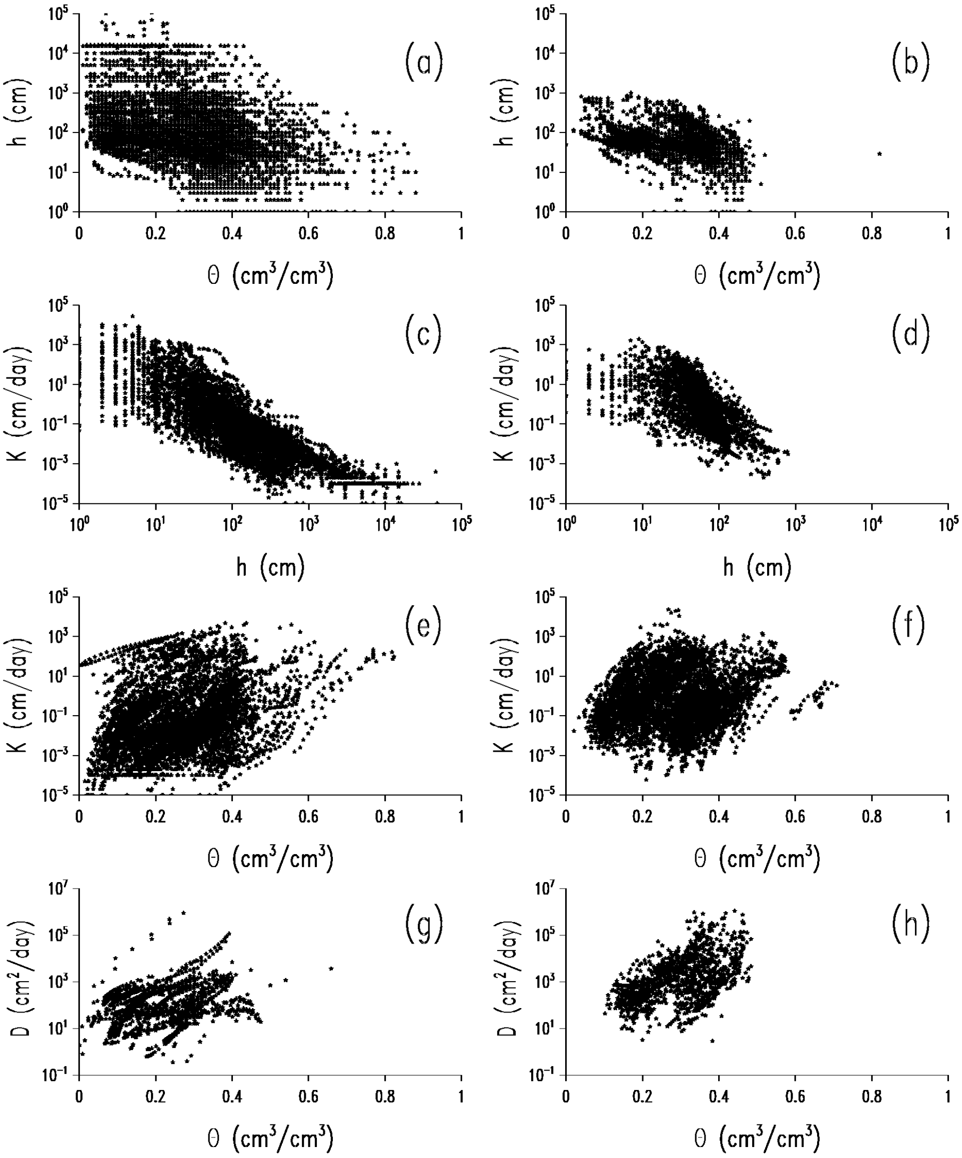


Figure 3 Scatter-plot of the hydraulic data in UNSODA V2.0. From top to bottom: water retention, h(θ), hydraulic conductivity, K(h) and K(θ), and soil water diffusivity D(θ) data. Laboratory measurements are shown on the left-hand side while field measurements are given on the right-hand side.

Database operations

The database allows a number of operations such as searching according to 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.

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 (Standard Query Language) 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 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.

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 the pre-set 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.

Extraction of data

Results from a query and the contents of 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 more complex calculations 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 pre-defined 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.

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 of the use of the UNSODA database; examples of more elaborate applications can be found in the references given in the Introduction.

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 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.

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 sizes 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 at 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 $-{}^{2}\log$[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 200 μms, $\Delta\phi$ 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 re-used to evaluate multiple $\Delta\phi$ distances.

Estimated values for the 50-μm fraction were 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:

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

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.

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*	*1.0*
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*	*19.0*
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*	*0.0*	*6.8*	*2.0*	*1.7*	*0.7*	*0.0*	*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*	*0.7*	*0.0*	*0.7*	*0.0*	*0.0*	*3.3*		**100**

Results of the test are shown in Tables 3 to 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. Some texture groups were poorly or not represented however. 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 for the European. One expects that for the European data set the estimation is not worse than for the non-European data set. The overall RMSE of 10.2 for the European data set is high, however, because of the much higher representation of medium textured soils in that data set, for which the evaluation is much worse.

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 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. Most incorrectly classified soils, however, were classified as a similar group of the textural triangle. Even small estimation errors could lead, however, 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.

Table 5 Original vs. predicted texture groups for the non-European (A) and European (B) subsets. Figures in 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%**

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 one 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.

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 92507-4617, 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 (EAR-9804902) and ARO (39153-EV) while we all thank support for travel by NATO (CLG 975761).

Chapter 4

Unsaturated Soil Hydraulic Database of Hungary: HUNSODA

Based on the paper published in: Agrokémia és Talajtan 51(1-2): 17-26. (2002)

Unsaturated Soil Hydraulic Database of Hungary: HUNSODA

A. Nemes

Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Herman Ottó u. 15, H-1022 Budapest, Hungary

Introduction

Protecting soils and water from being damaged by agricultural and industrial pollution is a major concern. Numerical simulation models are increasingly being used to manage agricultural production, to predict the behaviour of soil contaminants, and to simulate transport processes in the vadose zone. These models typically require a large variety of input data, in particular, soil hydraulic data. The quality of these inputs often substantially affects the quality of the simulation results (Leij and van Genuchten, 1999).

Advances are being made to measure soil hydraulic properties (cf. Gee and Ward, 1999). Despite these efforts, the unfortunate fact remains that the hydraulic parameters are still always notoriously difficult to measure, especially for undisturbed field soils. The methodology to determine soil water retention, θ (h), and, especially, the unsaturated hydraulic conductivity, $K(h)$ or $K(\theta)$, is often perceived as inadequate for many applications, especially when a large number of samples are required (Wösten et al., 2001a).

An alternative to direct measurement of the unsaturated hydraulic properties is the use and/or generalisation of experimental data that are already available. These alternative approaches require reliable sets of soil hydraulic data, as well as basic soil properties such as soil texture, bulk density, and organic material content or other information that may have an influence on the hydraulic behaviour of soils.

The use of models for research and management has shown that many input data have to be quantified in order to be able to make reliable predictions at regional, national or supra-national scales. Unfortunately, these data are often fragmented, of different degree of detail, of varying reliability and are held in different databases or in different institutes (Wösten et al., 1999). The role and importance of proper national and international scale databases in resolving agricultural and environmental issues is often emphasised (e.g. Leij and van Genuchten, 1999). Many electronic databases with soil information exist already, however, the emphasis of most of the databases is on soil taxonomy and they often have limited unsaturated soil hydraulic data. With soil hydraulic data in focus, the international UNsaturated SOil DAtabase (UNSODA) (Leij et al., 1996; Nemes et al., 2001) and the European database of soil hydraulic properties (HYPRES) (Wösten et al., 1999) were developed. Both databases contain a wealth of information on soil hydraulic data, measurement methods and other relevant soil data and provided data to numerous studies.

Similar soil data have also been collected in Hungary. Many studies took place that used smaller or more extensive data sets of soil hydraulic properties. Várallyay et al. (1979) described water-retention curves by fitting different functions and polynomials to measured data. Várallyay et al. (1980) established a new categorisation system and developed a 1:100.000 scale map of soil water management properties for Hungary. Rajkai (1984) calculated capillary conductivity of soils using their water-retention curves, and compared them with measured data. Rajkai et al. (1981) calculated water-retention data from other – easily measurable - soil physical properties, which techniques Rajkai and Kabos (1999) later improved further.

This paper describes a newly developed data source of physical and hydro-physical properties of Hungarian soils. At present, this searchable database holds information on 840 individual soil horizons. The structure and contents of the database are discussed and derived class pedotransfer functions (PTFs) are provided for the benefit of future users.

The database

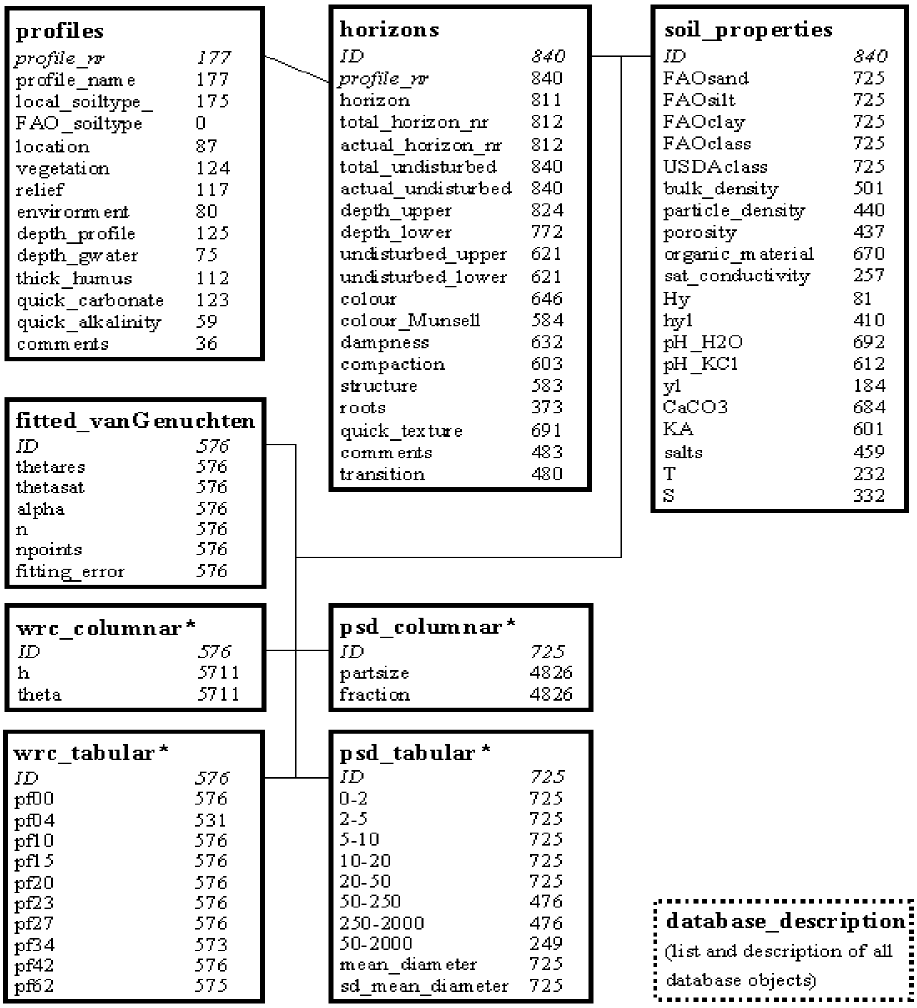
Database structure

HUNSODA is a compilation of the available Hungarian soil hydraulic data in Microsoft Access-97 format. MS Access has been chosen because it is widely available and allows management of data on ‘stand-alone’ computers as well as through computer networks. This format provides flexibility in data entry, manipulation and retrieval as well as data output and interfacing with other applications. Furthermore, MS Access has extensive user-friendly query and graphics capabilities for the examination of 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 8 tables as logical groups of fields. Each of the 840 soil horizons in HUNSODA received a unique identification number that is stored in a separate field named “ID”. The numbering begins from 1 and continues with an increment of 1 in the order of data entry. The tables are linked (“indexed”) through this field – except the “profiles”, table, where this field is not applicable (see later). 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 in the database are usually indicated with a “No data” entry.

The structure of the database, names of tables, links between tables as well as the amount of available data for each field are outlined in Figure 1. The two main tables of HUNSODA are called “profiles” and “horizons”. The “profiles” table holds basic information about the soil profiles, such as their geographic location, classification and some features of their environment. The soil type is originally given according to the Hungarian classification system (field “local_soiltype”). Later, the FAO soil type has been roughly determined from the available information using the FAO World Reference Base (FAO, 1998) and is given in a separate field (“FAO_soiltype”). Some diagnostic features that were observed or determined at the site using quick tests can also be consulted here. The “horizons” table holds the diagnostic description of all 840 horizons in the database. The horizon notations are given according to the Hungarian soil classification system (Stefanovits, 1963; Szabolcs, 1966). The table can be consulted for information on where the horizon is situated within the profile and on the depth from where the undisturbed sample was taken. Features such as colour, Munsell-colour (Munsell Color Company, 1975), dampness, compaction status, soil structure, presence of plant residues (roots) and the transition into the next horizon are also stored in this table. Since every profile (may) have several horizons, it is apparent that the above-described "ID" number can not link these two tables. A 3-digit unique profile identification number (field "profile_nr") has been introduced to enable a proper link between these tables. Obviously, more horizons have the same "profile_nr" entry in the "horizons" table, but duplicate numbers are not allowed in the "profiles" table to keep unanimous correspondence.



profiles	
profile_nr	*177*
profile_name	177
local_soiltype_	175
FAO_soiltype	0
location	87
vegetation	124
relief	117
environment	80
depth_profile	125
depth_gwater	75
thick_humus	112
quick_carbonate	123
quick_alkalinity	59
comments	36

horizons	
ID	*840*
profile_nr	840
horizon	811
total_horizon_nr	812
actual_horizon_nr	812
total_undisturbed	840
actual_undisturbed	840
depth_upper	824
depth_lower	772
undisturbed_upper	621
undisturbed_lower	621
colour	646
colour_Munsell	584
dampness	632
compaction	603
structure	583
roots	373
quick_texture	691
comments	483
transition	480

soil_properties	
ID	*840*
FAOsand	725
FAOsilt	725
FAOclay	725
FAOclass	725
USDAclass	725
bulk_density	501
particle_density	440
porosity	437
organic_material	670
sat_conductivity	257
Hy	81
hy1	410
pH_H2O	692
pH_KC1	612
y1	184
CaCO3	684
KA	601
salts	459
T	232
S	332

fitted_vanGenuchten	
ID	*576*
thetares	576
thetasat	576
alpha	576
n	576
npoints	576
fitting_error	576

wrc_columnar*	
ID	*576*
h	5711
theta	5711

psd_columnar*	
ID	*725*
partsize	4826
fraction	4826

wrc_tabular*	
ID	*576*
pf00	576
pf04	531
pf10	576
pf15	576
pf20	576
pf23	576
pf27	576
pf34	573
pf42	576
pf62	575

psd_tabular*	
ID	*725*
0-2	725
2-5	725
5-10	725
10-20	725
20-50	725
50-250	476
250-2000	476
50-2000	249
mean_diameter	725
sd_mean_diameter	725

database_description
(list and description of all database objects)

Figure 1 The database structure and the available data in HUNSODA. The boxes indicate tables (table names are in bold), which include the fields and the number of samples with available data. Indexed fields (*in italics*) are used to link the tables as depicted with the lines between the tables. Boxes with dotted borders indicate auxiliary tables.

The "soil_properties" table contains laboratory determined physical and chemical data for each horizon. There are 4 tables that contain data with a functional relation between an independent and a dependent variable. Soil water retention (WR) data are stored as h-θ curves. The absolute value of the soil matric or pressure head, h, is given in cm (hPa); and the water content, θ, is expressed in cm^3/cm^3. WR data are stored in two ways: in a column format and in a table format. Table "wrc_columnar" has only three fields: one for the horizon ID number, and two for the actual h-θ data pairs. Since water contents were always measured at standard pressure head values for these Hungarian soils, the same data are stored in a different way also. Table "wrc_tabular" stores data in such way, that for each applied pressure-head value, a separate field is allocated. The particle-size distribution (PSD) tables store the fraction of soil mass as a function of the particle size or diameter for soil particles smaller than 2 mm. Tables "psd_columnar" and "psd_tabular" store PSD data in the same manner as described for WR data above. Experience shows that storing data in both ways can help the user in executing quick selections using various formats.

Table "fitted_vanGenuchten" contains information derived from measured data. The four parameter van Genuchten model (θ_r, θ_s, α, n) (van Genuchten, 1980) – as one of the most commonly used parameterised input for numerical simulation models - was fitted to the water-retention data, using a genetic algorithm that minimises the fitting error. The four fitted parameters are stored for each horizon where water-retention data were available. The number of considered water-retention data pairs and the error of the fit averaged for the curve were also logged here.

Information on measurement methodology can be found in the 'Description' of each field, where applicable. Those can be consulted once the appropriate table is opened in design view.

Two auxiliary tables are included in HUNSODA. A table titled "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 to help the user in familiarising with the database.

Figure 2 Map of the sampling location of soils in HUNSODA

The data

HUNSODA at present holds information on 840 soil horizons collected by and analysed at the Research Institute for Soil Science of the Hungarian Academy of Sciences (RISSAC) and contributed either by individual scientists or as data collected in the framework of joint research projects. The number of records available for each field in the database can be seen in Figure 1. The "wrc_columnar" table should be interpreted as follows: the database contains 576 soil samples with water-retention data having a total of 5711 h-θ data pairs. Figures for the "wrc_tabular" table show how many samples of the database have measured values at the particular pressure-head value. The particle-size tables should be interpreted likewise. Figure 2 shows the geographical location of the profiles. Data of these profiles were not specifically collected in the framework of a single major project, so the location of profiles does not reflect any systematic sampling. However, most of the lowland areas that are of major agricultural or environmental interest are represented in the database. Figure 3 shows the USDA-SCS texture triangle and the distribution of the 725 samples with original measured data. Most of the textural groups are well represented. Soils with very fine texture and those that are a mixture of very fine and coarse particles (sandy Clays) are however somewhat underrepresented – which may correspond with the probability of their natural occurrence. Table 1 shows the number of topsoil and subsoil horizons with available water-retention data in each of the texture classes according to the FAO and USDA classification systems. Any soil horizons which were classified as 'A' horizons were considered as topsoils, all others were grouped together as subsoils. The poor representation of the 'very fine' group (FAO) is apparent. There are considerably more SiL, SiC and SicL soils with available water-retention data than soils in other USDA textural classes. Unlike some other databases, HUNSODA does not show bias in favour of coarse-textured soils, which is partly due to the natural occurrence of soil textural types and partly may reflect experimental interests (e.g. salinity research) towards soils with heavier texture. In principle, topsoils represent one third of the database.

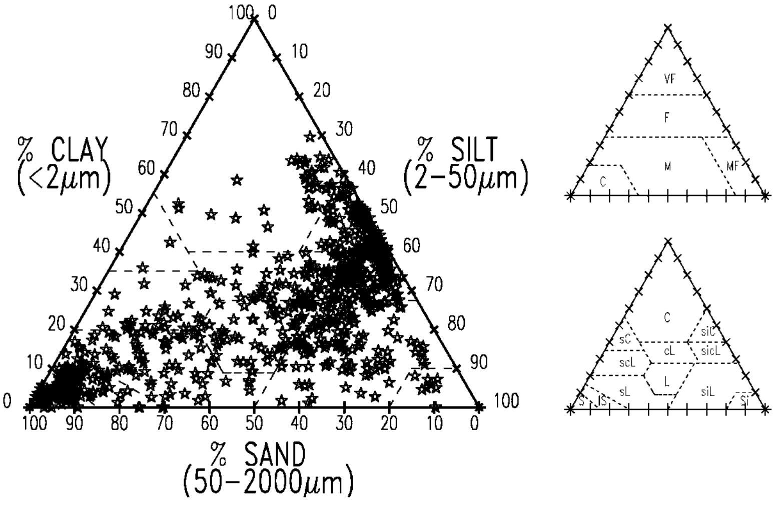


Figure 3 Distribution of soils in HUNSODA across the USDA-SCS soil textural triangle

Table 1 Textural distribution of soils with measured water-retention and particle-size distribution data in the HUNSODA database: (a) according to the FAO classification system and (b) according to the USDA classification system. (*n*: number of samples; *N*: total number of samples)

(a)

FAO class	Subsoil	Topsoil	n	*% of N*
C	49	19	68	*12.8*
M	128	78	206	*38.8*
MF	57	25	82	*15.4*
F	120	43	163	*30.7*
VF	7	5	12	*2.3*

(b)

USDA class	Subsoil	Topsoil	n	*% of N*
S	27	9	36	*6.8*
lS	7	6	13	*2.4*
sL	30	11	41	*7.7*
scL	7	2	9	*1.7*
C	18	10	28	*5.3*
cL	23	11	34	*6.4*
L	24	21	45	*8.5*
SiL	65	43	108	*20.3*
Si	6	3	9	*1.7*
sicL	87	34	121	*22.8*
siC	67	20	87	*16.4*
n	361	170	531	
% of N	*68*	*32*		

There are 576 available water-retention curves measured in the laboratory. Water contents were determined at 10 standard pressure head values which are 0, -2.5, -10, -30, -100, -200, -500, -2500, -15000 and $-1.5*10^6$ hPa (that are equivalent to pF0, pF0.4, pF1.0, pF1.5, pF2, pF2.3, pF2.7, pF3.4, pF4.2 and pF6.2). Some of the hydraulic data are geometric averages for replicate curves for the same horizon/sample. The available data covers a 40-50% (v/v) wide band from 0 to -15000 hPa. The number of points on each water-retention curve and the wide range in θ and h makes the database a suitable data source for the optimisation of mathematical expressions or the development of pedotransfer functions.

Database operations

The database allows a number of operations such as searching according to 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.

Specific information can be obtained from the database by running a 'query' to retrieve data from one or more tables according to user-defined criteria. Queries can be specified in SQL (Standard Query Language) or they can be entered through a graphical interface, which does not require detailed knowledge of SQL. Through the graphical interface 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.

HUNSODA is not write-protected, thus the user is able to modify the contents of the database. Editing data is simply a matter of overwriting the proper fields. When the same changes are desired for a lot (or all) of the entries, considerable effort can be avoided by defining an 'update query' that will search and overwrite the contents of all entries that meet the defined criteria.

New data can be entered in different ways. One can simply open the targeted table in datasheet view and start typing the new data at the end of the appropriate field(s). 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 from one or more tables simultaneously. Forms can display additional information on for example, the required data format or other constraints.

Results from a query and the contents of 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 more complex calculations 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 (printable) output based on data selected by queries. One such report is provided that summarises all available information for each horizon.

Applications

A soil physical and hydro-physical database may be a data source for a variety of studies related to agriculture and environment. In the following we demonstrate a possible way of deriving extra information from the database contents, through the development of class PTFs.

The term "pedotransfer function" (PTF) has been introduced by Bouma and van Lanen (1987) to describe empirical functions that relate different soil characteristics and/or properties to one another or to land qualities. Class PTFs use horizon designations, soil type or soil textural class as regressed variables. Grouping soils according to their texture often proved useful, as different particle-size distributions can be associated with different pore-size distributions, which in turn mean different soil hydraulic properties. Several databases or data sets have been used in the past to derive such functions (e.g. Rajkai and Kabos 1999, Wösten et al., 1999, Wösten et al., 2001b).

We simultaneously fitted a function to hydraulic data of soils grouped by soil texture. Textural classification of soils is however not applied uniformly throughout the World. Different particle-size limits are used to distinguish clay/silt/sand fractions (Filep and Ferencz, 1999; Nemes et al., 1999), but even with the same particle-size limits, class boundaries may be considered at different particle fractions, e.g. the USDA-SCS (USDA, 1951) and the FAO (FAO, 1995) classification systems. Moreover, in some cases soil textural classification is not solely based on soil particle-size distribution data. In Hungary other indicators, like soil hygroscopy and the sticky point index of Arany (Arany, 1943) are used along with the fraction of particles smaller than 0.02μm to define soil texture. These properties are determined according to Várallyay (1978) and Búzás (1993). To achieve internationally compatible and comparable results, we derived class PTFs according to the two classification systems that are most commonly used in current international research.

The FAO and USDA texture classification systems are based on three-poled particle-size distributions, which define clay as the fraction of particles <2μm, silt as the fraction of particles between 2 and 50 μm and sand as the fraction between 50 and 2000μm. Along with the distribution of the HUNSODA soils in the USDA texture triangle, Figure 3 shows the boundaries of the different texture classes according to the FAO and USDA-SCS classification systems.

Table 2 Re-fitted Mualem – van Genuchten parameters of each class pedotransfer function: (a) according to the FAO classification system and (b) according to the USDA classification system.

(a)

FAO class	θsat	θres	n	alpha	RMSR
C	0.00966	0.414814	0.027478	1.534133	0.010877
M	0.00000	0.438973	0.009746	1.228564	0.023715
MF	0.00000	0.447729	0.002281	1.251066	0.030531
F	0.00000	0.450373	0.000823	1.254555	0.032474
VF	0.00000	0.525737	0.000883	1.226032	0.027481

(b)

USDA class	θsat	θres	n	alpha	RMSR
S	0.01300	0.408743	0.023771	1.875734	0.012861
lS	0.00000	0.413930	0.022367	1.412027	0.011259
sL	0.00000	0.424590	0.016445	1.251622	0.016461
scL	0.00000	0.430524	0.029298	1.192810	0.025381
C	0.00000	0.498629	0.000670	1.252291	0.028852
cL	0.00000	0.430199	0.002402	1.246581	0.030105
L	0.00000	0.423860	0.006519	1.245827	0.022730
SiL	0.00000	0.458333	0.009931	1.230832	0.024158
Si	0.00000	0.463677	0.003128	1.282823	0.027745
sicL	0.00000	0.435508	0.001765	1.239395	0.032917
siC	0.00000	0.453244	0.000854	1.246492	0.031838

Water-retention data of the database were measured at ten standard pressure-head values in the laboratory (see earlier). Class PTFs were developed by first calculating the geometric mean water contents at each of the ten pressure-head values separately for each texture class. Next the van Genuchten model was fitted to the mean values using a genetic algorithm, which minimises the fitting error. Table 2 lists the fitted van Genuchten parameters for each texture class, along with the root mean squared residuals of the fits in comparison with the geometric mean values at the 10 pressure-heads. Fits are more accurate for classes with coarser texture and worse for classes with finer texture. Figure 4 shows, for the FAO classes, the geometric mean values connected to form a curve and the standard deviation (SD) of the mean values at the ten pressure-heads also connected to form a band. Standard deviations are generally larger towards the wet end of the curves. Figure 5 shows the fitted curves using the van Genuchten parameters in Table 2. From this figure, differences between FAO texture groups become more apparent, with the coarse textured soils clearly showing the lowest average water contents throughout the entire water-retention curve and the very fine textured class showing the highest water contents at each pressure-head. Some differences between the Medium, Medium-Fine and Fine groups can also be seen, however, considering their SD-s from Figure 4, there are considerable overlaps among these groups. Considering the USDA classification, – apart from the coarsest and finest textured classes – class average curves run relatively close to each other – with some of them crossing over. It may suggest that there is no need for a lot of small classes while deriving class PTFs, but rather should a very distinctive but not so detailed system be used.

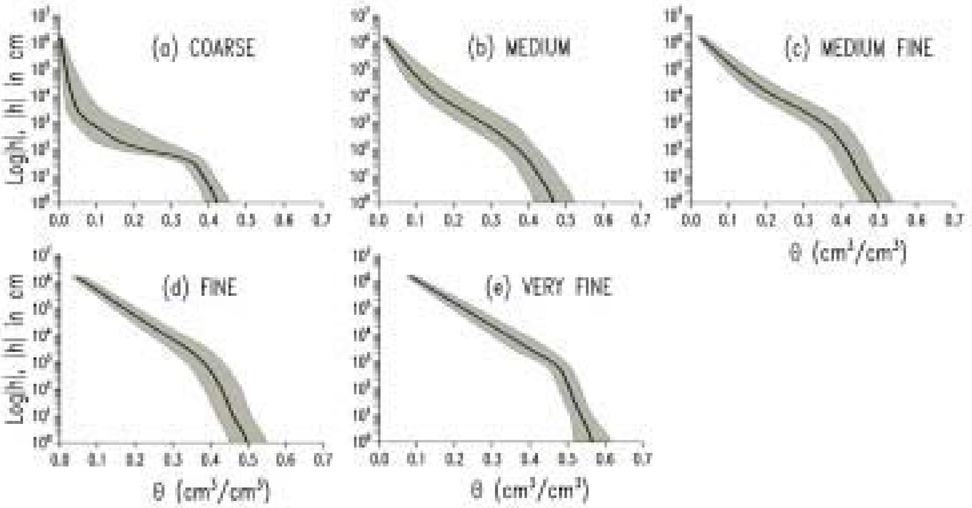


Figure 4 Geometric mean water-retention curve ±1SD for the FAO soil texture classes. Data are based on measurements at the ten pressure-heads and connected (interpolated) in between.

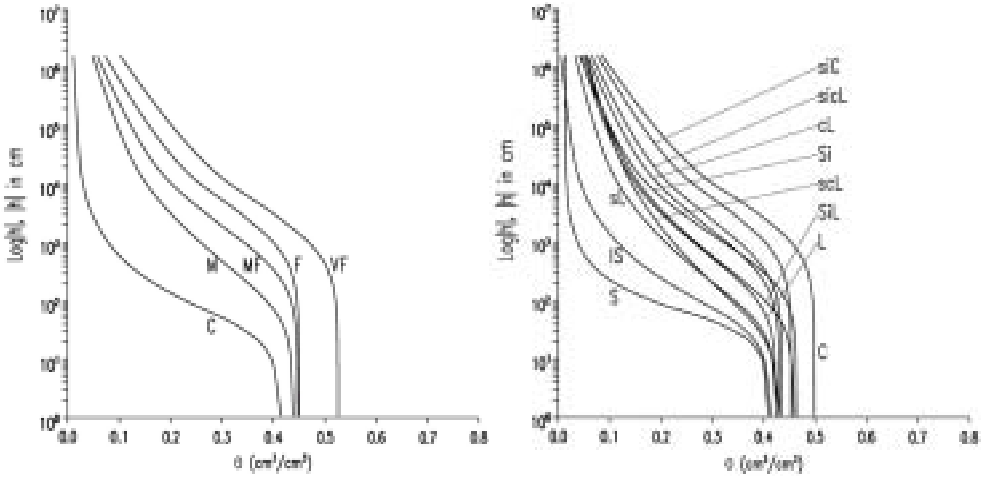


Figure 5 Class pedotransfer functions according to the FAO (left) and USDA (right) soil texture classes

Summary and recommendations

Databases with information on soils often constitute a basic source of information for studies in production agriculture, environmental engineering, and remote sensing. A new, searchable, relational database of soil physical and hydro-physical properties has been developed for Hungarian soils. The database is stored in MS-Access format. The database is compatible with most of the popular software and can be run on a personal computer as well as on computer networks. It currently stores data of 840 soil horizons, 576 having soil hydraulic data, measured using nationally and internationally accepted standard methodology. An outline of the structure and the contents of HUNSODA, as well as a basic guide to its operation has been given; and class pedotransfer functions according to alternative texture class definitions were developed. When more classes are distinguished, PTFs of several classes tend to be very similar to each other.

Developing this database may enhance the ability of Hungarian soil scientists to address many more environmental issues of concern, however, further expansion of the database is recommended. Additional data can easily be included in the database. Expansion of the database may help in making it more widely applicable.

Future research is needed to include other factors, like soil structure and soil compaction status as possible variables to distinguish groups within soil texture classes.

Availability of the database

HUNSODA currently requires 1.5 Mbytes of disk space (~0.4 Mbytes when condensed) to store the database. The database can be requested by regular mail (Director’s Office, MTA TAKI, Herman Ottó u. 15, H-1022 Budapest, Hungary). 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 author wishes to thank the Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences and all individuals who contributed data to the database. The author wishes to specially thank Dr. Kálmán Rajkai and Gyöngyvér Motsai for their extensive help in accessing old, archived data.

Chapter 5

Functional Evaluation of Pedotransfer Functions Derived from Different Scales of Data Collection

Accepted for publication by Soil Science Society of America Journal

Functional Evaluation of Pedotransfer Functions Derived from Different Scales of Data Collection

A. Nemes[1], M.G. Schaap[2] and J.H.M. Wösten[3]

[1] Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Herman Ottó u. 15, H-1022 Budapest, Hungary

[2] George E. Brown Jr. Salinity Laboratory, USDA-ARS, 450 W. Big Springs Road, Riverside, CA 92507, USA

[3] ALTERRA Research Institute for Green World Research, P.O. Box 47, 6700 AA Wageningen, The Netherlands

Abstract

Soil water and solute transport models require soil hydraulic properties as input. Estimation of these properties by pedotransfer functions (PTFs) can be an alternative to troublesome and expensive measurements. New approaches to develop PTFs are continuously being introduced, however, PTF applicability in locations other than those of data collection has been rarely reported. In this study, three databases were used to develop PTFs using artificial neural networks (NNs). Data from Hungary were used to derive *national scale* soil hydraulic PTFs. The HYPRES database was used to develop *continental scale* PTFs. Finally, a database containing mostly American and European data was used to develop *intercontinental scale* PTFs. For each database, 11 PTFs were developed that differed in detail of input data. Accuracy of the estimations was tested in two ways, using independent Hungarian data. First, soil water retention was estimated at nine values of soil matric potential. Root mean squared residuals (RMSRs) using different inputs ranged from 0.02 to 0.06 m^3m^{-3} for national scale PTFs, while international scale PTFs had RMSRs from 0.025 to 0.088 m^3m^{-3}. Larger errors were found mostly for soils that were underrepresented in the data of PTF development. Estimated water retention curves (WRCs) were then used to simulate soil moisture time series of seven Hungarian soils. Small differences were found among the PTFs derived from different scale data collections. Root mean squared residuals during a growing season ranged from 0.065 to 0.07 m^3m^{-3}, while simulations using laboratory-measured WRCs had RMSR of 0.061 m^3m^{-3}. Such small differences in the accuracy of simulations using PTFs and measured WRCs make international PTFs an alternative to national PTFs and measurements. However, testing of the international PTFs with a specific model for specific soil and land use remains desirable because of uncertainty in soil representation in such databases.

Introduction

Modeling water and solute transport has become an important tool in simulating agricultural productivity as well as environmental quality. The use of models, however, is often limited by the lack of accurate information on soil hydraulic properties. Soil water and solute transport models typically require data on soil water retention and hydraulic conductivity. Measurement of these properties is relatively time-consuming and costly, especially when data are needed for large areas of land. For many applications, the estimation of hydraulic characteristics with pedotransfer functions (PTFs) can be an alternative as most of the PTFs use input data that are easily and/or routinely collected.

A prerequisite for the development of PTFs is the availability of a source database that contains potential predictors as well as hydraulic properties. Most PTFs available in the literature use soil texture, bulk density (BD), organic matter contents (OM) as predictors; additional parameters are rarely used (Rawls et al., 1991; Wösten et al. 2001a). Estimation of hydraulic characteristics is mostly limited to water retention points or parameters and saturated hydraulic conductivity. A small number of PTFs were proposed for the estimation of unsaturated hydraulic conductivity. Pedotransfer functions are usually published in a tabular form for particular soil classes, as linear or nonlinear regression equations, or more recently, distributed as computer codes resulting from neural network analysis (e.g. Schaap et al., 2001). For an overview of the current status of PTFs we refer to Wösten et al. (2001a).

Many PTFs have been developed in recent decades. Independent databases have been used to evaluate various PTFs that were developed elsewhere. For example, Tietje and Tapkenhinrichs (1993) and Kern (1995) evaluated different PTFs for the estimation of water retention. Tietje and Hennings (1996) tested PTFs for the estimation of saturated conductivity. Some recent comparisons include Imam et al. (1999), Cornelis et al. (2001), and Wagner et al. (2001). Imam et al. (1999) compared three PTFs to compute the water holding capacity of inorganic soils. Cornelis el al. (2001) compared nine PTFs to estimate the soil moisture retention curve. Wagner et al. (2001) evaluated the performance of eight PTFs to estimate unsaturated soil hydraulic conductivity. The latter two studies ranked PTFs, noting that the PTF performance in both cases could be influenced by the geographical preference of the source data sets.

A limitation of most studies that evaluate PTFs is that it remains unclear what the main sources of the estimation errors are. In those studies it is not clear whether differences between data sets used to derive PTFs (size, origin, reliability), differences between the algorithms of PTF development (e.g., different regression types vs. NN models) or differences among the predictors cause a particular PTF to perform better than others. Schaap and Leij (1998) cross-validated NN models by developing PTFs using the same algorithm and the same predictors on data of three independent databases. They found that PTFs derived from one database gave systematically different estimations for the other two data sets, but that estimations improved somewhat when PTFs were derived from all available data. They concluded that the performance of a PTF depends on both the derivation and evaluation data sets and that origin, size, and other data characteristics may determine the performance of PTFs.

Water retention and hydraulic conductivity are not the final aim but are intermediate characteristics needed to calculate other soil properties with more practical meaning. Functional evaluation of estimated soil hydraulic data helps to characterize the contribution of such data to the inaccuracy and uncertainty of simulations. A number of studies used soil water simulation models to evaluate the performance of estimated soil hydraulic characteristics through the simulation of different aspects of soil behavior (e.g., Wösten et al., 1995; Espino et al., 1996; Hack-ten Broeke and Hegmans, 1996; van Alphen et al., 2001; Soet and Stricker, 2002).

Many countries or regions in the World do not have a sufficient amount of soil hydraulic data for agricultural or environmental modeling purposes or to develop PTFs. Nations or regions with appropriate resources, however, have collected considerable amounts of soil information. Most often, these data are used in individual studies at local or national scale. However, in recent years, development of international databases (e.g., Wösten et al., 1999; Nemes et al., 2001) enabled, among other possibilities, the development of international PTFs. A potential benefit of having extensive international databases is that those may include, but are not limited to, soils that are similar in their properties and were developed under similar soil-forming conditions to the soils of the area of planned application.

The objective of this study was to test the hypothesis that a PTF developed from an international database cannot be used at smaller, national scale, and country- or region-specific PTF is necessary. We developed PTFs for the estimation of water retention using data stored in two international databases and developed similar country-specific PTFs using data from Hungary. All PTFs were tested using a second, independent Hungarian data set. The performance of international and national PTFs were first tested on sum of square residuals between measured and estimated retention characteristics. Next, we used different PTF estimates to simulate soil moisture time series of seven Hungarian soils. Sum of square residuals were evaluated between simulated water contents and water contents observed in the field.

Materials and Methods

Data sets

Three databases were used to provide soil hydraulic data collected at three different scales. The HUNSODA database (Nemes, 2002) comprises soil data collected solely in Hungary. The database holds soil water retention characteristics for 576 soil horizons. The HYPRES database (Wösten et al., 1999) contains data of 12 countries of Europe, and holds measured soil hydraulic characteristics for 4030 soil horizons. A third intercontinental database provided data from several parts of the World. This database was previously used for the development of the ROSETTA pedotransfer program (Schaap et al., 2001) and contains 2134 soil samples derived from the UNSODA database (Nemes et al., 2001) and two other databases that originate from the United States (cf. Schaap and Leij, 1998).

All three databases were filtered to select soils that are (i) mineral soils; (ii) have data available on soil texture, BD and OM content; and (iii) have at least four measured soil water retention points [$\theta(h)$]. This selection left us with Hungarian (N = 471), European (EUR, N = 2464) and Intercontinental (ICO, N = 1347) data sets. The Hungarian data set was further split randomly to generate a data set to develop PTFs (HUN, N = 235) and a data set to test PTFs (HUNTEST, N = 236). The HUN, EUR, and ICO data sets were used to formulate two additional data sets to develop PTFs. Because Hungarian data were not present in either of the international scale databases, the assumption is that any information developed from these databases would not be valid for Hungary. To investigate the effect of data from a country being represented in an international database we also combined a copy of the HUN data set with a copy of each of the two International data sets to form two new data sets later referred to as EUR + HUN (*N* = 2699) and ICO + HUN (*N* = 1582).

Figure 1 shows the number of samples and the textural composition in the HUN, EUR, and ICO data sets that are used to develop PTFs and in HUNTEST, the data set used for testing PTF performance. Texture classes are not represented equally in all data sets. The two international data sets hold a considerably larger proportion of loamy sand, sandy loam, sandy clay loam samples, whereas soils with silty clay and silty clay loam texture are better represented in the Hungarian data. Sandy clays and silts are, however, poorly represented in all three data sets.

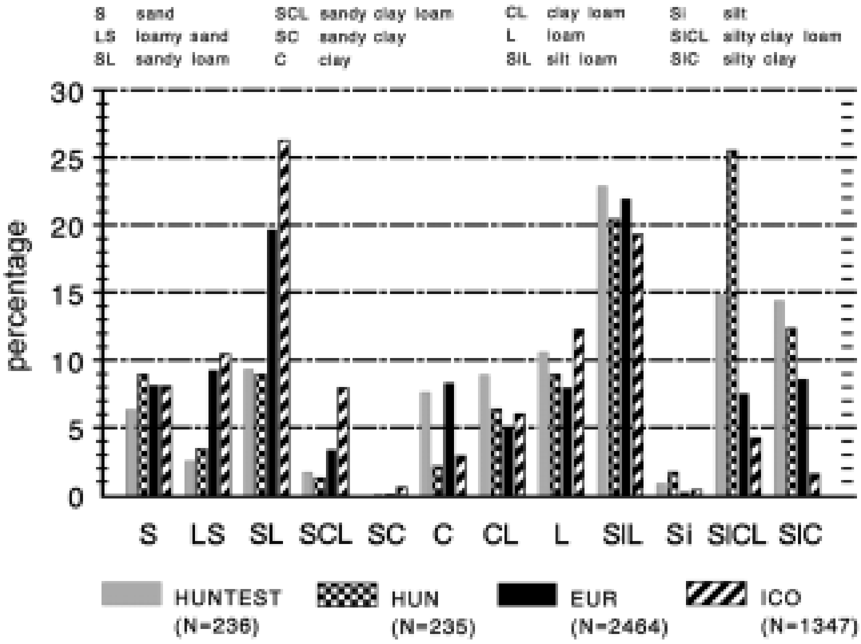


Figure 1. Textural composition of the test data set (HUNTEST) and the Hungarian (HUN), European (EUR), and Intercontinental (ICO) data sets of pedotransfer function development according to the USDA soil textural classification system (USDA, 1951). The number of samples (*N*) in each data set are in parentheses.

Table 1 shows the summary statistics of some variables of the above data sets that were used in the PTF development and evaluation. Differences in texture and OM content are apparent, with decreasing average silt, clay, and OM contents from the Hungarian data through the European to the intercontinental scale data. Data in the different data sets were obtained from different sources and had different numbers and positions of points at the WRCs. To obtain uniform description of all the WRCs, the volumetric soil water content, θ, as a function of matric potential, h, was described with the van Genuchten equation (van Genuchten, 1980):

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1+(\alpha h)^n]^m} \qquad [1]$$

where subscripts r and s refer to residual and saturated values, and α, n and m are curve shape parameters, where $m = 1 - 1/n$. Parameters of the above equation were fitted to the individual WRCs using the Simplex method (Nelder and Mead, 1965). Average water contents at four different matric potentials (0 kPa, -10 kPa, -33 kPa, -1500 kPa) show a decreasing trend from the Hungarian to the intercontinental data, as can also be expected from the differences in texture and OM content. Besides being geographically more diverse than the EUR data set, the ICO data set is also the one being more distant from the HUN data set in its properties, based on Table 1.

Table 1. Mean, standard deviation, and median of some soil properties of the test data set (HUNTEST) and the Hungarian (HUN), European (EUR), and Intercontinental (ICO) data sets used to develop pedotransfer functions.

		HUNTEST			HUN			EUR			ICO		
Property	Unit	mean	SD	med.	mean	SD	med.	mean	SD	med.	mean	SD	med.
USDA Sand	%	29.00	26.38	21.15	30.79	28.76	20.80	40.12	31.09	31.66	51.78	26.49	51.90
USDA Silt	%	43.72	17.99	47.85	44.40	19.93	51.40	37.75	21.93	35.85	31.59	21.82	30.50
USDA Clay	%	27.28	16.32	26.11	24.81	14.68	25.80	22.14	17.17	18.00	16.63	12.36	14.20
Org. Matter	g kg^{-1}	17.72	16.66	16.45	18.43	16.67	17.20	15.93	19.31	9.50	13.82	22.07	3.45
Bulk Density	Mg m^{-3}	1.43	0.16	1.47	1.44	0.16	1.46	1.45	0.22	1.49	1.40	0.25	1.44
θ_{sat}†	m^3 m^{-3}	0.47	0.06	0.46	0.46	0.06	0.45	0.43	0.08	0.42	0.41	0.09	0.41
θ(-10 kPa)†	m^3 m^{-3}	0.40	0.09	0.40	0.38	0.08	0.39	0.33	0.12	0.34	0.31	0.12	0.32
θ(-33 kPa)†	m^3 m^{-3}	0.35	0.10	0.36	0.33	0.10	0.34	0.29	0.12	0.29	0.25	0.11	0.26
θ(-1500 kPa)†	m^3 m^{-3}	0.20	0.09	0.20	0.18	0.08	0.19	0.17	0.10	0.16	0.11	0.06	0.10

† derived from fitted van Genuchten parameters

Neural networks

Pedotransfer functions have been developed using a wide variety of techniques (cf. Rawls et al., 1991; Wösten et al., 2001a). One recently used technique is the analysis by artificial NNs (e.g. Pachepsky et al., 1996; Tamari et al., 1996; Minasny et al., 1999). As most studies found that the predictive capabilities of NN PTFs were equivalent or superior to different regression-type PTFs, we used NNs in this study.

A NN model consists of many simple computing elements (neurons or nodes) that are organized into subgroups (layers) and are interconnected as a network by weights. A model typically consists of an input layer, an output layer, and one (or more) *hidden* layer(s) that connect(s) the input and output layers. The number of nodes in the input and output layers correspond to the number of input and output variables of the model, the number of hidden nodes can be varied freely. Data flow goes from the input layer through the hidden layer(s) to the output layer. A node in the hidden and output layers receives multiple inputs, typically from all nodes of the previous layer. Within the node, each input is weighted and combined to produce a single value as the output of that node, which is then directed to all the nodes of the next layer, or outputted if it was a node of the output layer. The weight matrices are obtained through a calibration (training) procedure, which can then be used to make estimations on independent data. For a more thorough description on NNs, we refer the reader to Hecht-Nielsen (1990) or Haykin (1999).

Following Schaap and Leij (1998), we used a three-layer back-propagation NN model. The number of nodes in the hidden layer was set to six. Eleven different models were developed to estimate water retention through the parameters of the van Genuchten equation, separately from each of the five data sets outlined above. This is to avoid a possible bias while applying one particular set of input parameters. Models were built up gradually, from simple to more complex, using the most commonly used predictors. These predictors were sand, silt, and clay contents, BD, OM contents as well as water retention points [$\theta(h)$] at three different matric potentials (h= -10, -33, -1500 kPa). Details on the input parameters used in each model can be seen in Table 2.

Table 2. Input parameters of the various neural network models. (SSC: sand, silt, and clay content, [%]; BD: Bulk Density, [Mg m^3]; OM: Organic Matter content, [g kg^{-1};]; θ(x): soil water content [m^3 m^{-3}] at matric potential x [-kPa])

Model	Input variables
M1	SSC
M2	SSC + BD
M3	SSC + BD + OM
M4	SSC + BD +θ(1500)
M5	SSC + BD +θ(33)
M6	SSC + BD +θ(10)
M7	SSC + BD +θ(10),θ(33)
M8	SSC + BD +θ(10),θ(1500)
M9	SSC + BD +θ(33),θ(1500)
M10	SSC + BD +θ(10),θ(33),θ(1500)
M11	SSC + BD + OM +θ(10),θ(33),θ(1500)

NNs were combined with the data selection procedure of the bootstrap method (Efron and Tibshirani, 1993) to generate internal calibration-validation data set pairs for an early stopping procedure. The bootstrap method is a nonparametric technique that simulates alternative (replica) data sets out of a single data set. Given a data set of size N, the bootstrap method generates replica data sets, also of size N, by random selection with replacement. Some samples are included more than once, while others are not selected into a particular replica data set. The replica data set is used to calibrate the NN model while data not in the replica data set are used for validation to stop the calibration process when a minimum error is reached. Multiple realizations of subsets can help to avoid bias toward any particular calibration-validation data set pairs. We generated 50 replica data sets, each of which was used to calibrate the NN models. This procedure provided 50 subestimates that could be slightly different from each other. The final estimate of a PTF, for each value, was then calculated by averaging the 50 subestimates of the value. All NN modeling was performed with the Neural Network Toolbox in MATLAB (Demuth and Beale, 1992).

Functional evaluation using simulated water content time series

Data on water regime in seven Hungarian soils were used to evaluate PTFs in their ability to provide parameters for water transport modeling. Soil water contents were measured in a Dystric Haplustept (GDL; Farkas et al., 1999), an Aquic Kandiustalf (DHR2), and two Aquic Calciustepts (DHR3 and DHR4; Czinege, 2000), an Udertic

Haplustoll (KM1), a Pachic Udertic Haplustoll (KM1K; Tóth and Várallyay, 2001), and a Leptic Natrustoll (NYL249; Tóth and Kuti, 2002). These profiles represent arable land as well as pasture. Five different crops covered the seven fields. The profiles had three to five distinct genetic horizons. Data on soil properties and land use are collected in Table 3.

Table 3. Properties of soils used in the simulations.

Soil	Soil Taxonomy *[World Reference Base]* *[Hungarian classification]*	Crop	USDA Sand		USDA Silt		USDA Clay		Organic Matter		Bulk Density	
			min.	max.	min.	max.	min.	max.	min.	max.	min.	max.
			%						$g\ kg^{-1}$		$Mg\ m^{-3}$	
GDL	Dystric Haplustept *[Chromic Cambisol]* *[Brown forest soil]*	corn	63.2	65.4	19.0	20.5	14.1	16.7	15.0	28.1	1.56	1.66
DHR2	Aquic Kandiustalf *[Gleyic Calcisol]* *[Alluvial meadow soil]*	alfalfa	21.5	56.9	24.6	44.9	18.4	33.6	6.6	25.3	1.58	1.68
DHR3	Aquic Calciustept *[Aridic Calcisol]* *[Calc. multilayer humous sand]*	alfalfa	30.1	95.9	1.8	35.5	1.5	34.5	1.0	19.1	1.48	1.74
DHR4	Aquic Calciustept *[Haplic Arenosol]* *[Calc. multilayer humous sand]*	alfalfa	50.6	96.1	1.8	26.8	2.1	22.7	1.0	13.2	1.49	1.58
KM1	Udertic Haplustoll *[Vertic Chernozem]* *[Meadow Chernozem]*	sunflower	2.9	4.5	53.9	57.2	39.3	41.7	9.4	53.2	1.39	1.53
KM1K	Pachic Udertic Haplustoll *[Calcaric Phaeozem]* *[Lowland Chernozem]*	winter wheat	3.3	5.1	55.2	60.7	35.9	40.3	0.7	64.8	1.40	1.52
NYL249	Leptic Natrustoll *[Verti-Mollic Solonetz]* *[Meadow Solonetz]*	pasture	3.6	7.3	54.4	67.1	27.2	42.1	2.7	15.6	1.46	1.66

Soil water contents were measured at five to 10 depths per profile (to a maximum depth of 80-100 cm), eight to 10 times a year, using auger. Simulations of one-dimensional flow were carried out using the SWAP model (version. 2.07d) (van Dam et al., 1997). The lower boundary conditions were set as: free drainage for GDL, and measured groundwater levels given as input for the other six profiles. Groundwater levels were measured at variable intervals. Linear interpolation was performed to derive daily values of groundwater levels as input. The upper boundary conditions were controlled by daily weather data collected from nearby weather stations. Simulation year was 1997 for GDL, 2000 for the three DHR soils and 1999 for KM1, KM1K, and NYL249. The simple crop growth routine of SWAP was used for each field. Factors to characterize plant growth were derived by adjusting general factors suggested by van Dam et al. (1997), Tiktak et al. (2000), and therein to local conditions. Soil hydraulic characteristics were described according to the Mualem-van Genuchten model. Measured saturated hydraulic conductivity (*Ks*) coupled with an assumed $L = 0.5$, as suggested by Mualem (1976), was used to describe unsaturated hydraulic conductivity [$K(h)$] for each soil. Simulations were run to calculate the soil-water profile of each soil, using different WRCs, as estimated by the various PTFs, but keeping two of the parameters (i.e. *Ks* and *L*) constant for each run with the same soil. As a control, we also ran the same simulations using laboratory-measured WRCs, and applying the same *Ks* and *L* values as above. Simulated moisture content data were then compared with measured data at corresponding depths and dates. Two weights were introduced to account for the different number of layers and days at which moisture contents were measured, to allow each soil to contribute equally to the averaged values.

Evaluation criteria

The calibrated NN models were used to make estimates of the van Genuchten parameters of the 236 soils of the HUNTEST data set. In turn these parameters were converted to water contents at the matric potentials that correspond to those available for the original WRC measurements. Accuracy of the estimations was evaluated using two measures. The mean residual (MR) can quantify systematic errors between measurements and estimations and the RMSRs can give the accuracy of the estimations in terms of standard deviations. These measures are calculated as:

$$MR = (1/N)\sum_{i=1}^{N}(\theta_i - \hat{\theta}_i) \qquad [2]$$

and

$$RMSR = \sqrt{(1/N)\sum_{i=1}^{N}(\theta_i - \hat{\theta}_i)^2} \qquad [3]$$

where N is the number of estimated and measured values, θ and $\hat{\theta}$ are measured and estimated water contents, respectively.

Results and Discussion

Estimation of water retention curves

Results of estimations with the 11 models are summarized in Figure 2. Root mean squared residuals in Figure 2a exhibit a trend of improvement from Model M1 to Model M11. More input information generally leads to better estimates. When one retention point was included in the list of inputs, the water content at –1500 kPa was the worst additional predictor as compared with water contents at –33 and –10 kPa. A possible reason for this is that water content at –1500kPa is more dependent on soil texture than on soil structure, whereas water contents at –33 and –10 kPa are also largely influenced by soil structure. Soil structure is known to have influence on water conditions and as soil texture is among the basic input parameters for all models in our study, it seems to be more beneficial to include such additional input parameters that have a relation to soil structure as well. Estimations using data from the HUN data set show RMSR values ranging from 0.02 to 0.06 m^3m^{-3}, with only the soil texture (M1) model having RMSR >0.045 m^3m^{-3}. These are relatively accurate estimations when compared to estimations that appear in literature (Wösten et al. 2001a). For models M1 to M4, estimations using the EUR data set provide RMSR values that are 0.02 m^3m^{-3} greater than the RMSR using the HUN data set. Difference between the RMSR of the EUR and HUN data sets using any other models was ≈0.01 m^3m^{-3}. The accuracy of estimations was improved marginally for all models when Hungarian data were added to the European data set (i.e., the EUR + HUN data set was used). With the ICO data, RMSR values range from 0.045 to almost 0.09 m^3m^{-3}, that is, 0.02 to 0.04 m^3m^{-3} greater than RMSR values of PTFs developed using the HUN data set. There are, however, large improvements in some cases when Hungarian data were added to the ICO data set (ICO + HUN). This is especially visible for M3, where OM content was added to the models. Although the representation of Hungarian data in the ICO + HUN data set is ≈15%, thus greater than in the EUR + HUN data set, it only improved estimations somewhat, but even then it never performed better than any of the smaller scale data sets. This is presumably due to the fact that many soils in the ICO data set come from areas where conditions that govern soil development may be far from the Hungarian conditions. Because of geographical proximity, soil-forming conditions of other European countries may differ less from the Hungarian conditions. This may result in smaller differences in soil properties, as shown in Table 1, explaining the better estimations.

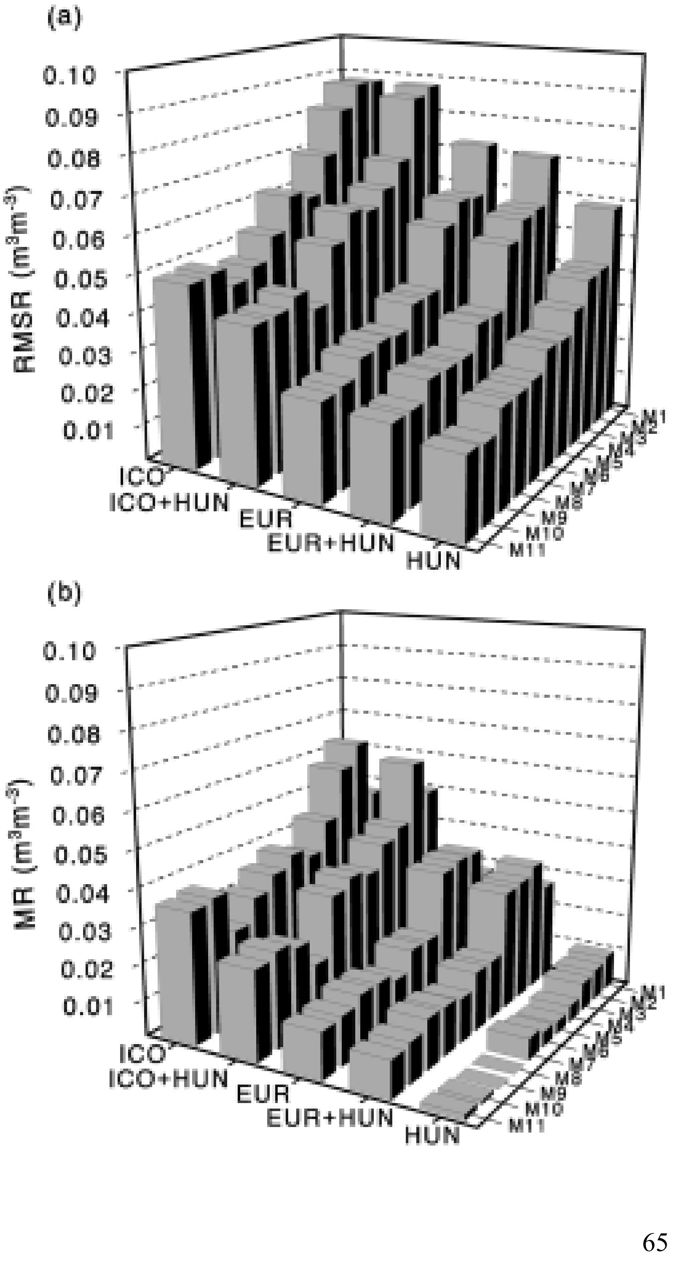


Figure 2 Estimation errors in terms of (a) root mean squared residuals (RMSRs) and (b) mean residuals (MRs) of each of the eleven models and five data sets used to develop pedotransfer functions. HUN, Hungarian data set; EUR, European data set; ICO, Inter-continental data set. Input parameters of models M1,M2 … M11 are listed in Table 2.

Mean residual values also show a trend of improvement as the list of input variables increases (Figure 2b). Most bias is introduced by the PTFs developed using the ICO (and ICO + HUN) data sets (MR is between 0.016 and 0.063 m^3m^{-3}). PTFs developed using the European scale data sets showed bias from 0.009 to 0.036 m^3m^{-3}, with slightly more accurate estimations when Hungarian data were included. Bias for the HUN data set always remained <0.01 m^3m^{-3}. It is interesting to see that bias always remained positive, reflecting an underestimation of water contents (cf. Eq. [3]). This may be caused by the considerable differences among data sets in OM contents (see Table 1). In general, soils with greater OM contents tend to retain more water at the same matric potentials than soils with lesser OM contents, which may be a direct effect of greater OM contents or an indirect effect through the improvement in soil structure stability. For this reason, data sets with mostly lesser OM contents may estimate lesser water contents. This seems to be justified when we notice that improvements in MR caused by the inclusion of Hungarian data in the larger scale data sets are largest when OM was one of the input parameters (i.e., Models 3, 11). It is especially clear in the example of the ICO set with considerably lesser average OM contents than the HUN (and HUNTEST) data set (mean: 18.43 vs. 13.82 g kg^{-1}; median: 17.20 vs. 3.45 g kg^{-1}).

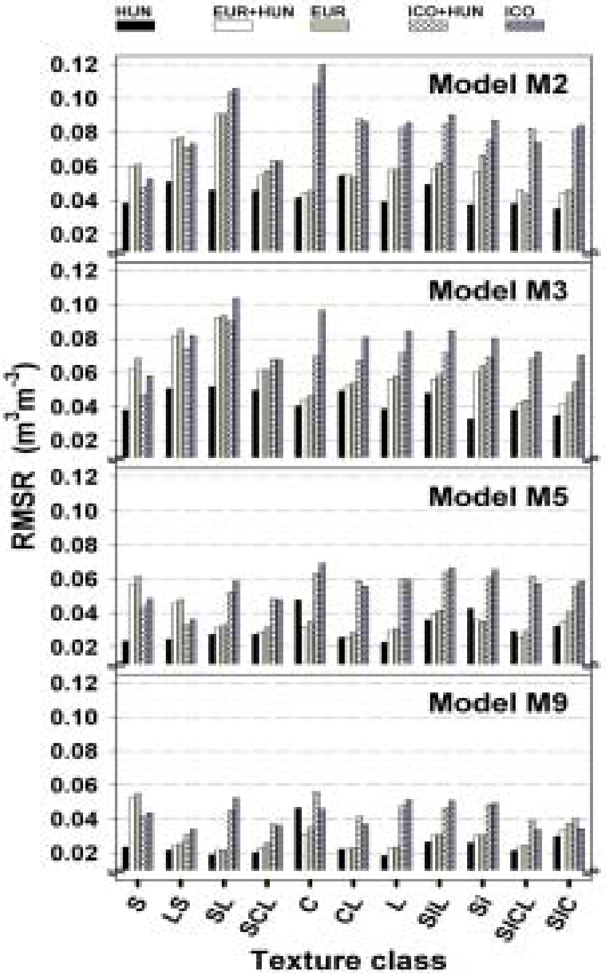


Figure 3. Root mean squared residuals (RMSRs) for each USDA soil texture class using four selected models (M2, M3, M5, M9) and each of the data sets used to develop pedotransfer functions. HUN, Hungarian data set; EUR, European data set; ICO, Intercontinental data set. Input parameters of the four models are listed in Table 2.

In Figure 3, RMSR values are stratified by USDA soil texture classes. Results of four selected models (M2, M3, M5, M9) are shown as examples. A general trend of improvement can be seen with increasing input to the models. For most texture classes, accuracy of PTF estimations was best using the HUN data set, followed by the EUR+HUN, EUR, ICO + HUN and ICO data sets. Addition of Hungarian data to the two international sets resulted in marginal improvement only. Pedotransfer functions using the EUR and EUR + HUN data sets provided worst estimations at the coarse end, thus for sand, loamy sand, and sandy loam soils. The large difference between the average OM contents of the HUNTEST and EUR data sets for the sand and loamy sand classes (Table 4) are one possible reason for a systematic overestimation. In fact, inclusion of OM content (as in model M3) made estimations worse for the above texture classes and the sandy clay loam class, as compared with estimations by model M2. Estimations by PTFs developed from the ICO and ICO + HUN data sets with models M2 and M3 that do not use water retention points as input, are considerably worse for sandy loam and clay soils than for other texture classes. The texture of Hungarian clay soils is heavier than that of the two international data sets, and their OM content is the greatest of all texture classes (Table 4); thus these soils are largely different from the clay soils of the ICO data set. Besides, the ICO data set is under-represented in clay soils (Figure 1). As for the sandy loams, the fact that PTFs from both international data sets make large errors may indicate that Hungarian soils of this texture are, in some way, different from those of the international data sets without the basic soil data explaining it. This is supported by the fact that for these classes estimations improve considerably using models M5 and M9 (thus models that use water retention data also) as compared with models M2 and M3. Measured water contents in the Hungarian data set for the sandy loam texture class are significantly (0.08-0.10 m^3m^{-3}) greater throughout the entire WRC than in the international data sets, without differences in the soil survey data explaining it.

Table 4. Average organic matter (OM) contents [g kg^{-1}] by USDA texture classes of the test data set (HUNTEST) and the Hungarian (HUN), European (EUR), and Intercontinental (ICO) data sets.

	Organic Matter Content			
Texture class	HUNTEST	HUN	EUR	ICO
	-------------------- g kg^{-1} ----------------------			
sand	1.81	3.05	14.70	11.11
loamy sand	3.48	6.68	12.98	15.05
sandy loam	15.86	14.06	14.99	14.23
sandy clay loam	9.80	13.43	12.91	5.45
sandy clay	--	--	9.40	1.94
clay	25.34	32.26	21.44	7.32
clay loam	22.60	17.00	22.98	5.20
loam	19.84	23.00	17.58	17.97
silt loam	17.97	20.85	11.40	20.75
silt	21.10	23.98	3.66	3.72
silty clay loam	14.91	21.34	18.37	5.08
silty clay	23.05	20.76	22.49	12.61

In Figure 4, RMSR values are shown for the same four models as in Figure 3, stratified by matric potential values. Clearly, there is improvement in estimations throughout the entire WRC with using more input in the models. In general, a hierarchy of PTFs can be observed: using the HUN data set as input being the best and PTFs using the ICO data set being the worst, with only some exceptions. A trend of special interest is the relatively large RMSR for water contents at saturation and at -0.25kPa. While accuracy of estimations at saturation using the HUN data set remains similar to the accuracy at other points in the wet range of the WRC, all PTFs developed on international data sets had larger RMSRs at saturation. Differences in OM content that were detailed above may explain part of the larger errors; however, different structural development and macroporosity of soils in the different data sets used for PTF development may also have influenced estimations. In Figure 4, the RMSRs of the direct fitting of the van Genuchten model to the measured water retention data of the HUNTEST data set are also shown. The possible presence of macroporosity, which is not accounted for in the van Genuchten equation, is reflected by the higher RMSR of the direct fit at saturation. As the comparison of measured and estimated WRCs was made through the use of the van Genuchten model, calculated errors will also include an element of error that result from the nonperfect fit of the van Genuchten model to the measured water retention points.

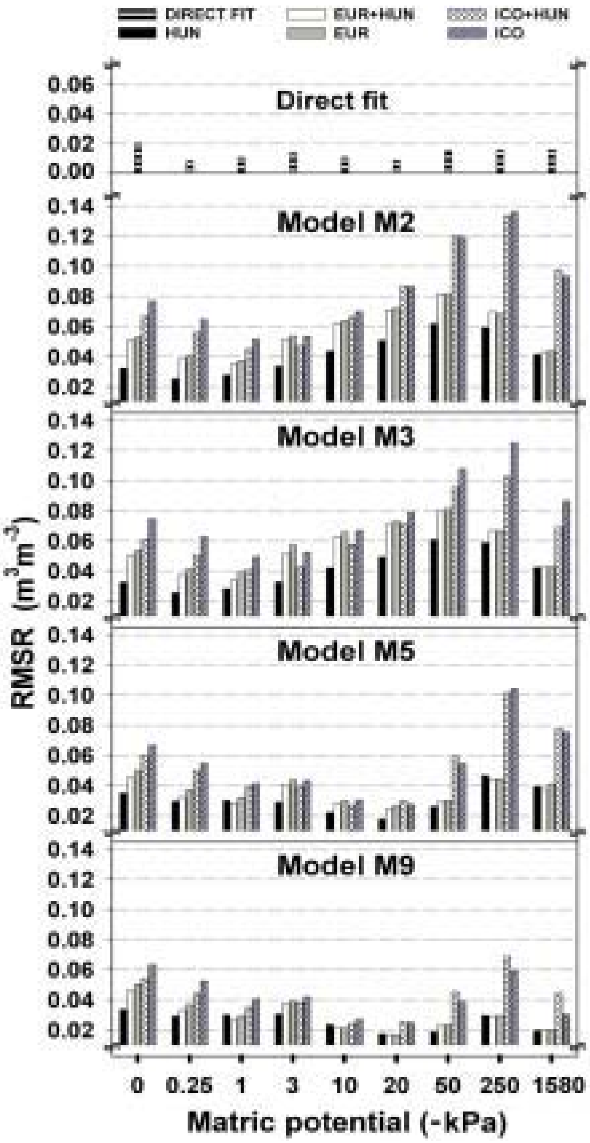


Figure 4. Root mean squared residuals (RMSRs) for several matric potential values for the direct fitting of the van Genuchten equation and using four selected models (M2, M3, M5, M9) and each of the data sets used to develop pedotransfer functions. HUN, Hungarian data set; EUR, European data set; ICO, Intercontinental data set. Input parameters of the four models are listed in Table 2.

Large RMSRs for the dryer range of the water retention mean an obvious failure of the ICO (and ICO + HUN) data sets. We found that the largest part of this error originated from soils with textures that were underrepresented in the ICO data set: clay, silt, silty clay, and silty clay loam (c.f. Figure 1). We note that smaller RMSRs at some matric potentials may have resulted from the fact that some NN models included one, two, or three water retention values at matric potentials close to the estimated values. This is the reason why major improvement is seen from M3 to M5 at matric potentials of –10, -20 and –50kPa and from M5 to M9 at –1580 kPa.

Functional evaluation on water content time series

One possible application of estimated WRCs is their use in numerical models for simulations of water content and solute transport dynamics. Three of the 11 PTFs (M2, M5, and M9) were further used in combination with all five data sets to simulate the soil moisture profiles of seven Hungarian soils. M2 was selected as a widely applicable very simple model that requires only texture and BD as input. M5 and M9 were selected as the best models that use one (M5) or two (M9) additional water retention points as input. Water retention points required by the selected models are often measured, as those are in common use for the calculation of water holding capacity. Simulated and measured water contents were then compared at each depth for each available date and for each profile. Root mean squared residuals and MR were calculated as defined earlier, with the replacement of measured and estimated WRC data with field-measured and simulated soil water contents respectively. As a control, allowing further comparisons, simulation was also run using laboratory-measured WRCs, and the same measures were calculated as above.

Table 5 summarizes the results. Averaged RMSR values ranged from 0.046 to 0.093 m^3m^{-3} considering the different data sets of PTF development, and from 0.048 to 0.090 m^3m^{-3} using the measured WRCs. While the focus should be on the averaged errors obtained using the particular data set or NN model, we would like to point out the case of two soils that show the worst results using any input data: GDL and NYL249. One possible cause of that for the GDL soil is the large heterogeneity of the soil (Cs. Farkas, 2002, personal communication), which makes sampling an intricate issue. The RMSR for NYL249 soil was probably greater because it is a soil type (Natrustoll) in which the salt content alters its physical properties. That factor was not accounted for in our PTFs and model runs, and is one of the potential pitfalls of most current PTFs. The average of all profiles is between 0.065 and 0.07 m^3m^{-3} for the five data sets, with the HUN set the best, and the ICO + HUN set the worst. Using measured data, the average RMSR is 0.061 m^3m^{-3}. Errors by the international PTFs were only a fraction larger (0-0.005 m^3m^{-3} on annual average) than errors by the national scale PTFs, and remarkably, the different scale PTFs were only 0.004 to 0.009 m^3m^{-3} worse on an annual basis than laboratory-measured WRCs, which may be a strong argument for PTFs. Inaccuracy of simulations even using original measured data may reflect inaccurate settings of some other input parameters (e.g., vegetation data), non-optimal numerical solution(s) in the simulation model, or it may as well suggest that our conventional laboratory techniques produce errant WRC data compared with true field conditions (Pachepsky et al., 2001). If errors are grouped by models, M2 to M9, there is improvement, but there is only a difference of 0.002 m^3m^{-3} in water content between the best (M9: 0.065 m^3m^{-3}) and the worst (M2: 0.067 m^3m^{-3}).

Table 5. Summary of the simulation results in terms of root mean squared residuals (RMSRs) and mean residuals (MRs) [m^3m^{-3}].

	measured	values averaged per data set					averaged per model		
Soil	WRC	HUN	EUR	EUR+HUN	ICO	ICO+HUN	M2	M5	M9
	------------------------ $m^3\ m^{-3}$ ------------------------								
				RMSR					
AVERAGE	**0.061**	**0.065**	**0.065**	**0.066**	**0.067**	**0.070**	**0.067**	**0.066**	**0.066**
GDL	0.067	0.090	0.093	0.088	0.084	0.093	0.084	0.090	0.094
DHR2	0.052	0.064	0.047	0.047	0.064	0.054	0.065	0.055	0.045
DHR3	0.053	0.057	0.084	0.083	0.063	0.065	0.064	0.072	0.076
DHR4	0.049	0.046	0.057	0.057	0.050	0.052	0.047	0.052	0.058
KM1	0.065	0.057	0.049	0.051	0.057	0.055	0.054	0.056	0.051
KM1K	0.053	0.066	0.057	0.061	0.058	0.077	0.067	0.063	0.061
NYL249	0.090	0.074	0.068	0.076	0.093	0.093	0.091	0.076	0.075
				MR					
AVERAGE	**-0.009**	**0.004**	**0.020**	**0.018**	**0.005**	**0.020**	**0.021**	**0.007**	**0.012**
GDL	0.049	0.075	0.081	0.076	0.071	0.080	0.071	0.077	0.081
DHR2	0.000	-0.032	-0.003	-0.003	-0.038	-0.027	-0.035	-0.011	-0.016
DHR3	-0.013	0.005	0.032	0.030	0.016	0.022	0.022	0.019	0.022
DHR4	-0.016	0.003	0.023	0.021	0.005	0.014	0.012	0.009	0.019
KM1	-0.034	-0.029	-0.015	-0.017	-0.035	-0.016	-0.019	-0.029	-0.019
KM1K	0.010	0.021	0.025	0.027	0.022	0.041	0.033	0.024	0.025
NYL249	-0.061	-0.016	0.000	-0.006	-0.008	0.022	0.063	-0.040	-0.029

The trend is similar for the MRs, with GDL and NYL249 showing the largest bias in the simulations with estimated as well as with measured WRCs. On average, PTFs using the HUN and the ICO data sets result in biases that are even smaller than the bias of the measured input data (0.009 m^3m^{-3}), but for all data sets, bias remains at or under 0.02 m^3m^{-3}. Regarding the grouping by M2, M5, and M9 models, M2 shows a somewhat larger bias (0.021 m^3m^{-3}) whereas the other two remain at or under 0.012 m^3m^{-3}. It is interesting to note that there are large differences among the biases of M2, M5, and M9 for the NYL249 Natrustoll. Model M2, which does not use measured water retention data as input, had a bias that was ≈0.1 m^3m^{-3} greater than M5 and M9. Differences between RMSR values were much smaller, which suggests that M2 was able to find the correct shape of the WRC but not the correct position of the air entry value.

The presentation of averaged values, as in Table 5 for example, may hide large differences between summer and winter periods, wet and dry periods, top- and subsoils, or among certain soil types. Such possible differences were examined. We found no significant difference between simulation errors for wet and dry or for winter and summer periods. There were differences among soils, but no correlation could be shown with texture or any other examined soil properties besides the influence of salinity in one soil. The only significant and systematic difference was the increasing underestimation (i.e., lesser MR) with depth. The change was irregular; however, for most soils MR was more negative by 0.04 to 0.06 m^3m^{-3} below 50 cm, than in the top 50 cm. This could be observed with all PTFs, but also when laboratory-measured WRC was used in the simulation model, which indicates that usage of PTFs were not the particular reason for such deviation.

Figure 5 shows an example of the simulations of water content time series for all seven depths of one soil (KM1). Five curves in each subplot show daily model outputs of soil water contents using estimated WRCs from each of the five data sets using the M5 model. An additional curve shows the daily outputs when laboratory-measured WRCs were used in the simulation model.

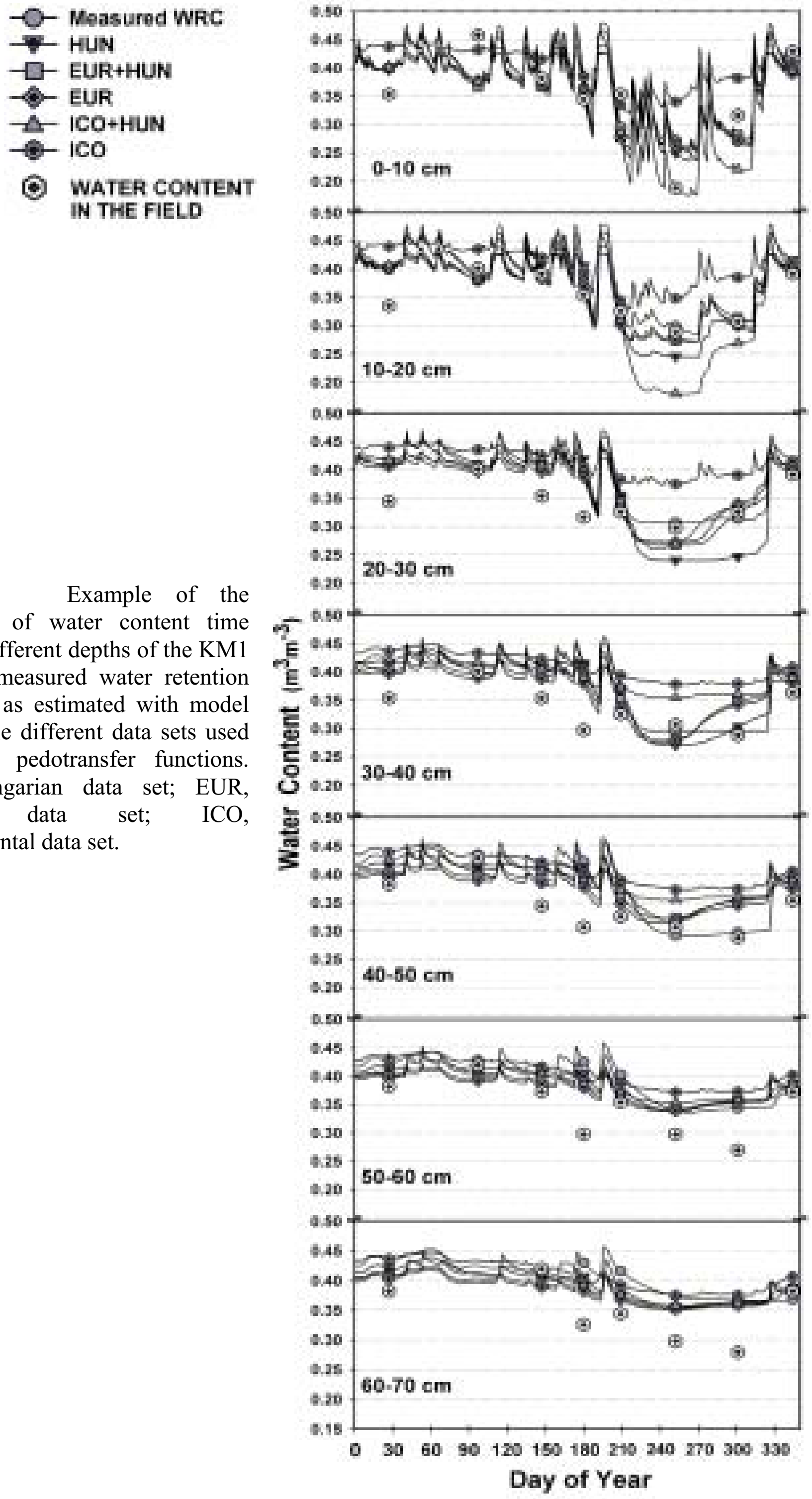


Figure 5. Example of the simulations of water content time series for different depths of the KM1 soil using measured water retention curves and as estimated with model M5 from the different data sets used to develop pedotransfer functions. HUN, Hungarian data set; EUR, European data set; ICO, Intercontinental data set.

Simulated water contents are marked with symbols at those days for which field-measured water contents were available. Day-to-day changes in the simulated moisture contents in the top layers are apparent. These changes gradually weaken with depth. Differences between simulations using WRCs obtained from the different PTFs (and the laboratory measurement) are much larger in the top layers, specially in the dry period of the year. For this soil, at most depths, curves representing PTFs developed from the ICO and ICO+HUN data sets show largest deviations from the field-measured water contents. For the subsoil layers, simulations continuously overestimated water contents during drier periods, no matter which WRC was used. At this site, temperatures <0° Celsius often occur in the first 30 to 60 days of the year. Large overestimation of field water content at Day 27 in the top horizons probably occurred due to the nonoptimal solution applied in the simulation model for conditions when, in reality, infiltration and soil water flow is limited by temperatures below freezing point (J.G. Kroes, 2002, personal communication). As also suggested in the previous paragraph, any of the presented characteristics of simulations were not necessarily the same for the other six soils.

Figure 6 summarizes all such measured/simulated data pairs using estimations by the M5 model, as an example, for all depths of all test soils and by all five data sets. Deviation of the regression line from the 1:1 line is very small. The slope and intercept parameters do not differ from 1 and 0, respectively, at a 95% confidence level. Table 6 lists parameters and coefficients of determination (R^2) of linear regressions, obtained as in Figure 6, for each group between measured and simulated water contents. For any of the data sets used to develop PTFs, the regression equation introduces very small bias (offset). Bias found for the different development data sets was smaller then the bias using the measured WRCs. For all data sets (and the measured WRCs), the intercept parameter did not differ from 0 at a 95% confidence level. The slope parameter of the line remains around (but always below) one, meaning a slight underestimation in wet periods. This parameter differed from 1 significantly (95% confidence) for the EUR, ICO, and ICO + HUN data sets. When data are grouped by models, superiority of M5 and M9 is apparent over M2, which despite of its small bias shows the largest deviation from the unit slope (0.87). Both parameters of M2 and the intercept parameter of M9 differed from the optimal 0 (intercept) and 1 (slope) at the 95% significance level. Regarding the coefficients of determination, the simulations using measured WRC data show largest R^2 (0.709), but simulations with estimated WRC data provide in most cases only somewhat lesser values. Worst of the data sets in this respect is ICO + HUN with an R^2 of 0.627.

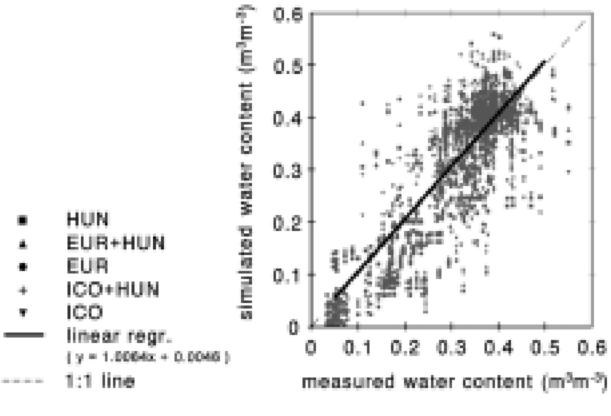


Figure 6. Measured vs. simulated water contents. All depths of all seven soils are shown using water retention curves estimated with model M5 from each data set of pedotransfer function development. HUN, Hungarian data set; EUR, European data set; ICO, Intercontinental data set.

Table 6. Parameters of linear regression and coefficients of determination (R^2) describing the relationship between measured and simulated soil water contents, grouped by data sets and by models.

	regression coefficients				
	slope		intercept		
data group	mean	SE	mean	SE	R^2
measured WRC	0.9807	0.0313	0.0186	0.0102	0.709
HUN	0.9937	0.0191	-0.0014	0.0063	0.691
EUR	0.9634	0.0184	-0.0063	0.0060	0.694
EUR+HUN	0.9671	0.0190	-0.0053	0.0062	0.683
ICO	0.9484	0.0196	0.0123	0.0064	0.659
ICO+HUN	0.9377	0.0208	-0.0024	0.0068	0.627
M2	0.8715	0.0150	0.0135	0.0049	0.626
M5	1.0064	0.0150	-0.0046	0.0049	0.690
M9	1.0082	0.0147	-0.0108	0.0048	0.701

In summary, differences in simulation accuracy, as a result of PTFs derived from different scale data sets, were much smaller than one would expect from differences obtained while evaluating WRCs only.

Conclusions

Soil hydraulic PTFs to estimate water retention characteristics were developed using Hungarian and international data sets and were evaluated using a separate Hungarian data set. Pedotransfer functions developed using a European data set provided somewhat larger errors than PTFs developed using only Hungarian data. PTFs developed using an intercontinental data set (containing data mainly from the USA and Europe) provided much larger errors. Inclusion of data from Hungary in the international data sets resulted in only small improvement of the PTFs. It suggests that having a small set of relevant data, when available, is better than using a large but more general data set.

Surprisingly, the differences among national and international PTFs largely disappeared when the estimated WRCs were used for simulations of soil water contents. Our findings support the success and effectiveness of PTFs. It has to be decided whether the accuracy of simulations that produce RMSR of 0.065 to 0.07 m^3m^{-3} using PTF-based soil hydraulic properties is satisfactory for a particular application. However, using measured WRCs lead to RMSR <0.01 m^3m^{-3} better than PTFs, which is an argument for PTFs. These indicate that for the presented case study, differences between estimations using different scale data sets (or measured WRCs) were not the main source of error in the simulation model, PTFs errors were overwhelmed by errors resulting from other factors. One should, however, still apply PTFs with care, keeping in mind limitations that were discussed: the fact, that not all soils are equally represented in any of the databases and other potential pitfalls set by different structure, clay mineralogy, salinity, and other factors that largely influence soil water status and that are not accounted for in most PTFs.

Acknowledgements

The authors wish to thank NATO (CLG 975761) for financial support for travel. One of us (MGS) was supported by the SAHRA science and technology center under a grant from NSF (EAR-9876800). A. Nemes wishes to thank the Ministry of Agriculture, Nature Management and Fisheries for financial support that was provided through the International Agricultural Centre (IAC) of Wageningen, The Netherlands. We thank three anonymous reviewers for their constructive comments.

Chapter 6

Soil Water Balance Scenario Studies Using Predicted Soil Hydraulic Parameters

Submitted for publication to Hydrological Processes

Soil Water Balance Scenario Studies Using Predicted Soil Hydraulic Parameters

A. Nemes[1], J.H.M. Wösten[2], J. Bouma[2] and G. Várallyay[1]

[1] Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences, Herman Ottó u. 15, H-1022 Budapest, Hungary

[2] ALTERRA Research Institute for Green World Research, P.O. Box 47, 6700 AA Wageningen, The Netherlands

Abstract

Pedotransfer functions (PTFs) have become a topic drawing increasing interest within the field of soil and environmental research because they can provide important soil physical data at relatively low cost. Few studies, however, explore which contributions PTFs can make to *exploratory* planning, in terms of examining the expected outcome of certain changes in soil and water management practices. This paper describes three scenario studies to demonstrate how PTFs may help improve land management practices. We use *exploratory* modeling to: (i) evaluate benefits and risks when irrigating a field, and the impact of soil heterogeneity; (ii) examine which changes can be expected – in terms of soil water balance and supply - if bulk density and/or organic matter content are changed as a result of alternative management systems; and (iii) evaluate the risk of leaching to deeper horizons in some soils of Hungary. Using an *exploratory* research approach, quantitative answers are provided to “what if” questions, allowing the distinction of trends and potential problems and successes, which may contribute to the development of sustainable management systems.

Introduction

The quality of our environment faces constant pressure and is endangered by various kinds of human (among others, agricultural) activities. Agricultural production – if not performed in a sustainable way – may adversely effect soil and water resources. Chemicals and nutrients are transported through and/or stored in our soils. Sophisticated simulation models are being developed to describe water and solute movement in the unsaturated zone of the soil. Other models are being developed to simulate heat and water fluxes at the soil surface, to assist research on possible climate change. One of the common input needs of these models are the soil hydraulic properties, namely water retention and saturated and unsaturated hydraulic conductivity. However, measurement of these properties is relatively time-consuming and costly, especially when data are needed for large areas of land. Such data are usually measured with different degrees of detail and reliability and are stored in different institutes and/or databases.

Research on predictive functions that derive difficult or expensive-to-measure soil properties from easily or routinely measured soil properties is on the rise. Bouma (1989) introduced the name ‘pedotransfer function’ (PTF) for such predictive functions. Recently published reviews on PTF development and use include e.g. Pachepsky et al. (1999) and Wösten et al. (2001a). Databases of different sizes, scales and detail are available to develop PTFs that predict soil hydraulic properties (c.f. Wösten et al., 2001a). Many studies compare and/or validate the performance of different PTFs. Recent publications on the topic include Kern (1995), Tietje and Hennings (1996), Schaap and Leij (1998), Imam et al. (1999), Cornelis et al. (2001), Wagner et al. (2001) and Minasny and McBratney (2002) among others. Some studies go further and evaluate the functionality of pedotransfer functions (e.g. Wösten et al., 1995; Hack-ten Broeke and Hegmans, 1996; van Alphen et al., 2001; Soet and Stricker, 2002; Nemes et al., 2003). Studies that belong to the first group, evaluate how well certain functions predict water retention (or hydraulic conductivity), whereas the second group of studies evaluates the performance of predicted soil hydraulic characteristics through the simulation of some practical aspects of soil behavior.

Scenario studies are important in planning and prevention. Changes in management or in the application regimes of certain substances can be planned and optimized, while measures to prevent undesired or hazardous effects of certain management practices can be implemented. Field experimentation representing different management possibilities would be time consuming, costly and sometimes even risky. Exploratory (‘what if?’) modeling offers an alternative that is quicker and easier to execute, and may give at least indicative answers about trends that are expected to occur. Justified PTFs can assist and enhance such modeling, as they can provide low-cost and low-risk input data without causing actual changes to our environment, which could result from on-site experimentation. The possibility to compare such simulations with (field) measurements is limited, so one has to be careful with the interpretation of results. Bouma et al. (1999) used plant growth and nutrient use as indicators for the functionality of different structural types and sizes. They performed exploratory modeling to identify the ‘ideal’ structure for a certain soil in a ‘pro-active’ approach. They suggest a desired pair of aggregate size and bulk density, and leave it to future practicioners to design and construct it. Droogers and Bouma (1997) analysed the degree of sustainability of different management scenarios on three different ‘phenoforms’ (management types) of the same ‘genoform’ (soil type). They used exploratory modeling to define a range of management options to choose from, realizing – for each case – a different balance between production level and environmental risks.

In this study we demonstrate possible uses of soil hydraulic pedotransfer functions. We show three examples in which PTFs are used in *exploratory* modeling studies. The effect of different irrigation practices on different soils is examined in the first study, pointing out difficulties being faced, when a particular field is heterogeneous. In the second study, we try to quantify the expected effect of changes in bulk density and organic material content on the soil water balance. The third example evaluates the probability and possible risk of a downward flux in ten different Hungarian soils. The three examples are derived from what seem to be realistic problems being faced in Hungary along with many other countries.

Materials and Methods

Location of the soil profiles and texture of the soil horizons used in the three studies are shown in Figure 1.

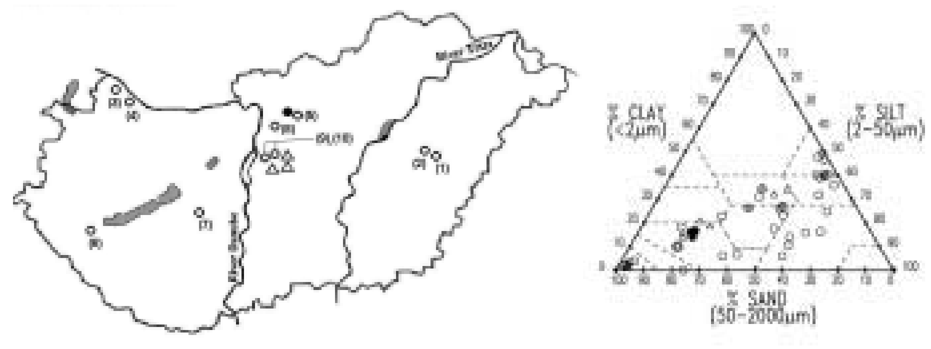


Figure 1. Location of profiles and textural distribution of the soil horizons used in the paper. Dark spots: soil in Study 1; triangles: soils in Study 2; circles: soils in Study 3, with the Profile No. in brackets (c.f. Table 1).

Study 1

The objective of the first scenario study is to examine the effect of different irrigation scheduling practices on soil water balance components and the soil's ability to supply the vegetation with water. Further, we show – using a simple comparison - the effect of soil heterogeneity on those water balance components.

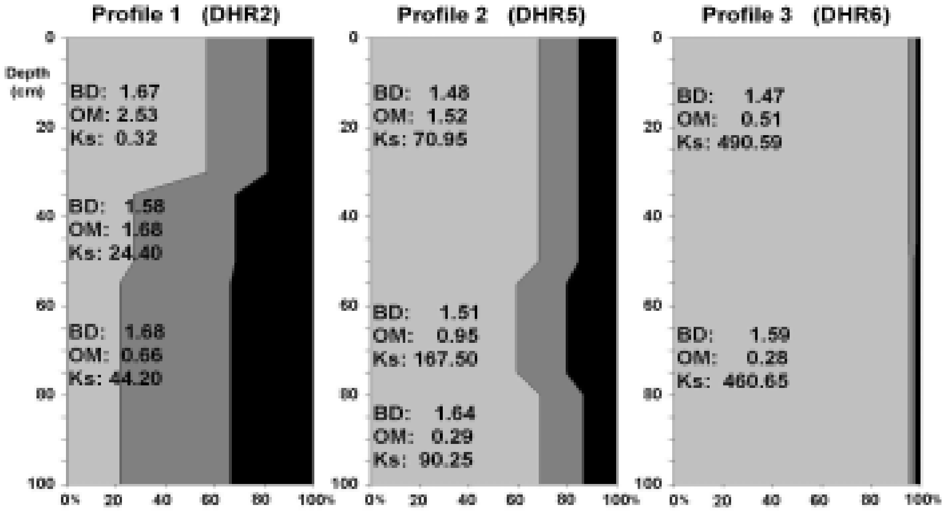


Figure 2. Texture and other physical properties of the top 1 meter of the three soils used in Study 1. Light grey: sand; dark grey: silt; black: clay content. BD is in gcm-3, OM is in % and Ks is in cm/day.

Three soil profiles were sampled within the same field near Dunaharaszti in the central part of Hungary (Czinege, 2000). This field has a history of major heterogeneity in crop yield. The sampled locations fall within 50 meters from each other. The profiles had 2 to 4 distinct horizons with measured basic soil physical properties. Textural differences among the top 1 meter of the three profiles and further details can be seen in Figure 2. Textural differences among the three profiles are large.

Simulations of one-dimensional flow were carried out on a daily output basis, using the SWAP simulation model (van Dam et al., 1997). Groundwater levels measured at variable time intervals were used as lower boundary condition for the simulation model. Depth of groundwater table changed between 0.6 and 1.7 meters throughout the year. Linear interpolation was performed to derive daily values of groundwater levels as input. The upper boundary conditions were controlled by simulated daily weather data. The stochastic weather generator of Semenov and Barrow (1997) was used to generate a one-year-long data set of daily average temperature, solar hours and amount of precipitation. These data were in turn used to calculate daily values of potential evapotranspiration using an empirical equation developed for the middle part of the Hungarian Great Plain area (Fodor, personal comm.). A drier and warmer than average year was simulated which both had a probability of occurrence of 25-25% (sum of annual precipitation: 442 mm; annual average temperature: 11.24 °C, calculated sum of annual potential evapotranspiration: 1106mm) to allow emphasis on precipitation/irrigation. The simple crop growth routine of SWAP was used with grass as cover crop. Factors to characterise plant growth were derived by adjusting general factors suggested by van Dam et al. (1997), and Tiktak et al. (2000) to local conditions. Soil hydraulic characteristics were described according to the Mualem-van Genuchten model. The national-scale M10 pedotransfer function of Nemes et al. (2003) was used to predict soil water retention for all soil horizons. This PTF uses neural network optimization to predict the van Genuchten parameters (van Genuchten, 1980) from soil texture, bulk density and two measured water retention points (h=-33 kPa and -1500 kPa). Soil texture, bulk density and organic matter content were used as input to a neural network model (Nemes, 2003) to predict saturated hydraulic conductivity (K(sat)). Predicted K(sat), coupled with an assumed $L=0.5$ – as suggested by Mualem (1976) - was used to describe unsaturated hydraulic conductivity ($K(h)$) for each soil horizon.

Multiple simulations were run for each of the three profiles to account for different irrigation scheduling. A "simple" (empirical) irrigation scheduling technique – as used in practice - was based on simple decision criteria that relied on daily observed weather data, as follows: (1) During the period from May to September, 20mm of irrigation water was applied on a particular day, if the average daily temperature of the preceding five days had been above 20 Celsius, and the sum of the precipitation for the same period did not exceed 1mm; (2) The same amount of irrigation was applied when daily average temperature fell between 17-20 Celsius for the preceding 8 days, with less than 1mm of precipitation. This way, irrigated days for this field were at least 6 days apart, which may be a realistic situation in practice, considering the rotation of irrigation between fields. In total, 260 mm of irrigation water was applied during the irrigated period. The alternative ("advanced") irrigation scheduling technique used decision criteria based on a preceding run of the simulation model. Using the measured weather data, the model calculated when the vegetation suffered from water stress. When the daily actual plant transpiration fell below 85% of potential transpiration, we applied 20mm of irrigation water, if there was no other irrigated day in the preceding five days (i.e. considering the same rotation of irrigation as above). In this scheme, a total of 340 mm of irrigation water had been applied – 80 mm more than in the simple irrigation scheduling system, giving an early indication that the "simple" system does not supply enough water to cover the needs of the vegetation.

As the three profiles are close enough to each other to fall within the same irrigation unit in reality, we calibrated the latter scheduling technique to the profile with the intermediate texture (Profile 2), and then applied the resulting irrigation schedule to the remaining two soil profiles as well. As a control, we also ran simulations for each profile without any irrigation.

Study 2

The second study aims to examine the expected effect – in terms of water balance components - of some changes typically introduced by tillage or land management practices. The area of the brown forest soil profile used in this study is located in the North-Central part of Hungary near Gödöllő and was described in detail by Farkas et al. (2000). It is a relatively homogeneous, sandy Loam textured soil, with 63-65% sand and 14-16% clay content throughout the profile. Bulk density decreases from 1.66 g/cm^3 to 1.56 g/cm^3 from the top of the soil to the depth of 70 cms. The organic matter contents are 2.81, 2.3 and 1.5 % in the A, B and C horizons respectively.

We assumed two simple changes that can both be introduced by different tillage or land management practices. In the first case, we assumed that the top 40 cms (the A and B horizons) have been loosened to a bulk density (BD) of 1.4 gcm^{-3}. This can be achieved in practice for instance by plowing. The other change we examined can only be introduced in a longer period of time: we doubled the organic matter (OM) content of the top 40 cms – the A and B horizons- to 5.6 and 4.6% respectively. This can be a long term effect of no-till systems and of applying (green) manure. Besides, we applied the combination of these two changes as well. We have to note, that while applying such changes, we remain within a range that is reasonable for Hungarian conditions, and is included within the calibration range of the applied pedotransfer functions.

Simulations of one-dimensional flow were carried out as in Study 1, with the following changes. Assigned groundwater levels at variable time intervals were used as lower boundary condition for the simulation model, with the depth of the groundwater table changing between 1 and 1.8 meters throughout the year. The upper boundary conditions were controlled by simulated daily weather data as above in Study 1. Two one-year-long data sets were generated: a drier and a wetter than average year were simulated (sum of annual precipitation: 442 and 561mm; annual average temperature: 11.24 and 11.22 °C, calculated sum of annual potential evapotranspiration: 1204 and 1228mm respectively) to allow the examination of the effect under both drier and wetter conditions. The simple crop growth routine of SWAP was used with maize as cover crop in the period 1 May – 30 September. Factors to characterise plant growth were derived by adjusting general factors suggested by van Dam et al. (1997), Tiktak et al. (2000) to local conditions. The M3 pedotransfer model of Nemes et al. (2003) - derived from the national scale database - was used to predict water retention for all soil horizons. This neural network PTF predicts soil water retention from soil texture, bulk density and organic matter content. Usage of the relatively simple M3 model was determined by the availability of input data. The same procedure as in Study 1 was applied to obtain saturated and unsaturated hydraulic conductivity for these imaginary situations. Outputted water balance components were the same as above in Study 1, with two differences. First, all measures were summarized for the entire simulation year as organic matter and bulk density changes are not limited to the summer period only. Second, for the maize crop, a pressure head of –60 kPa was suggested in literature as the starting point of limited water uptake.

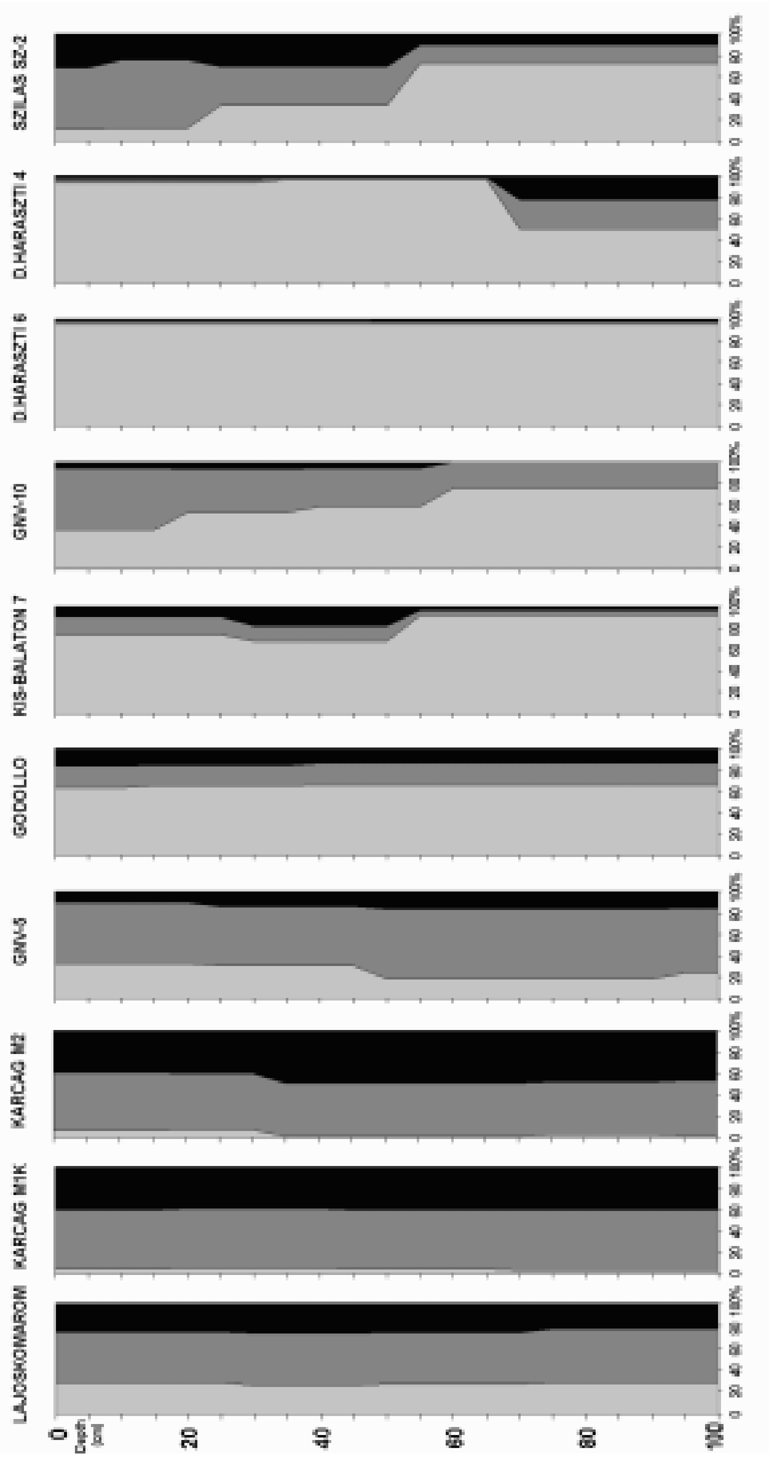


Figure 3. Textural composition of the top 1meter of the ten soils used in Study 3. Light grey: sand; dark grey: silt; black: clay content.

Study 3

The objective of this scenario study was to evaluate and quantify the risk of potential nutrient and pesticide leaching in major Hungarian soil types. The risk of such leaching is higher when downward fluxes of water are more likely to occur. Ten typical soil profiles have been selected to represent ten major soil types in Hungary. These selections took place on the basis of soil genetic classification (c.f. Stefanovits, 1963) with most of the basic categories represented from the hydrophysical categorization of Hungarian soils, as established by Várallyay (1989b). The profiles had 3 to 5 distinct horizons with measured basic soil physical properties. Texture of the ten profiles to the depth of 1 meter can be seen in Figure 3. Details on some soil properties of each site and the significance of the soil types in Hungary (according to Várallyay et al., 1980) can be found in Table 1. Simulations of one-dimensional flow were carried out as in Study 1, with the following changes. We kept the depth of groundwater table below 3 meters from the soil surface throughout the year. This is to achieve a standardized input for bottom boundary conditions, and may not reflect reality completely in all cases. Pedotransfer models used in Study 2 have been applied in this study also to predict water retention and saturated hydraulic conductivity for each soil horizon. Input parameters – for both pedotransfer models – were soil texture, bulk density and organic matter content.

Twenty simulation runs have been executed for each soil using twenty sets of one-year-long simulated weather data. The weather generator program had been calibrated to provide data that is representative for the Southern Great Plain agricultural area of Hungary (simulated average annual precipitation: 508mm; average temperature: 10.75 °C). We examined the frequency and amount of leached water for each calendar day at two depths in the soil profiles.

Table 1. List of some properties and the significance of the soils used in Study 3. BD is bulk density and OM is organic matter content.

Profile No.	Classification Hungarian/FAO	Coverage by the main soil type (Hung.) (% of Hungary *(Várallyay et al., 1980)*)	Name of profile	Depth (cm)	BD g/cm3	OM (%)
1	Lowland Chernozem	4.9	Karcag M1K	0-	1.52	6.48
	Calcaric Phaeozem			21-	1.46	2.99
				46-	1.40	2.54
				68-	1.49	1.83
2	Meadow soil	8.2	Karcag M2	0-	1.41	5.61
	Gleyic Vertisol			20-	1.45	3.87
				36-	1.44	2.91
				75-	1.52	1.99
				95-	1.50	1.22
3	Meadow Chernozem	6.9	GNV-5	0-	1.45	3.02
	Gleyic Chernozem			24-	1.46	2.78
				50-	1.36	-
				97-	1.46	-
4	Alluvial soil with humus accumulation	2.6	GNV-10	0-	1.51	2.68
	Humic Fluvisol			18-	1.50	2.85
				42-	1.43	-
				62-	1.43	-
5	Ramann type Brown Forest Soil	9.4	Gödöllő	0-	1.66	2.81
	Chromic Cambisol			15-	1.63	2.32
				40-	1.56	1.50
6	Brown forest soil with clay accumulation	16.2	Kis-Balaton 7	0-	1.43	3.35
	Haplic Luvisol			30-	1.55	0.85
				56-	1.48	-
7	Chernozem	5.0	Lajoskomárom	0-	1.30	2.91
	Calcic Chernozem			30-	1.46	3.69
				50-	1.40	1.59
				68-	1.48	-
				85-	1.42	-
8	Alluvial-Meadow soil	8.4	Szilas Sz-2	0-	1.10	4.90
	Gleyic Fluvisol			10-	1.13	2.60
				25-	1.32	1.90
				53-	1.52	0.30
9	Coversand	4.2	Dunaharaszti 4.	0-	1.49	0.46
	Haplic Arenosol			35-	1.58	0.10
				70-	1.53	1.32
10	Multilayer humous sand	3.7	Dunaharaszti 6.	0-	1.47	0.51
	Calcaric Arenosol			50-	1.59	0.28

Results and Discussion

Study 1

We used different output measures to evaluate the result of different irrigation schemes. Water deficit to the vegetation (expressed in mm) was added up for the period from 15th May to 15th October. Some characteristics of the flux at one meter depth in the profiles were summarized for the same period: flux balance, the number of days with downward flux, and the sum of daily amounts of net downward flux. Additionally, days were counted, when the average pressure head in the top 50 cm of soil – representing the root zone - exceeded the value that is suggested as the starting point of water stress for grass (–80 kPa, Tiktak et al., 2000).

Table 2. Water deficit (in mm) for the vegetation in the period May 15 – October 15 under different irrigation schemes. The number of days with water deficit are shown in brackets.

	Profile 1		Profile 2		Profile 3	
no irrigation	505.7	(150)	377.7	(128)	121.5	(44)
simple	304.5	(136)	140.5	(57)	3.3	(0)
advanced	281.3	(124)	95.8	(29)	2.3	(0)

Table 2 shows that for each soil, irrigation significantly decreased waterstress for the vegetation. For the heaviest of the three soils (Profile 1) the introduction of the simple irrigation decision criteria brought most improvement – about 40% - in terms of the water deficit. For Profile 2 it is nearly 60%, and for the coarsest soil – Profile 3 - it nearly removes the entire water deficit. The switch to the advanced irrigation scheme caused considerable improvement for the soil with intermediate texture, but not so much for the other two. In terms of water-stressed days for the vegetation, Profile 1 still has a deficit for 4 out of 6 months (124 days), even after the more sophisticated irrigation scheme was applied. This is because the scheduling of irrigation was optimized to Profile 2, a profile with coarser texture. For Profile 2, the number of stressed days decreased by 50 and 75% respectively, applying the simple or the advanced irrigation systems. The simple irrigation system was enough to remove all stressed days of Profile 3, the soil with coarsest texture.

Table 3. The number of days with soil water pressure head below –80 kPa in the top 50 cm of the soil in the period May 15 – October 15 under different irrigation schemes.

	Profile 1	Profile 2	Profile 3
no irrigation	153	143	53
simple	153	132	1
advanced	137	81	0

Table 4. Flux balance (FB; [mm], negative downwards), net leaching (NL; [mm]) and the number of days with leaching in the period May 15 – Oct 15 under different irrigation schemes.

	Profile 1			Profile 2			Profile 3		
	FB	NL	days	FB	NL	days	FB	NL	days
no irrigation	0.02	-13.51	69	76.51	-6.73	37	249.15	-33.42	13
simple	-0.04	-13.50	69	75.84	-6.94	39	127.93	-73.46	42
advanced	-0.03	-13.53	67	75.22	-7.05	39	77.23	-102.7	30

The number of days with the average pressure head in the top 50 cm below -80 kPa is summarized in Table 3. The trends are very similar to those in Table 2. One difference is, however that for Profile 2 the number of days with pressure head below –80 kPa in the top 50 cm is considerably higher than the number of days with simulated water deficit – in both cases when irrigation has been applied (132 vs. 57 and 81 vs. 29). This suggests, that this soil has a good capacity to store water in underlying (below 50 cms) slightly finer textured horizons, which then makes it available as excess water supply when upper horizons dry out.

Table 4 summarizes flux data that occurred at 1-meter depth. Water (solute) that flows below this depth can be potentially lost water (and soluble, such as chemicals and nutrients) that will end up in – and thus pollute - the subsoil and/or ground water. Only Profile 3 with its very coarse texture in the top 1 meter showed to be sensitive for the irrigation treatments. Applying the two irrigation schemes, the amount of water leached below 1 meter increased from 33 through 73 to 103 mm in the examined five months. This shows an increase in the risk of subsoil/groundwater pollution. The advanced irrigation scheme made an improvement in terms of reducing the number of days with leaching by almost 30% as compared to the simple "empirical" irrigation scheme for Profile 3. Thus, the increase in the amount of leached water (103 vs 73 mm) was simply due to the increased amount of irrigation water and not due to any lesser quality in the advanced irrigation scheme. In all examined ways, the simulated response of the other two soils remained virtually the same. None of the applied irrigation schemes had any negative influence on the leaching pattern of the other two soils.

We succeeded to give a quantitative estimate on the benefits achieved and risks posed by irrigation. A simulation-assisted "advanced" irrigation system may be more complicated, but may bring further improvement as compared to the simple "empirical" irrigation system. We showed what dilemmas large soil heterogeneity within the same soil management unit may cause: one action optimal for part of the unit may be far from optimal for other parts.

Study 2

Figure 4 shows the water retention curves – and the predicted saturated hydraulic conductivity value - of the "A" horizon, with changed BD and/or OM content. Decrease in BD resulted in a more than 6% increase in water content at saturation, which difference gradually decreases to zero by 10kPa. The increase in OM content resulted in slight but consistent changes: A half percent increase in water holding capacity at saturation gradually turns to a slight (0.2-0.3%) decrease in water holding capacity by drying to about 100kPa. Changes in the "B" horizon followed the same trends, and are not shown.

The expected amount of water deficit to the vegetation is summarized in Table 5. Decreasing BD caused – on average – a 23.2% reduction in plant water deficit, and a 3.3% drop in the evaporation deficit. The effect of the increase of OM content further induced water stress: plant water deficit increased by 36%, and soil evaporation deficit was increased by 8.4%. Changes in both directions – and in all cases - were slightly stronger in the wet simulation year.

The differences between wet and dry-year transpiration deficits are much lower than the differences in soil evaporation. While the crop can take up water from below the soil surface, evaporation deficit can be higher due to a very dry soil surface and low hydraulic conductivity. Most of this deficit is cumulated in the summer months (not shown). Similar increase/decrease trends can be seen in Table 6 for limiting soil water pressure heads. The redistribution of water is slowed by lowered hydraulic conductivity when OM content is increased, resulting in a higher number of days with drier soil. The opposite trend is evident from these data for the case when BD is decreased. Interestingly, there were about 10% more 'limiting' days in the wet year than in the dry year. We think, this was only caused by the better distribution of the precipitation in the dry year in our particular case.

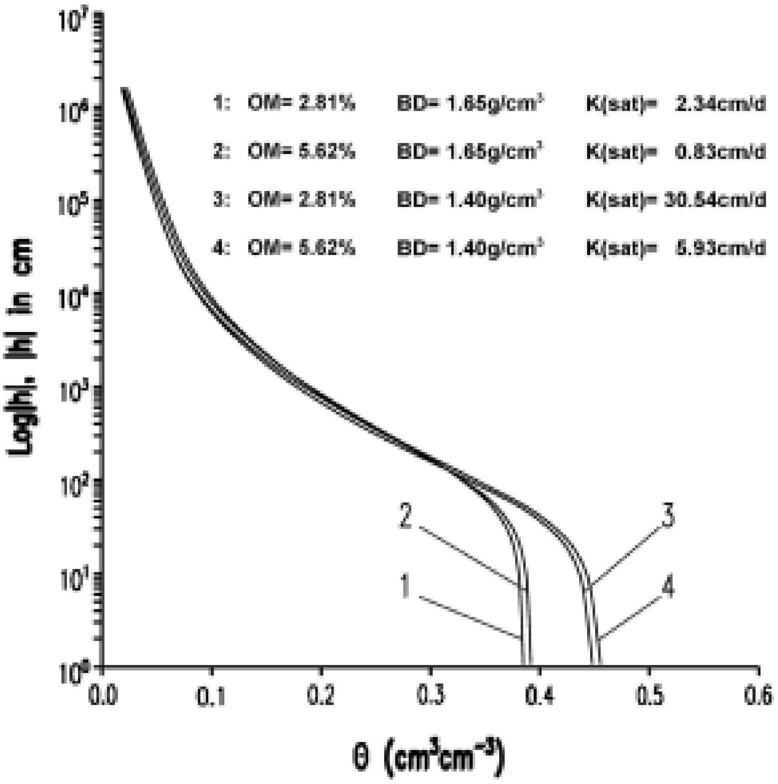


Figure 4. Water retention and saturated hydraulic conductivity of the 'A' horizon (top 15 cm) of the soil near Gödöllő before and after applying hypothetical changes to some physical soil properties.

Table 5. Evaporation and transpiration water deficits (in mm) during the year under the different hypothetical soil conditions.

	DRY YEAR		WET YEAR	
	transpir.	evapor.	transpir.	evapor.
ORIGINAL (1)	80.26	339.74	79.78	259.53
OM INCREASED (2)	102.06	362.61	112.48	287.67
BD DECREASED (3)	61.88	334.94	58.99	247.88
BOTH CHANGED (4)	86.00	356.30	80.85	271.99

Table 6. The number of days with soil water pressure head below –60 kPa in the top 50 cm of the soil throughout the year.

	DRY YEAR	WET YEAR
ORIGINAL (1)	102	113
OM INCREASED (2)	126	147
BD DECREASED (3)	84	87
BOTH CHANGED (4)	106	117

Table 7. Flux balance (FB; [mm], negative downwards), net leaching (NL; [mm]) and the number of days with leaching throughout the year, under the different hypothetical soil conditions.

	DRY YEAR			WET YEAR		
	FB	NL	days	FB	NL	days
ORIGINAL (1)	372.25	-43.13	64	363.67	-73.81	87
OM INCREASED (2)	332.38	-25.28	81	337.71	-56.07	113
BD DECREASED (3)	394.07	-48.39	64	386.93	-79.95	81
BOTH CHANGED (4)	349.13	-33.95	83	344.82	-66.91	105

Table 7 summarizes flux data that occurred at the depth of 1 meter. Throughout the year, upward flux was dominant, with at least 330 mm of water moving towards the soil surface. There was no significant difference between the sum of flux in the dry and wet year. The decreased BD resulted in a 5% increase in total upward flux throughout the year as compared with the original BD cases. The increased OM content however decreased the total upward flux by exactly 10%. Regarding net leaching, we found a 19% (or 7.8 mm) increase while BD was lowered, and a 28% (or 15.8 mm) decrease when OM content was doubled. Net leaching - on average – was almost twice as much (69.2 mm) in the wet year than in the dry year (37.7 mm). The distribution of the increased or decreased leaching changed in a different manner. Increased leaching caused by the decreased BD takes place in a somewhat lower number of days, indicating, that the flow intensity has increased. A possible explanation is that, despite better water-holding properties, increased soil water conductivity allowed higher flux rates both upwards and downwards. Increased OM content, however, increased the number of days with downward flux by 29%. Coupled with the decreased total amounts of flux and downward flux, it means, that the flow rate and intensity decreased dramatically, which can be a result of increased water retention and decreased hydraulic conductivity. It allows better storage for precipitation/irrigation water.

In summary, reduced BD (e.g. by plowing) increased flow rates, which may enhance the water supply capacity of soils, reduce water deficit, but in the meantime increase the risk of leaching into deeper soil horizons. Increased OM content had the opposite effect, with the benefit of reduced leaching – as a result of reduced flow rates -, but with the disadvantage of increased water deficit to the vegetation. When the combined effect of the above two factors is simulated, the two opposite trends basically cancel each other. Of course, changing a management system can result in much more complex changes as well (e.g. macroporosity and bypass flow) but those effects were not accounted for in this study, even though they can be well modeled by now.

Study 3

Twenty one-year long simulations were performed on each of the ten soils. Figure 5 shows the sum or average of the most relevant climatic factors per calendar week, over the 20 years.

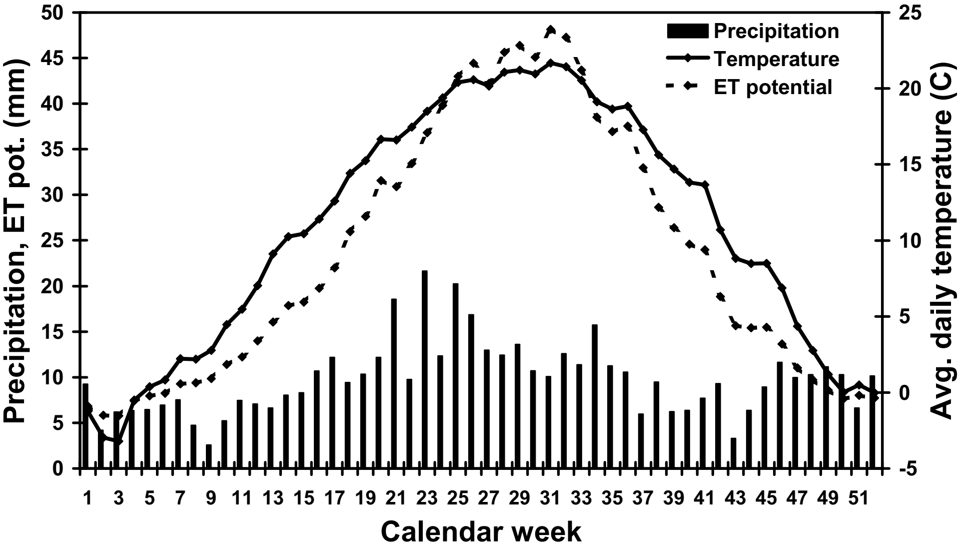


Figure 5. Averaged climatic characteristics used in the simulations of Study 3. Characteristics are averaged over the 20 years. Weekly sums (precipitation, ETpot.) or averages (temperature) are shown.

Flux output was generated with daily intervals. The amount of downward flux (later: Dflux) that occurred at the depth of 40 cm was averaged for each calendar day over the twenty years. We considered the 40 cms depth as the effective root-zone of an old grassland. The frequency at which Dflux occurred was also calculated. Results are shown in Figure 6. These ten soils can be grouped into three groups based on their Dflux pattern at 40 cm.

Group one consists of the two soils: Profiles 1 and 2 (c.f. Table 1). These are the heaviest textured soils throughout the profile (c.f. Figure 2). Dflux occurred only in certain periods of the year. For Profile 1, there was no Dflux after day 63. For Profile 2, the no-flux period was about 225 days. The maximum probability of Dflux was reached – for both soils – in the late winter period (January-February) at 10% and 15% respectively. This means, that Dflux occurred in two and three years out of the possible twenty respectively. The (daily) averaged amount of Dflux always remained below 2.5 mm/day (with the exception of a one time one day event for Profile 2). Basically, Dflux occurs only in distinct periods of the year, and at a low flow rate.

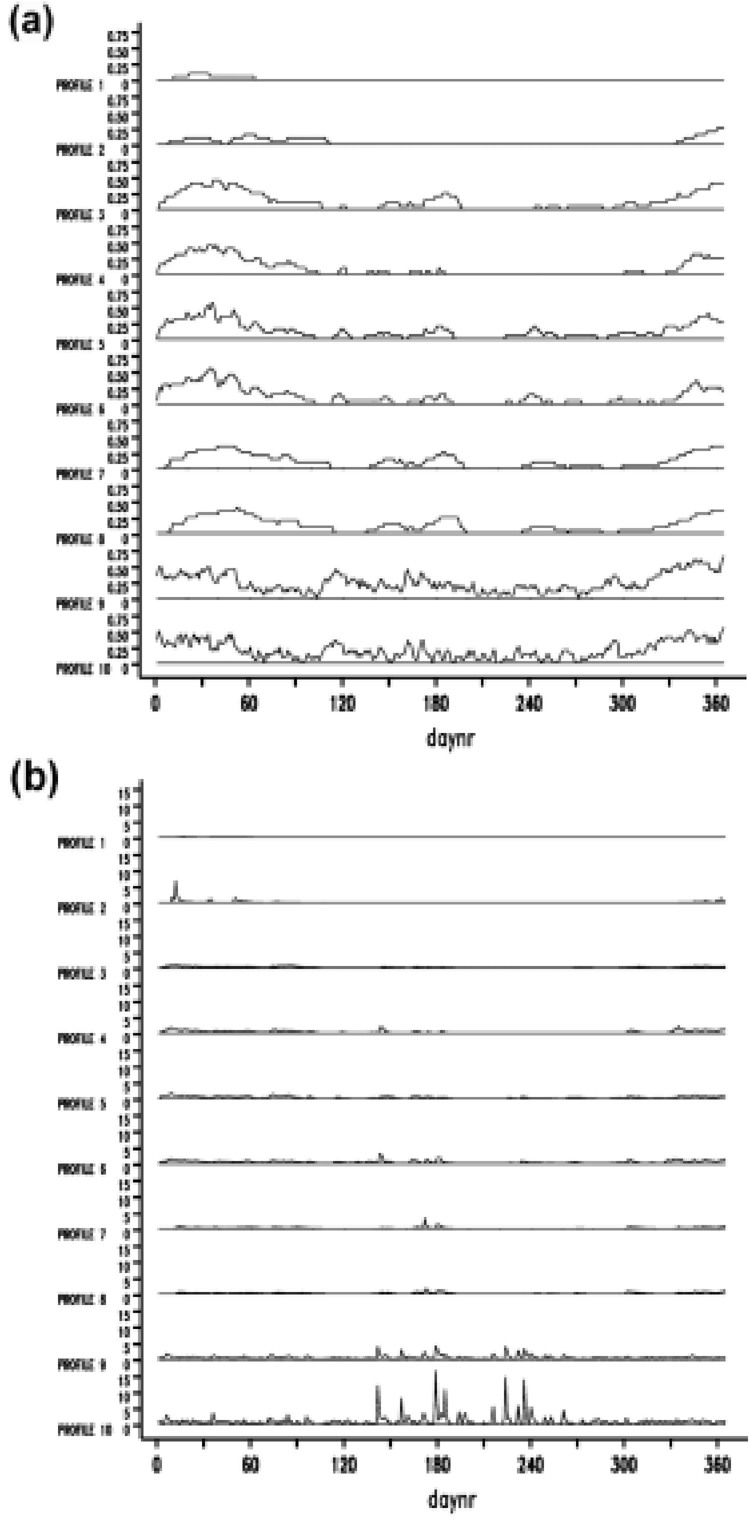


Figure 6. Frequency (a) and amount (b) (in mm) of downward flux in the ten soils at the depth of 40 cm. Frequency is expressed as ratio: the number of years - out of the possible 20 - with downward flux on a particular calendar day. Amounts are averaged based on the net downward flux that occurred in those 20 years.

Group two consists of six soils: Profiles 3 to 8. These are medium to medium-fine textured soils, with relatively small textural change in the profile, except for Profile 8, that gets considerably coarser with depth. Profiles 7 and 8 have finer texture in the upper 50 cms – with clay content between 20-25% - but these two soils have lower bulk density (BD<1.32) in the top 30-40 cms. We have to note, that – although these soils are not the only ones that have a plowed top horizon – this low bulk density may be due to recent plowing prior to sampling. Dflux in these soils occurred on most calendar days at least once out of the twenty years. There is a large (30 to 50%) probability that Dflux occurs in the December-February period. Outside of those three months, the probability of Dflux remained below 30%, but is expected to occur at least once out of twenty years on many of the days. The average daily amount of Dflux remained below 2.5-3 mm/day. However, larger fluxes occur in the summer period (May-June), when the probability of Dflux is lower. June is the wettest month, and the higher Dflux rates show the sensitivity of these soils to large amount of precipitation (or irrigation!).

Group three contains the remaining two soils: Profiles 9 and 10. These two soils from the Dunaharaszti area have coarse texture in the top 70 cm, with Profile 9 getting finer below. Dflux occurred with more than 25-30% probability many times throughout the entire year, often exceeding 50% in the winter periods. The shape of the probability curve is more sew-saw like than for the other eight soils, reflecting the sensitivity of these soils to day-to-day weather (precipitation) conditions and hinting – what is known in general about coarse textured soils – that they cannot store water well, but water rather drains through them at a high flow rate. This is also reflected in the amounts graph. Larger and quasi-immediate (averaged) Dflux follows each day with larger amount of (averaged) precipitation. The second precipitation peak (in August, c.f. Figure 5) clearly shows in the fluxes found for these soils. The amount of Dflux exceeds (an average of) 5 mm/day many times throughout the year.

Conclusions

Three studies were performed that used predicted soil hydraulic properties in *exploratory* simulation modeling. Studies of these kind can be used to explore the effect of changes being planned in soil and water management, under conditions where measured soil hydraulic characteristics are not available or measurements are not feasible. Expert knowledge can further be fine-tuned in the context of such studies, as it gives a quantitative dimension to the outcome of implemented changes. Of course, simplification is an essential part of the procedure and many factors are not accounted for. Still, the first two studies produced some clear messages, about “do’-s and don’t-s”.

Study 1 quantifies a major cause of the observed crop yield heterogeneity, and shows that while one part of the field is very sensitive for leaching, other parts are sensitive for drought. Besides quantifying different risk factors, such studies can also support the delineation of new management units within fields that best balance contradictory demands. In Study 2, we gave a simple example how hypothetical situations can be examined using these tools. The advantages, disadvantages and associated risks of different management plans and the extent of their effect could be evaluated. A study like Study 3 is essential in deciding, for instance, how contrary irrigation requirements, in terms of water supply and associated leaching, can be balanced.

Combining these kinds of studies could give an extended evaluation as to how different soils would react to different combinations of management practices and irrigation conditions. This picture can be further broadened by involving simultaneous yield estimation and estimation of cost-effectiveness, which make these tools powerful in planning a sustainable, productive and environmentally sound management system.

Acknowledgements

The authors thank NATO (CLG 975761) for financial support for travel. A. Nemes wishes to thank the Ministry of Agriculture, Nature Management and Fisheries for financial support which was provided through the International Agricultural Centre (IAC) of Wageningen, The Netherlands.

Chapter 7

General Conclusions

General Conclusions

1 The often used loglinear interpolation was shown to be a sub-optimal method to standardize soil particle-size distribution (PSD) data. Use of that procedure in the future is therefore not recommended. A novel procedure that is based on the similarity of PSD curves to each other performed best, along with the direct fit of a non-parametric spline function to particle-size data. The spline procedure has the advantage that it is simple and needs no external data, however, it is more sensitive to low data density on the PSD curve. The performance of the 'similarity procedure' depends heavily on the availability of a large, external reference data set. It was demonstrated that a set of 9607 soil horizons from the Netherlands had a sufficiently large variation in particle-size distributions to be successfully used to estimate PSD data for soils outside the Netherlands. The combination of these two methods was used to standardize particle-size data of soils in the HYPRES database.

2 International soil hydraulic databases can be valuable alternatives to smaller (e.g. national) databases when seeking solutions to soil water management related problems with international significance. UNSODA is a public domain database that holds unsaturated soil hydraulic data from around the world. The novel UNSODA Version 2.0 is compatible with most of the popular software and can be run on a personal computer. Currently lacking from the database are data for tropical soils. Uneven geographical distribution of the available data and the variability in measurement techniques used to obtain the data require caution from future users. Using data of UNSODA V2.0, the applicability of the 'similarity procedure' to interpolate soil PSD in a cross-continental manner between Europe and North America was studied. It was hypothesized that the procedure would not perform equally well between soils of these different geographical areas, however, the study rejected this hypothesis.

3. It was identified that a large amount of experimentally measured soil physical and hydro-physical data exist in Hungary, but that those are used and stored - in many cases - only through individual project reports. Recognizing the lack of a searchable, relational database, the HUNSODA database has been developed. The database is bilingual (English-Hungarian) and uses SQL (Standard Query Language) which makes it compatible to more of the international soil hydraulic databases of the day. It currently stores data of 840 soil horizons, of which 576 have soil hydraulic data. Class pedotransfer functions according to two different international texture standards (FAO and USDA) were developed and are provided in the form of parameters of the van Genuchten equation. If many smaller classes are distinguished (as in the USDA system), there is no significant difference between average curves of most classes. A less detailed classification system (e.g. FAO) yields more distinction between class functions.

4. Soil hydraulic pedotransfer functions (PTFs) to estimate soil water retention were developed from Hungarian and international data sets. The same methodology and the same sets of predictors were used to allow the source database to be the only variable that is changed. PTFs developed using a European and an intercontinental data set resulted in larger residuals than PTFs developed using only Hungarian data, while estimating water retention of an independent set of Hungarian soils. Larger residuals were found mostly for soils that were underrepresented in the data of PTF development. Results suggest that having a small set of relevant (local) data - when available - is better than using a large but more general data set.
   Functional evaluation of the various PTFs was performed by simulating time-series of soil water contents of seven Hungarian soils. Differences among national and international PTFs largely disappeared when the various estimated water retention curves (WRCs) were used in the simulation model. Using WRCs measured in the laboratory lead to root mean squared residuals that were only marginally (<0.01 m^3m^{-3}) better on an annual basis than those using PTF estimates, which is an argument for PTFs. For the presented case study, differences between estimations using different scale data sets (or measured WRCs) were not the main source of error in the simulation model, errors introduced by different PTFs were overwhelmed by errors resulting from other factors.

5. Scenario studies can be used in planning and prevention to fine-tune expert knowledge. The role of PTFs in such studies is that they allow the examination of potential risks and benefits under conditions where measured soil hydraulic characteristics are not available (e.g. when modeling the effect of hypothetical changes) or measurements are not feasible (e.g. for large areas of land). Three scenario studies were performed using exploratory modeling to demonstrate how PTFs may help improve land management planning. Results of the different studies are:

   (i) A quantitative estimate could be given of the benefits achieved and risks posed by different irrigation plans. Problems caused by soil heterogeneity within the same soil management unit were described, such as: while one part of a field is very sensitive for leaching, other parts are sensitive for drought. Such studies can support the delineation of new management units within fields that best balance contradictory demands.

   (ii) Imaginary changes in land management, such as decreasing bulk density (e.g. plowing) and increasing soil organic matter content, have opposite effects on soil behavior, in terms of flow rates, plant water deficits, and risks of water (and nutrient) leaching into deeper soil horizons.

   (iii) Using 20 years of simulation, ten Hungarian soils could be classified into three groups, based on the daily amounts and annual pattern of downward water flux at 40 cm depth.

   Using an exploratory research approach, quantitative answers were provided to "what if" type of questions, allowing the distinction of trends in potential problems and successes.

Chapter 8

References

References

Arany, S. 1943. Vezérfonal a mezőgazdasági főiskolai hallgatók laboratóriumi gyakorlatához (Laboratory practicals for agricultural college students: Guidelines). Debrecen, Hungary. (in Hungarian)

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

Batjes, N.H. 1996. Development of a world data set of soil water retention properties using pedotransfer rules. Geoderma 71: 31-52.

Batjes, N.H. and V.W.P. van Engelen (Editors). 1997. Guidelines for the compilation of a 1:2.5 million SOTER database (SOVEUR Project). Report 97/06, International Soil Reference and Information Centre, Wageningen, The Netherlands.

Bouma, J. 1989. Using soil survey data for quantitative land evaluation. Adv. Soil. Sci. 9: 177-213.

Bouma, J. and J.A.J. van Lanen. 1987. Transfer functions and threshold values: From soil characteristics to land qualities. In: K.J. Beek, P.A. Burrough and D.E. McCormack (Editors): Proceedings of the international workshop on quantified land evaluation procedures, Washington, D.C., USA, 15-18 May 1986, pp. 106-110. ITC Publication 6. Enschede, The Netherlands.

Bouma, J., P. Droogers and P. Peters. 1999. Defining the "ideal" soil structure in surface soil of a Typic Fluvaquent in the Netherlands. Soil Sci. Soc. Am. J. 63: 343-348.

Buchan, G.D. 1989a. Applicability of the simple lognormal model to particle-size distribution in soils. Soil Sci. 147(3): 155-161.

Buchan, G.D. 1989b. Comments on 'A unifying quantitative analysis of soil texture: Improvement of precision and extension of scale'. Soil Sci. Soc. Am. J. 53: 594.

Buchan, G.D., K.S. Grewal and A.B. Robson. 1993. Improved models of particle-size distribution: An illustration of model comparison techniques. Soil Sci. Soc. Am. J. 57: 901-908.

Búzás, I. (Editor). 1993. Talaj- és agrokémiai vizsgálati módszerkönyv 1.: Fizikai, vízgazdálkodási és ásványtani talajvizsgálati módszerek. (Methodology for soil- and agrochemical analysis I: Methods for soil physical, soil water management and mineralogy analysis). 37-41. INDA 4231 Publisher. Budapest, Hungary. (in Hungarian)

Campbell, G.S. 1985. Soil physics with BASIC. Developments in Soil Science, 14. Elsevier, Amsterdam.

Commission of the European Communities (CEC). 1985. Soil map of the European Communities 1:1.000.000. CEC-DGVI, Luxembourg.

Cornelis, W.M., J. Ronsyn, M. van Meirvenne and R. Hartmann. 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. Soil Sci. Soc. Am. J. 65(3): 638-648.

Czinege, E. 2000. A termőhely-specifikus műtrágyázási szaktanácsadás talajtani alapjainak bővítése. (Extension of soil information and knowledge of site-specific nutrient management consulting). PhD Thesis. University of Gödöllő, Gödöllő, Hungary. (in Hungarian)

Demuth, H. and M. Beale. 1992. Neural Network Toolbox Manual. MathWorks Inc. Natick, MA., USA.

Droogers, P. and J. Bouma. 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Sci. Soc. Am. J. 61: 1704-1710.

Efron, B. and R.J. Tibshirani. 1993. An introduction to the bootstrap. Monographs on statistics and applied probability. Chapman and Hall, New York, USA.

Eijsackers, H.J.P. and T. Hamers (Editors). 1993. Integrated soil and sediment research: a basis for proper protection. Selected Proceedings of the First European Conference On Integrated Research for Soil and Sediment Protection and Remediation (EUROSOL), Maastricht, The Netherlands, 6-12 September 1992.

Espino, A., D. Mallants, M. Vanclooster and J. Feyen. 1996. Cautionary notes on the use of pedotransfer functions for estimating soil hydraulic properties. Agricultural Water Management 29: 235-253.

FAO. 1990. Guidelines for soil descriptions. (3rd ed.). FAO/ISRIC, Rome, Italy.

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.

FAO. 1998. World reference base for soil resources. FAO World Soil Resour. Rep. 84. Rome, Italy.

Farkas, Cs., Cs. Gyuricza and P. László. 1999. Egyes talajfizikai tulajdonságok vizsgálata talajművelési tartamkísérletekben gödöllői barna erdőtalajon. (Studies on certain physical soil properties in long-term soil cultivation experiments on a brown forest soil in Gödöllő). Növénytermelés 48(3): 323-336. (in Hungarian)

Farkas, Cs., Cs. Gyuricza, P. László and M. Birkás. 2000. Study of the influence of soil tillage on soil water regime. in: R. Horn, J.J.H. van den Akker and J. Arvidsson (Editors): Subsoil compaction. Advances in Geoecology 32., Catena Verlag, pp. 251-257.

Filep, Gy. and G. Ferencz. 1999. Javaslat a magyarországi talajok szemcseösszetétel szerinti osztályozásának pontosítására. (Recommendation for improving the accuracy of soil classification on the basis of particle composition). Agrokémia és Talajtan 48. 305-320. (in Hungarian)

Finke, P.A. 1995. Bodeminventarisatie: van gegevensverzameling tot informatie-verstrekking. (Soil inventarization: from data collection to providing information) In: P. Buurman and J. Sevink (Editors): Van bodemkaart tot informatiesysteem. Wageningen Press, Wageningen, The Netherlands, pp. 111-125. (in Dutch).

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

Hack-ten Broeke, M.J.D. and J.H.B.M. Hegmans. 1996. Use of soil physical characteristics from laboratory measurements or standard series for modelling unsaturated water flow. Agricultural Water Management 29: 201-213.

Hastie, T.J. and R.J. Tibshirani. 1990. Generalised additive models. Chapman and Hall, London, UK.

Haykin, S. 1999. Neural Networks: a comprehensive foundation. 2nd ed. Prentice Hall, London, UK.

Hecht-Nielsen, R. 1990. Neurocomputing. Addison-Wesley, Reading, MA, USA.

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

Imam, B., S. Sorooshian, T. Mayr, M.G. Schaap, J.H.M. Wösten and R.J. Scholes. 1999. Comparison of pedotransfer functions to compute water holding capacity using the van Genuchten model in inorganic soils – Report to IGBP-DIS Soil Data Tasks. IGBP-DIS WORKING PAPER #22. IGBP-DIS, Toulouse, Cédex, France.

International Society of Soil Science. 1929. Minutes of the first Commission Meetings, International Congress of Soil Science, Washington, 1927. Proceedings of the Congress of the International Society of Soil Science 4: 215-220.

Jamagne, M., D. King, C. Le Bas, J. Daroussin, A. Burrill and P. Vossen. 1994. Creation and use of a European Soil Geographic Database. 15th International Congress of Soil Science, Transactions, Commission V, Symposia, Acapulco, Mexico, Vol. 6a, pp. 728-742.

Johnson N.L. and S. Kotz. 1970. Distributions in statistics. Continuous Univariate Distributions 2(3), Wiley, New York.

Kabat, P. and M.J.D. Hack-ten Broeke. 1989. Input data for agrohydrological simulation models: some parameter estimation techniques. Report No. 9., The Winand Staring Centre, Wageningen, The Netherlands. 18p.

Katschinski, N.A. 1956. Die mechanische Bodenanalyse und die Klassifikation der Böden nach ihrer mechanischen Zusammensetzung. Rapports au Sixiéme Congrés de la Science du Sol. Paris, B, pp. 321-327.

Kern, J.S. 1995. Evaluation of soil water retention models based on basic soil physical properties. Soil Sci. Soc. Am. J. 59: 1134-1141.

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

Láng, I. 1994. The ecological foundation of land use in Hungary. Agrokémia és Talajtan 43(3-4): 264-268.

Leij, F.J., W.J. Alves, M.Th. van Genuchten and J.R. Williams. 1996. Unsaturated Soil Hydraulic Database, UNSODA 1.0 User's Manual. Report EPA/600/R-96/095, U.S. Environmental Protection Agency, Ada, OK, USA, 103 p.

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

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

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

Minasny, B. and A.B. McBratney. 2002. The Neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc.Am. J. 66: 352-361.

Moran, C.J., A.B. McBratney, A.J. Koppi. 1988. A micromorphometric method for estimating change in the volume of soil induced by weathering. Journal of Soil Science 39: 357-373.

Mualem, Y. 1976. A new model predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12: 513–522.

Munsell Color Company. 1975. Munsell soil colour charts. Munsell Color. Kollmorgan, Maryland, USA.

National Plan for Water Management (Országos Vízgazdálkodási Keretterv). 1964. OVH, Budapest, Hungary (in Hungarian).

Nelder, J.A. and R. Mead. 1965. A simplex method for function minimization. Comput. J. 7: 308-313.

Nemes, A. 2002. Unsaturated soil hydraulic database of Hungary: HUNSODA. Agrokémia és Talajtan 51(1-2): 17-26.

Nemes, A. 2003. Magyarországi talajok telítettségi vízvezetőképességének becslése egyszerűbben mérhető talajfizikai tulajdonságokból. (Pedotransfer function to estimate saturated hydraulic conductivity for Hungarian soils). Submitted to Agrokémia és Talajtan. (in Hungarian)

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

Nemes, A., M.G. Schaap, F.J. Leij and J.H.M. Wösten. 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. J. Hydrol. 251(3-4): 151-162.

Nemes, A., M.G. Schaap and J.H.M. Wosten. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. (in press)

Nováky, B., A. Bussay and P. Domonkos. 1996. Éghajlati változások hatása az öntözővízigényre (Effects of climatic changes on the irrigation water demand of plants). Climatological and Agrometeorological Papers 5., OMSZ (Hungarian Meteorological Service), Budapest, Hungary. (in Hungarian).

Pachepsky, Ya.A., D. Timlin and G. Várallyay. 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60: 727-773.

Pachepsky, Ya.A., W.J. Rawls and D.J. Timlin. 1999. The current status of pedotransfer functions: their accuracy, reliability and utility in field- and regional-scale modelling. In: Corwin, D.L., K. Loague and T.R. Ellsworth (Editors): Assessment of non-point source pollution in the vadose zone. Geophysical Monograph 108. pp. 223-234. American Geophysical Union, Washington, D.C., USA.

Pachepsky, Ya.A., W.J. Rawls and D. Gimenez. 2001. Comparison of soil water retention at field and laboratory scales. Soil Sci. Soc. Am. J. 65: 460-462.

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

Rajkai, K. 1984. A talaj kapilláris vezetőképességének számítása a pF-görbe alapján. (A method for calculating the capillary conductivity of the soil on the basis of the water retention (pF) curve). Agrokémia és Talajtan 33: 50-62. (in Hungarian)

Rajkai, K., Gy. Várallyay, Ya.A. Pachepsky and R.A. Scserbakov. 1981. pF-görbék számítása a talaj mechanikai összetétele és térfogattömege alapján. (Calculation of water retention data from the texture and the bulk density of soils). Agrokémia és Talajtan 30: 409-438. (in Hungarian)

Rajkai, K. and S. Kabos. 1999. A talaj víztartóképesség-függvény (pF-görbe) talajtulajdonségok alapján történő becslésének továbbfejlesztése. (Estimation of soil water retention characteristics (pF curves) from other soil properties). Agrokémia és Talajtan 48: 15-32. (in Hungarian)

Rawls, W.J., T.J. Gish and D.L. Brakensiek. 1991. Estimating soil water retention from soil physical properties and characteristics. Adv. Soil Sci. 16: 213-234.

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

Rousseva, S.S. 1987. Studies on the soil erodibility of calcaric chernozems and cinnamonic forest soils. PhD Thesis, N. Poushkarov Institute of Soil Science and Agrogeology, Academy of Agricultural Sciences, Sofia, Bulgaria. (in Bulgarian).

Rousseva, S.S. 1997. Data transformations between soil texture schemes. Eur. J. Soil Sci. 48: 749-758.

Schaap, M.G. and F.J. Leij. 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163(10): 765-779.

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

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

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(3-4): 163-176.

Semenov, M.A. and E.M. Barrow. 1997. Use of a stochastic weather generator in the development of climate change scenarios. Climatic Change 35: 397-414.

Shirazi, M.A. and L. Boersma. 1984. A unifying quantitative analysis of soil texture. Soil. Sci. Soc. Am. J. 48: 142-147.

Shirazi, M.A., L. Boersma and J.W. Hart. 1988. A unifying quantitative analysis of soil texture: improvement of precision and extension of scale. Soil Sci. Soc. Am. J. 52: 181-190.

Soet, M. and J.N.M. Stricker. 2002. Functional behaviour of pedotransfer functions in soil water flow simulation. Hydrological Processes. (in press)

Soil Survey Staff. 1975. Soil Taxonomy: A basic system of soil classification for making and interpreting soil surveys. USDA/SCS Agricultural Handbook No. 436, U.S. Government Printing Office, Washington, D.C., USA.

Stefanovits, P. 1963. Magyarország talajai (Soils of Hungary). Akadémiai Kiadó. Budapest, Hungary. (in Hungarian)

Szabolcs, I. (Editor). 1966. A genetikus üzemi talajtérképezés módszerkönyve. (Methodology for genetic soil mapping at farm scale). OMMI Genetikus Talajtérképek. Ser. 1. No. 9., Budapest, Hungary. (in Hungarian)

Szabolcs, I. 1974. Salt-affected soils in Europe. Martinus Nijhoff, The Hague, The Netherlands.

Szabolcs, I., K. Darab and Gy. Várallyay. 1969. Methods of predicting salinization and alkalization processes due to irrigation on the Hungarian Plain. Agrokémia és Talajtan 18(Suppl.): 351-376.

Szalai, Gy. (Editor). 1989. Az öntözés gyakorlati kézikönyve (Handbook of irrigation) Mezőgazdasági Kiadó, Budapest, Hungary. (in Hungarian)

Tamari, S., J.H.M. Wösten and J.C. Ruiz-Suárez. 1996. Testing an artificial neural network for predicting soil hydraulic conductivity. Soil Sci. Soc. Am. J. 60: 771-774.

Tietje, O. and M. Tapkenhinrichs. 1993. Evaluation of pedotransfer functions. Soil Sci. Soc. Am. J. 57: 1088-1095.

Tietje, O. and V. Hennings. 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69: 71-84.

Tiktak, A., F. van den Berg, J.J.T.I. Boesten, M. Leistra, A.M.A. van der Linden and D. van Kraalingen. 2000. Pesticide Emission Assessment at Regional and Local Scales: User Manual of FOCUS Pearl version 1.1.1. RIVM Report 711401008, Alterra Report 28, RIVM, Bilthoven, The Netherlands. 142 pp.

Tóth, T. and L. Blaskó. 1998. Secondary salinization caused by irrigation. In: A. Rodríguez Rodríguez, C.C. Jiménez Mendoza and M.I. Tejedor Salguero (Editors): The soil as a strategic resource: degradation processes and conservation measures. Geoforma Ediciones, Logroño. pp. 229-253.

Tóth T. and Gy. Várallyay. 2001. Egy mintaterület talajának variabilitása a sófelhalmozódás tényezői szerint. (Variability in the soil of a sample area according to salt accumulation factors). Agrokémia és Talajtan 50: 19-34. (in Hungarian)

Tóth, T., L. Kuti, L. Kabos and L. Pásztor. 2001. Use of digitalized hydrogeological maps for evaluation of salt-affected soils of large areas. Arid Land Research and Management 15: 329-346.

Tóth T. and L. Kuti. 2002. Numerical simulation versus repeated field instrumental measurements: a case study of monitoring salinity status in a native sodic grassland with shallow groundwater. Agrokémia és Talajtan 51: 243-252.

USDA (United States Department of Agriculture). 1951. Soil Survey Manual. U.S. Dept. Agriculture Handbook No. 18. Washington, D.C., USA.

van Alphen, B.J., H.W.G. Booltink and J. Bouma. 2001. Combining pedotransfer functions with physical measurements to improve the estimation of soil hydraulic properties. Geoderma 103: 133-147.

van Dam, O., J.W.M. van der Drift and C.A. Diepen. 1994. Estimation of available soil moisture capacity for the soil units of the EC soil map. Technical Document 20. DLO Winand Staring Centre, Wageningen, The Netherlands. 88p.

van Dam, J.C., J. Huygen, J.G. Wesseling, R.A. Feddes, P. Kabat, P.E.V. van Walsum, P. Groenendijk and C.A. van Diepen. 1997. SWAP version 2.0, Theory. Simulation of water flow, solute transport and plant growth in the Soil-Water-Atmosphere-Plant environment. Technical Document 45, DLO Winand Staring Centre, Report 71, Department Water Resources, Agricultural University, Wageningen.

van Genuchten, M.Th. 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci. Soc. Am. J. 44: 892-898.

Vanoni, V.A. 1980. Sedimentation Engineering. Sedimentation Committee, Hydraulic Division, American Society of Civil Engineers, New York.

Várallyay, Gy. 1978. A talaj fizikai és vízgazdálkodási tulajdonságainak vizsgálata. (Analysis of soil physical and hydro-physical properties). MÉM Standard. (MSZ-08, 0205-78). Budapest, Hungary. (in Hungarian)

Várallyay, Gy. 1987. A talaj vízgazdálkodása. (Soil water management). MTA Doktori Értekezés. Budapest, Hungary. (in Hungarian)

Várallyay, Gy. 1989a. Soil water problems in Hungary. Agrokémia és Talajtan 38(3-4): 577-595.

Várallyay, Gy. 1989b. Mapping of hydrophysical properties and moisture regime of soils. Agrokémia és Talajtan 38(3-4): 800-817.

Várallyay, Gy. 1994. Soil management and environmental relationships in Central and Eastern Europe. Agrokémia és Talajtan 43(1-2): 41-66.

Várallyay, Gy., K. Rajkai, Ya.A. Pachepsky and E.V. Mironenko. 1979. A pF-görbék matematikai leírása. (Mathematical description of water retention curves). Agrokémia és Talajtan 28: 15-38. (in Hungarian)

Várallyay, Gy., L. Szűcs, K. Rajkai, P. Zilahi and A. Murányi. 1980. Magyarországi talajok vízgazdálkodási tulajdonságainak kategóriarendszere és 1:100 000 méretarányú térképe. (Soil water management categories of Hungarian soils and the map of soil water properties (1:100 000)). Agrokémia és Talajtan 29: 77-112. (in Hungarian)

Wagner, B., V.R. Tarnawski, V. Hennings, U. Müller, G. Wessolek and R. Plagge. 2001. Evaluation of pedo-transfer functions for unsaturated soil hydraulic conductivity using an independent data set. Geoderma 102: 275-297.

Walker, P.H. and D.J. Chittleborough. 1986. Development of particle-size distributions in some Alfisols of southeastern Australia. Soil. Sci. Soc. Am. J. 50: 394-400.

Wösten, J.H.M., P.A. Finke and M.J.W. Jansen. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66: 227-237.

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

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

Wösten, J.H.M., G.J. Veerman, W.J.M. de Groot and J. Stolte. 2001b. Waterretentie- en doorlatendheidskarakteristieken van boven- en ondergronden in Nederland: de Staringreeks. (Water retention and hydraulic conductivity characteristics of top- and subsoils of the Netherlands: the Staring-series). ALTERRA Report Nr. 153. ALTERRA. Wageningen, The Netherlands. (in Dutch)

Yaalon, D.H. 1989. Comments on 'A unifying quantitative analysis of soil texture'. Soil Sci. Soc. Am. J. 53: 595.

Summary

Rational soil management and land use are important elements of sustainable agricultural and environmental management. One of the main elements in the sustainability approach is proper knowledge and management of the soil water and nutrient balance. Different agricultural management practices have been developed to optimise aeration and water and nutrient supply capacity of soils and to minimise their pollution. Movement of chemicals and nutrients is basically directed by soil-water flow in the form of solutes. There are an increasing number and variety of numerical computer models being developed throughout the world that simulate soil water and solute transport and plant growth. However, use of these models has often been limited by lack of accurate input parameters. Among the key input parameters for these models are the soil hydraulic characteristics, such as water retention and hydraulic conductivity. Several methods are available to obtain measured characteristics under field or laboratory conditions. However, the fact remains that the hydraulic parameters are still difficult to measure accurately, especially for undisturbed field soils. For many applications, estimation of soil hydraulic characteristics with pedotransfer functions may offer an alternative to costly and troublesome measurements.

This thesis contributes to hydrological and environmental research by focusing on improving the capability to support water and solute transport models with soil hydraulic input data. These data can be estimated from different data sources. The predictive capability and reliability of different estimates is studied, with special focus on Hungary as test country.

Environmental and agricultural management and pollution problems are identified worldwide. Many of such problems are not restricted to national boundaries and therefore require international co-operation if solutions are to be found. Often, these solutions require the ability to use soil data as input in simulation models, however, despite a number of recognised international standards, soil data are rarely compatible across national frontiers. This problem is encountered when compiling international databases. Only few of the contributing institutions/countries adhere strictly to a recognised international system such as e.g. FAO. Therefore, interpolation of data is needed. Particle-size distribution is among the key inputs to soil hydraulic pedotransfer functions. It is also a soil property for which national standards largely differ.

The performance of four different interpolation procedures was evaluated to achieve compatibility of cumulative soil particle-size data that were collected according to the standards of different countries. The accuracy of the different procedures was found to vary with size intervals between measured points of the particle-size distribution. The loglinear interpolation has previously been used in various studies but was found to give the least accurate estimation of the four procedures. Fitting the Gompertz curve, which is a special non-symmetrical type of curve described by a closed-form equation, showed less sensitivity to size intervals between measured points. However, interpolation within some

of the particle-size distributions was not sufficiently accurate and this procedure could not be applied to particle-size distributions where the number of measured size fractions was less than the number of model parameters. Fitting a non-parametric spline function to the particle-size distributions showed a considerable increase in accuracy of the interpolation with decreasing size intervals between measured points. As a novel approach, the 'similarity procedure' was introduced which does not use any mathematical interpolation functions. It uses an external source of soil information from which soils are selected with particle-size distributions that match the distribution of the soil under investigation. This similarity procedure was capable of giving the most accurate interpolations. Once an extensive external reference data set with well-quantified particle-size distributions is available, the similarity procedure becomes a very powerful tool for interpolations. Based on the number and distribution of measured points on the particle-size distribution curve, a general rule was formulated as when to fit a spline function or use the novel similarity procedure to estimate missing values **(Chapter 2)**.

A prerequisite for the development of PTFs is the availability of a source database that contains potential predictors as well as hydraulic characteristics. Many electronic databases with soil information exist already, however, the emphasis of most of the databases is on soil taxonomy and they often have limited unsaturated soil hydraulic data. The potential use of international soil hydraulic databases is twofold. First, they may serve as valuable alternatives to smaller (e.g. national) databases when answering research questions that are related to soil water management, especially those with international significance. They may provide a common language to scientists and other potential users in different countries/regions. Second, many countries or regions in the World do not have the resources to collect a sufficient amount of soil hydraulic data. In such cases, external data sources are the only option.

Focusing on soil hydraulic data, the international UNsaturated SOil hydraulic DAtabase (UNSODA) was developed to provide 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 for some 790 soil samples from around the world. Following a first MS-DOS version of the database, the second version (UNSODA V2.0) was released in 1999. The format and structure of the new database have been redesigned to provide additional 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 to show graphical output. The structure and contents of the database are reviewed, as well as the operations that can be performed on the different data types in UNSODA V2.0.

Two applications are shown that use data of UNDOSA V2.0. The retrieval of data is briefly illustrated. It is 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 'similarity procedure', introduced in Chapter 1, was tested for it's validity for soils of different geographical areas. While applying the procedure, a large European data set was used to help interpolate particle-size distribution of non-European soils. It was hypothesised that the procedure would not perform equally well for soils of other geographical areas, however, the study rejected this hypothesis **(Chapter 3)**.

Soils are particularly important natural resources and contribute significantly to the national economy in Hungary. Their rational utilisation and conservation as well as the maintenance of their functionality have particular significance in the economy and in environment protection, as almost 70% of the area of the country can potentially be used for agriculture. Many studies have been conducted in Hungary to answer questions of local soil hydrological phenomena. It was recognised that many soil hydraulic data have been collected in the framework of various independent projects. However, only a minor part of those were stored in a common database. Available data were fragmented and could only be gathered from various different sources. Therefore those were not easily accessible for present/future researchers and were only suitable for a limited range of applications.

The newly developed Hungarian UNsaturated SOil DAtabase (HUNSODA) is introduced in **Chapter 4**. Focusing on soil hydraulic data, the database also serves as storage for a range of other soil physical data and basic soil information on 840 soil horizons, of which 576 soil horizons have soil water retention data. The database was developed in a format that uses SQL (Standard Query Language) for communication which makes it compatible with many other databases. The structure and contents of the database are discussed, and a simple description is given for basic operations.

Data were then used to develop class pedotransfer functions for Hungarian soils according to both the FAO and USDA soil texture classes. Some differences between the Medium, Medium-Fine and Fine classes of the FAO classification could be seen, however, considering their standard deviations there is considerable overlap among these classes. In case the USDA classification was considered, apart from the coarsest and finest textured classes, all other class average curves run very close to each other, with some of them crossing over. It suggests that there is no need for many small classes when class PTFs are derived. Rather a very distinctive but not so detailed system should be used.

The relevance of any type of information derived from such databases has to be tested. Many PTFs have been developed in recent decades. Independent databases have been used to evaluate various PTFs that were developed elsewhere. There are two major limitations of most studies that evaluate PTFs. One is that it remains unclear what the main sources of the estimation errors are. In those studies it is not clear whether differences between data sets used to derive PTFs (size, origin, reliability), differences between the algorithm of PTF development (e.g. different regression types vs. neural network models) or differences among the predictors cause a particular PTF to perform better than others. The other limitation is that in most cases, testing of PTFs is restricted to testing the estimation of water retention parameters or points. However, water retention and hydraulic conductivity are not the final aim but are intermediate characteristics needed to calculate other soil properties with more practical meaning. Functional evaluation of estimated soil hydraulic data helps to characterize the contribution of such data to the inaccuracy and uncertainty of simulations.

Data from Hungary were used to derive “National scale” soil hydraulic PTFs. A data set selected from the HYPRES database was used to develop “Continental scale” PTFs. Finally, a data set containing mostly American and European data was used to develop “Intercontinental scale” PTFs. All three data sets – plus two more sets combined from the above data sets - were used to develop 11 different PTFs using artificial neural networks. The same algorithm and the same predictors were used for each data set to test the performance of each PTF. Models were built up gradually – from simple to more complex – using the most commonly used predictors. All PTFs were tested using a second, independent Hungarian data set. The performance of international and national PTFs were first tested on sum of square residuals between measured and estimated retention characteristics. Next, we used different PTF estimates to simulate soil moisture time series of seven Hungarian soils. Sum of square residuals was evaluated for simulated water contents and water contents observed in the field.

Estimations showed an improving trend as the list of input variables increased. Estimation of the water retention curves resulted in root mean squared residuals (RMSR) - using different sets of input parameters - ranging from 0.02 to 0.06 m^3m^{-3} for national scale PTFs, while international scale PTFs had RMSRs from 0.025 to 0.088 m^3m^{-3}. Larger residuals were found mostly for soils that were underrepresented in the data of PTF development. Results suggest that having a small set of relevant data – when available - is better than using a large but more general data set.

Estimated water retention curves were then used to simulate soil moisture time series of seven different Hungarian soils. Small differences were found among the PTFs derived from different scale data collections. Root mean squared residuals over a growing season ranged from 0.065 to 0.07 m^3m^{-3} while simulations using laboratory measured water retention curves had an error of 0.061 m^3m^{-3} in similar simulations. The differences among national and international PTFs largely disappeared when the estimated water retention curves were used for simulations of soil water contents. Differences between estimations using different scale data sets (or measured water retention curves) were not the main source of error in the simulation model **(Chapter 5)**.

Scenario studies are important in planning and prevention. Expert knowledge can further be fine-tuned in the context of such studies, as it gives a quantitative dimension to the outcome of implemented changes. Field experimentation representing different management possibilities would be time consuming, costly and sometimes even risky. Pedotransfer functions can provide important soil physical data at relatively low costs. Few studies, however, explore which contribution PTFs can make to exploratory planning, by examining the expected outcome of certain changes in soil and water management practices.

Three scenario studies are described to demonstrate how PTFs may help improve land management practices. The three examples were derived from what seem to be realistic problems facing Hungary as well as many other countries. Exploratory modeling was used to:

i. Evaluate benefits and risks when irrigating a field, including the impact of soil heterogeneity;

ii. Examine which changes can be expected – in terms of soil water balance and supply - if bulk density and/or organic matter content are changed as a result of alternative management systems; and

iii. Evaluate the risk of leaching to deeper horizons in some Hungarian soils.

Results of the different studies are:

i. It is possible to make a quantitative estimate of the benefits achieved and risks posed by different irrigation plans. Problems caused by soil heterogeneity within the same soil management unit are described. As an example, one part of a heterogeneous field is sensitive for leaching while other parts are sensitive for drought.

ii. Imaginary changes in land management, such as decreasing bulk density (e.g. plowing) and increasing soil organic matter content, have opposite effects on soil behavior, in terms of flow rates, plant water deficits, and risks of water (and nutrient) leaching into deeper soil horizons.

iii. Using 20 years of simulation, the ten Hungarian soils could be classified into three separate groups, based on daily amounts and annual pattern of downward water flux at 40 cm depth.

Using an exploratory research approach, quantitative answers are provided to "what if" type of questions. This approach allows distinction of trends in potential problems and successes, which may contribute to the development of sustainable management systems **(Chapter 6)**.

Apart from the outlined research, the candidate participated as a research student in the completion of the research project entitled "Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning" supported by the European Union (project no. CHRX-CT94-0639). This project resulted in the compilation of the European scale HYPRES (Hydraulic Properties of European Soils) database. The complete results of this project are not presented. The candidate co-authored a research paper (Wösten et al., 1999) and several conference papers – based on the above publication - which presented results of this research. Part of this research was presented in Chapter 2 and the database was later used in the work presented in Chapter 5.

Samenvatting

Verstandig bodem- en landgebruik zijn belangrijke elementen in concepten van duurzame landbouw en milieubeheer. De zorg betreffende het verstandig gebruik van onze bodems richt zich op een goed inzicht in en beheer van haar water- en nutriëntenbalans. Er zijn verschillende landbouwsystemen ontwikkeld om de aeratie en de water- en nutriëntenbeschikbaarheid van de bodem zo optimaal mogelijk te maken en de verontreiniging te beperken. Hierbij vindt transport van chemicaliën en nutriënten vooral plaats in de vorm van transport van opgeloste stoffen. Wereldwijd worden er steeds meer computermodellen ontwikkeld die water- en stoffentransport en plantengroei simuleren. Het gebruik van deze modellen wordt echter vaak beperkt door gebrek aan nauwkeurige invoergegevens. Cruciale invoergegevens voor deze modellen zijn onder meer de bodemfysische karakteristieken zoals waterretentie en doorlatendheid. Er zijn verschillende methoden beschikbaar om deze karakteristieken te meten onder veld- en laboratoriumomstandigheden, maar het feit blijft dat de bodemfysische karakteristieken moeilijk nauwkeurig te meten zijn, speciaal in ongestoorde gronden. Voor veel toepassingen kan de schatting van de bodemfysische karakteristieken met pedotransferfuncties een goed alternatief vormen voor de dure en lastige metingen.

Dit proefschrift richt zich op de verbetering van de bodemfysische invoergegevens in water- en stoffenmodellen binnen de hydrologie en milieuwetenschappen. Deze invoergegevens kunnen worden geschat uitgaande van verschillende gegevensbronnen. De voorspelkracht en betrouwbaarheid van de schattingen worden bestudeerd met speciale aandacht voor Hongarije als testland.

Wereldwijd worden er landbouwkundige en milieuproblemen gesignaleerd. Veel van deze problemen worden niet beperkt door nationale grenzen en vereisen voor hun oplossing dus internationale samenwerking. Voor de oplossing zijn vaak bodemgegevens nodig als invoergegeven voor modellen. Ondanks een aantal internationaal erkende standaards zijn bodemgegevens slechts zelden uitwisselbaar over nationale grenzen heen. Dit probleem doet zich ook voor bij het samenstellen van internationale gegevensbestanden. Maar een beperkt aantal bijdragende instituten/landen volgt strikt een internationaal erkend systeem zoals bijvoorbeeld de FAO. Hierdoor is interpolatie van gegevens noodzakelijk. Deeltjesgrootteverdeling is in dit verband een belangrijk invoergegeven in bodemfysische pedotransferfuncties. Het is ook een gegeven met erg verschillende nationale standaarden.

De geschiktheid is getest van vier verschillende interpolatieprocedures om uitwisselbaarheid te bewerkstelligen van de cumulatieve deeltjesgrootteverdeling zoals die volgens de standaard van verschillende landen zijn verzameld. De nauwkeurigheid van de verschillende procedures blijkt te verschillen afhankelijk van de afstand tussen meetpunten van de deeltjesgrootteverdeling. Loglineaire interpolatie is in het verleden in verschillende studies gebruikt, maar bleek van de vier geteste procedures de minst nauwkeurige schatting op te leveren. Gebruik van de Gompertz curve - dit is een speciaal type niet-symmetrische curve - toonde minder gevoeligheid voor de afstand tussen meetpunten. Interpolatie van een

aantal deeltjesgrootteverdelingen was echter niet voldoende nauwkeurig en deze procedure kon niet worden toegepast op deeltjesgrootteverdelingen waarvan het aantal gemeten klassefracties kleiner was dan het aantal modelparameters. Gebruik van een niet-parametrische splinefunctie toonde een aanzienlijke toename in nauwkeurigheid van de interpolatie met afnemende afstanden tussen meetpunten. Als nieuwe benadering is de ‘similarity procedure’ geïntroduceerd die geen enkele wiskundige interpolatiefunctie gebruikt. Het gebruikt externe bodeminformatie, waaruit gronden worden geselecteerd, met deeltjesgrootteverdelingen die overeenkomen met de verdeling van de betreffende grond. Deze similarity procedure gaf de meest nauwkeurige interpolatie. Indien er voldoende externe bodeminformatie beschikbaar is, vormt de similarity procedure een krachtig hulpmiddel voor interpolaties. Gebaseerd op het aantal en de verdeling van de meetpunten van de deeltjesgrootteverdeling is een algemene regel afgeleid wanneer het best de splinefunctie kan worden gebruikt en wanneer de nieuwe similarity procedure **(Hoofdstuk 2)**.

De beschikbaarheid van een gegevensbron die zowel potentiële schatters en bodemfysische karakteristieken bevat is een voorwaarde voor het afleiden van pedotransferfuncties (PTFs). Er bestaan al veel elektronische gegevensbronnen, maar de nadruk daarbij ligt meestal op classificatie en ze bevatten vaak maar weinig bodemfysische karakteristieken van de onverzadigde zone. Er zijn twee vormen van gebruik van de internationale bodemfysische gegevensbronnen. Allereerst kunnen ze een goed alternatief vormen voor kleinere (bijvoorbeeld nationale) gegevensbronnen bij het beantwoorden van onderzoeksvragen op het gebied van beheer van bodemwater met vooral een internationale dimensie. Ze kunnen een brug vormen tussen onderzoekers en andere potentiële gebruikers in verschillende landen/regio’s. Op de tweede plaats hebben veel landen of regio’s in de wereld niet de middelen om een voldoende aantal bodemfysische karakteristieken te verzamelen. In dergelijke gevallen zijn externe gegevensbronnen de enige mogelijkheid.

De internationale UNsaturated SOil hydraulic DAtabase (UNSODA) is ontwikkeld om praktijkmensen en onderzoekers te voorzien van bodemfysische en andere bodemkarakteristieken. De huidige gegevensbron omvat gemeten gegevens over waterretentie, doorlatendheid, water diffusiviteit en bodemkundige informatie over ongeveer 790 bodemmonsters van over de hele wereld. Als opvolger van de eerste MS-DOS versie van de gegevensbron is de tweede versie (UNSODA V2.0) in 1999 verschenen. De vorm en structuur van de nieuwe gegevensbron zijn aangepast om extra zoekopties mogelijk te maken, om de uitwisselbaarheid te vergroten met programma’s voor in- en uitvoer van gegevens en om gebruikers van dienst te zijn bij het zoeken en bij de grafische presentatie. De structuur en de inhoud van de gegevensbron worden kritisch beoordeeld, net zoals de bewerkingen die kunnen worden uigevoerd op de verschillende gegevens in UNSODA V2.0.

Er worden twee toepassingen getoond van het gebruik van de UNDOSA V2.0 gegevens. Het verwerven van gegevens wordt kort gedemonstreerd en wordt gevolgd door een meer gedetailleerd voorbeeld over de interpolatie van de deeltjesgrootteverdeling volgens verschillende nationale definities van deeltjesgrootteklassen. De in Hoofdstuk 1 geïntroduceerde ‘similarity procedure’ is getest met betrekking tot de geldigheid voor verschillende geografische gebieden. In de test is een grote Europese gegevensset gebruikt bij de interpolatie van de deeltjesgrootteverdeling van niet Europese gronden. De veronderstelling was dat de procedure niet goed zou voldoen voor gronden in andere geografische gebieden, maar deze veronderstelling bleek niet juist te zijn **(Hoofdstuk 3)**.

Bodems zijn belangrijke natuurlijke hulpbronnen en dragen in belangrijke mate bij aan de nationale economie van Hongarije. Verstandig gebruik van deze bodems en de instandhouding van hun functie zijn van belang voor de economie en voor het milieu, omdat bijna 70% van het landoppervlak landbouwkundig gebruikt wordt. In Hongarije zijn veel lokaal hydrologische studies uitgevoerd. In deze studies zijn veel bodemfysische karakteristieken verzameld. Echter maar een klein gedeelte hiervan is opgeslagen in een gegevensbron, waarbij de gegevens bovendien vaak incompleet zijn en uit verschillende bestanden komen. Hierdoor waren de gegevens slecht toegankelijk voor huidig en toekomstig onderzoek, terwijl ook het toepassingsbereik beperkt is.

De nieuwe Hongaarse UNsaturated SOil DAtabase (HUNSODA) wordt geïntroduceerd in **Hoofdstuk 4**. Naast nadruk op bodemfysische karakteristieken dient de gegevensbron ook als opslagmedium voor andere bodemgegevens van 840 bodemhorizonten waarvan 576 bodemhorizonten met waterretentiekarakteristieken. De gegevensbron gebruikt SQL (Standard Query Language) voor de communicatie, waarmee het uitwisselbaar is met veel andere gegevensbronnen. De structuur en de inhoud van de gegevensbron worden beschreven, inclusief een eenvoudige handleiding voor gebruik.

De gegevens zijn gebruikt om klassevertaalfuncties af te leiden voor Hongaarse gronden volgens de FAO en USDA textuurklassen. Alhoewel er verschillen bestaan tussen de functies voor de Medium, Medium-Fine en Fine textuurklassen van FAO zijn er ook belangrijke overeenkomsten tussen de functies voor deze klassen. Als de USDA classificatie wordt gebruikt komen de functies voor alle klassen, behalve grof en fijn, sterk met elkaar overeen. Dit duidt erop dat er voor het afleiden van PTFs geen behoefte is aan vele, gedetailleerde textuurklassen.

De toepasbaarheid van de verkregen informatie moet worden getest. Recentelijk zijn er veel PTFs ontwikkeld en onafhankelijke gegevens zijn gebruikt om de verschillende PTFs te testen. Er bestaan twee belangrijke beperkingen in het testen van PTFs. Allereerst blijft het onduidelijk wat de belangrijkste bronnen zijn van de schattingsfouten. In deze testen is het niet duidelijk of de ene PTF beter presteert dan de andere, doordat voor het afleiden van PTFs verschillende soorten (omvang, locatie, betrouwbaarheid) gegevens worden gebruikt, verschillende technieken (bijvoorbeeld regressie- analyse en neurale netwerken) worden gehanteerd of verschillende verklarende variabelen worden gebruikt. De tweede beperking is dat PTFs bijna uitsluitend worden getest aan de hand van waterretentieparameters of punten. Waterretentie en doorlatendheid zijn geen doel op zich maar zijn hulpkarakteristieken om meer praktische bodemeigenschappen te berekenen. Functionele evaluatie van de geschatte bodemfysische karakteristieken helpt om de bijdrage van deze karakteristieken aan de onzekerheid van de simulatieberekeningen te kwantificeren.

Hongaarse gegevens worden gebruikt om bodemfysische PTFs op nationale schaal af te leiden. Een set uit de HYPRES gegevensbron is gebruikt om continentale PTFs af te leiden. Tenslotte is een set bestaande uit vooral Amerikaanse en Europese gegevens gebruikt om intercontinentale PTFs af te leiden. De drie sets - plus nog twee andere sets verkregen door een combinatie van de drie sets – zijn gebruikt om met neurale netwerken, elf verschillende PTFs af te leiden. Voor elke set zijn hetzelfde rekenschema en dezelfde schatters gebruikt. De PTF modellen variëren van eenvoudig tot complex. Alle PTFs zijn getest door gebruik van een tweede, onafhankelijke, Hongaarse gegevensset. De geschiktheid van internationale en nationale PTFs is allereerst getest aan de hand van de correlatie tussen de gemeten en berekende waterretentiekarakteristieken. Vervolgens zijn de verschillende PTFs gebruikt om het vochtverloop te simuleren in zeven Hongaarse gronden. De correlatie is vastgesteld tussen de berekende en de in het veld gemeten vochtgehaltes.

Schattingen worden beter naarmate het aantal variabelen toeneemt. Schattingen van de waterretentiekarakteristiek resulteren in root mean squared residuals (RMSR) – bij gebruik van verschillende sets invoergegevens – variërend van 0.02 tot 0.06 m^3m^{-3} voor PTFs op nationale schaal, terwijl PTFs op internationale schaal RMSRs hadden van 0.025 tot 0.088 m^3m^{-3}. Voor gronden die ondervertegenwoordigd zijn in de PTF afleiding werden grotere fouten gevonden. Het resultaat toont aan dat het beschikbaar hebben van een kleine relevante gegevensset beter is dan het gebruik van een grote meer algemene gegevensset.

De geschatte waterretentiekarakteristieken zijn vervolgens gebruikt om het vochtverloop te simuleren in zeven verschillende Hongaarse gronden. Er worden kleine verschillen gevonden tussen de verschillende PTFs. Root mean squared residuals over het groeiseizoen variëren van 0.065 tot 0.07 m^3m^{-3} terwijl berekeningen met in het laboratorium gemeten waterretentiekarakteristieken een fout hadden van 0.061 m^3m^{-3}. De verschillen tussen de nationale en internationale PTFs verdwenen grotendeels als de geschatte waterretentiekarakteristieken worden gebruikt voor de simulatie van vochtgehaltes. Verschillen in schattingen bij gebruik van gegevensets op verschillende schalen waren niet de belangrijkste foutenbronnen in de simulaties **(Hoofdstuk 5)**.

Scenariostudies zijn belangrijk bij planning en preventie. Kennisexpertise kan binnen dergelijke studies verder worden verfijnd om een kwantitatieve inschatting mogelijk te maken van de voorgestelde veranderingen. Veldexperimenten voor verschillende managementsystemen zijn tijdrovend en soms zelfs onmogelijk. Pedotransferfuncties kunnen tegen relatief lage kosten bodemfysische karakteristieken voorspellen. Er zijn echter maar weinig studies die onderzoeken welke bijdrage PTFs kunnen leveren aan verkennende planning door het effect na te gaan van bepaalde veranderingen in het bodem- en waterbeheer.

Er worden drie scenariostudies beschreven om aan te tonen hoe PTFs kunnen helpen om landgebruiksystemen te verbeteren. De drie voorbeelden hebben betrekking op realistische problemen waarmee zowel Hongarije als veel andere landen te kampen hebben. Verkennende modellering is gebruikt om:

i. Na te gaan wat de voor- en nadelen zijn van irrigatie en de rol die bodemheterogeniteit hierbij speelt;
ii. Na te gaan welke veranderingen kunnen worden verwacht – in termen van de waterbalans – als de bulkdichtheid en/of het organisch stofgehalte veranderen tengevolge van alternatieve managementsystemen; en
iii. Na te gaan wat de kans op uitspoeling naar diepere horizonten is in een aantal Hongaarse gronden.

Het resultaat van de verschillende studies is dat:

i. Het mogelijk is een kwantitatieve inschatting te maken van de voordelen en de risico's van de verschillende irrigatieplannen. Het effect van bodemheterogeniteit binnen dezelfde bodemeenheid wordt duidelijk doordat delen van het heterogene veld erg gevoelig zijn voor uitspoeling en andere delen gevoelig zijn voor droogte.

ii. Veranderingen in landgebruik die leiden tot lagere bulkdichtheden (bijvoorbeeld ploegen) en toenemend organisch stofgehalte hebben een tegengesteld effect op de bodem in termen van stroomsnelheden, vochttekorten en de kans op uitspoeling van water (en nutriënten) naar diepere bodemhorizonten.
iii. Uitgaande van 20 jaar simulatie, konden de tien Hongaarse gronden worden gegroepeerd in drie groepen op basis van dagelijkse hoeveelheden en jaarlijkse patronen in de waterflux op een diepte van 40 cm.

Met verkennend onderzoek worden kwantitatieve antwoorden gegeven op het type ‘wat als’ vragen, waarbij trends kunnen worden onderscheiden en potentiële problemen en successen zichtbaar worden die kunnen bijdragen aan de ontwikkeling van duurzame managementsystemen **(Hoofdstuk 6)**.

Naast het beschreven onderzoek nam de kandidaat als onderzoeksstudent deel aan het project “Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning”, medegefinancierd door de Europese Unie (project no. CHRX-CT94-0639). Dit project resulteerde in de samenstelling van de HYPRES (Hydraulic Properties of European Soils) gegevensbron op Europese schaal. De volledige resultaten van dit project worden niet gepresenteerd. De kandidaat was medeauteur van een wetenschappelijk artikel (Wösten et al, 1999) en verschillende conferentieartikelen die gebaseerd zijn op dit onderzoek. Een deel van het onderzoek is beschreven in Hoofdstuk 2 en de gegevensbron is later gebruikt voor het werk beschreven in Hoofdstuk 5.

Curriculum Vitae

Attila Nemes was born on 14 November 1970 in Miskolc, Hungary. He graduated in Agricultural Sciences (specialised to Tropical and Sub-Tropical Agriculture) at the Agricultural University (at present Szent István University), in Gödöllő, Hungary. He completed his MSc thesis research at the DLO Winand Staring Centre (currently ALTERRA) in Wageningen.

In 1994 he was accepted to the PhD program of the University of Gödöllő and became employed at the Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences. He became infected by international research and - in 1996 - became a research student in the EU funded project entitled "Using existing soil data to derive hydraulic parameters for simulation models in environmental studies and in land use planning" (project no. CHRX-CT94-0639). Later, in 1999 he was admitted to the "Sandwich PhD programme" of the Wageningen Agricultural University (currently Wageningen University and Research Centre). He participated in a number of research projects in collaboration with Purdue University (West Lafayette, IN, USA), USDA George E. Brown Salinity Laboratory (Riverside, CA, USA), USDA Hydrology and Remote Sensing Laboratory (Beltsville, MD, USA), Moscow State University of Environmental Enginering (Moscow, Russia).

Time passed, and four out of five of his schools/institutes even changed their names, but he held on firmly to his PhD studentship. During his ever-lasting studies he was supported by 17 different fellowships and grants. Upon receiving one of these fellowships, he was awarded a certificate by then Minister of Education, Culture and Science of the Netherlands, Mr. L.M.H.A. Hermans, for summarizing the Dutch way of living – from a foreigner's standpoint - in the following 5 phrases:

Let's get wet!
Do it cheap!
Find your bike!
Eat some cheese!
Be Orange!

Currently he is a visiting postdoctoral scientist at the USDA-ARS Hydrology and Remote Sensing Laboratory in Beltsville, Maryland, USA.

Propositions

1. It is difficult to define a single optimal technique to interpolate soil particle-size distribution data.

   /this thesis/

2. International and internationally compatible national soil databases can help in finding solutions for environmental problems that are not restricted to manmade borders.

   /this thesis/

3. Pedotransfer function (PTF) comparisons often don't account for the fact that compared functions are results of different methodology, databases and inputs. A lot more can be learned if these factors are changed one by one.

   /this thesis/

4. Reported differences between estimates of different PTFs may not be so important when they are applied in simulations. Other sources of error may dominate PTF errors.

   /this thesis/

5. As long as the main alternative to PTFs is the use of laboratory measured soil hydraulic data while simulating field soil conditions we may not make a big mistake using PTF estimates.

   /this thesis/

6. Central and Eastern European countries possess a wealth of environmental data. Unfortunately, the continuity of data collection and data sharing has been hampered by recent changes in the financial situation of science.

7. The scientist is not a person who gives the right answers; he is one who asks the right questions.

   /Claude Levi-Strauss/

8. Most computer users owe their eyes a better lighting ergonomics. Monitor settings and lighting are often overlooked. Low-glare working environment, use of a local source of light instead of ceiling lights and changing background, brightness, contrast, and refresh-rate settings of the monitor should be considered.

9. People in democracies have the right not to believe what is said via the media. This right could be exercised more frequently.

10. Who points at someone with a finger should never forget that 3 of his fingers point at him!

11. Computers are good in many ways. But let's not forget to learn that the "Undo" button does not exist in real life.

12. By the time you learn how to write a thesis you realize that yours had already been written.

Attila Nemes – Multi-scale hydraulic pedotransfer functions for Hungarian soils
Wageningen, 1 April 2003